Developing an Objective Definition of Simulation Fidelity for Enroute Air Traffic Control

by

Colin Dow

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Applied Science

in

Systems Design Engineering

Waterloo, Ontario, Canada, 2015

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The domain of enroute Air Traffic Control (ATC) relies heavily on simulation for a variety of purposes. However, little research has been conducted in this particular domain to determine the link between fidelity and how simulation is used. This thesis introduces the first definition of simulation fidelity for the enroute ATC domain; it also presents a first standardized simulation environment categorization system. These are important foundational steps, as an online survey of 86 ATC industry professionals found that a significant majority believe that simulation fidelity is not well defined for enroute ATC.

An initial definition of simulation for enroute ATC was developed based on documentation regarding the current enroute ATC operational environment and previous research experience in the enroute ATC domain. This definition underwent a preliminary validation during semi structured interviews conducted at an air navigation service provider (ANSP), where all 13 interviewees believed that the definition capture the environment components that can affect the fidelity of enroute ATC simulation. Subsequently, almost 85% of the 86 industry professionals surveyed at least 'Agreed' with the components in the definition, with no significant differences with regards to this agreement within the demographic groups of nationality, primary use of simulation, gender and years of experience working simulation. The definition helps to reduce the ambiguity and confusion around the concept of simulation fidelity within the domain of enroute ATC, and potentially provide the foundation for further investigation into the links between fidelity and simulation use within the ATC industry.

A categorization system, similar to that used by the FAA for categorizing flight simulators, was then developed in order to operationalize the fidelity definition into five categories differentiating the fidelity of enroute ATC simulation environments. During the validation of this construct, a key limitation was identified in that, as it is currently structured, simulation environments can fall under more than one category. Potential modifications and future iterations of the categorization system are discussed.

In addition, industry perceptions regarding how simulation of varying degrees of fidelity ought to be used depending on the task to be accomplished are presented and discussed. The perceptions indicate a strong desire to rely heavily on higher fidelity simulation to accomplish training, testing new operational concepts and researching human factors issues with few instances of support for lower fidelity simulation. However, these perceptions do not necessarily represent best practices. This investigation is meant to stimulate discussion of how simulation is currently used within the industry as well as offer potential areas for further research to determine if there are other options to the status quo.

Acknowledgements

First and foremost I would like to acknowledge my supervisor, Dr. Jonathan Histon from the Systems Design Engineering Department at the University of Waterloo. Without his support, guidance and belief, this thesis could never have been completed to the highest standard. I am extremely grateful and privileged to have had him guide my progress as a human factors professional.

I would also like to thank all of the individuals from the air navigation service provider that I worked with during my research and who contributed to the work in this thesis. Their generous support and help made this research possible.

I would like to thank Dr. Robert Duimering and Dr. Stephen Waslander for being the readers of my thesis.

I gratefully acknowledge the National Science and Engineering Research Council of Canada for funding this research project through the Engage Grant and the Engage Plus Grant.

To Mom, Dad and my dearest Kathryn:

I am eternally grateful for your support and love, and would not be where I am today without you.

I dedicate this thesis to you.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ANSP – Air Navigation Service Provider

ATC – Air Traffic Control

ATM – Air Traffic Management

FAA – Federal Aviation Administration

FFS – Full Flight Simulator

FTD – Flight Training Device

KSAs – Knowledge, skills and abilities

ICAO – International Civil Aviation Organization

OJT – On-the-Job Training

SME – Subject Matter Expert

Chapter 1

Introduction

There are more aircraft in the air today than ever before as air traffic continues to grow. In 2007, the International Civil Aviation Organisation (ICAO) forecasted that passenger air traffic would grow at approximately 4.6% annually and freight air traffic at 6.6% through 2025 (ICAO, 2007). In 2001 there were 3.011 trillion passenger-kilometers performed, a number that by 2011 had grown to 5.062 trillion, an increase of approximately 68% in only 10 years (ICAO, 2012).

Air navigation service providers (ANSPs) around the world are currently developing and implementing air traffic management (ATM) systems that can handle these increases while remaining both safe and efficient. Some of the technologies and tools that are part of the new ATM system include but are not limited to: Automatic Dependent Surveillance-Broadcast (ADS-B), Controller-Pilot DataLink Communication (CPDLC), System Wide Information Management (SWIM), Collaborative Decision Making (CDM) and various decision support tools (FAA, 2012; Eurocontrol, 2012). For an in-depth analysis of how the operational environment is changing due to technological upgrades, see Durso and Manning (2008). With the proposed upgrades the controller in the future ATM system will have a greater amount of information at their disposal and it is believed that the system will use this information to generate safer and more efficient air travel.

However, these current and future air navigation system improvements will cause an increase in system automation and complexity, leading to significant changes in an air traffic controller's operational environment (Durso and Manning, 2008; Hilburn et al., 2006; Blanken, 2002). These changes to the operational environment are triggering the need for re-evaluation of the knowledge, skills and abilities (KSAs) required for successful controller performance in the modernized system. Compounding this issue is that the new tools and technologies are being implemented in the environment quicker than training programs can adapt to train individuals for the new working conditions.

It is important that the training programs and processes evolve with the system; however, this is not as simple as it seems given the scale and complexity of the changes being made to an air traffic controller's operational environment, not to mention the speed at which these changes are implemented. Independent reviews of the Federal Aviation Administration's (FAA) ATC training program highlighted the need for further research and work to improve their training processes. It was found that while they have made an effort to update their training processes, their current training program is still insufficient to handle all the demands of emerging issues created by the future ATM system (Barr et al., 2011; USGAO, 2008).

1.1 Simulation Use in Training

Simulation use has become ubiquitous in training programs for complex, safety-critical systems like ATC. It provides a more affordable and significantly safer environment in which to conduct training, thus leading to its increased use in domains such as aviation, medicine, military, and process control systems among others (Moroney and Lilienthal, 2008).

However, the term *simulation* includes a wide variety of different types of environments within a particular domain. Simply by changing the fidelity, or realism, of a simulation, an assortment of training environments become available. The following high-level definition of simulation fidelity is used as the foundation for the research conducted in this thesis:

> Simulation fidelity is the degree of similarity between the training situation and the operational situation which is simulated. It is a two dimensional measurement of this similarity in terms of: (1) the physical characteristics, for example, visual, spatial, kinesthetic, [auditory], etc.; and (2) the functional characteristics, for example the informational, and stimulus and response options of the training situation. (Hays and Singer, 1989, p.50)

An important open research question then is: what levels of simulation fidelity are required to achieve specific outcomes for the different uses of simulation and how should the various simulations be employed? There are two significant challenges in answering this question within a given domain.

First, there is the need for a standardized definition and understanding of simulation fidelity that applies to the domain. Without a standardized definition, individuals will create their own definitions of what affects fidelity and it is likely that variance across these definitions will exist. This variance introduces a certain amount of ambiguity in discussions about how simulation of varying degrees of fidelity ought to be used. A more consistent and shared definition of simulation fidelity helps to eliminate differing representations of simulation fidelity by providing a widely-accepted point of reference that enables a clearer discussion to take place regarding how fidelity affects a simulation's usefulness.

Second, there is a need to map the domain-specific levels of simulation fidelity to training particular skills. The small amount of literature that does begin to identify the specific effects of simulation fidelity on training outcomes is mostly domain-specific, making it challenging to generalize findings to other domains. Domain-specific research is required to objectively and clearly identify the effects of varying the fidelity of different components of the operational environment on learning specific skills, such as applying procedural knowledge or complex problem solving. A detailed discussion regarding the effects of simulation fidelity on training is provided in Chapter 2.

These two challenges tend to result in a variety of different approaches to using simulation and potentially create inconsistent and inefficient training. Answering the research question above would improve the selection and use of simulation in training programs and most likely improve overall training effectiveness and efficiency; this represents an important part of the development of a training program capable of training individuals for the new operational environment. Even once the two challenges above have been investigated, it must be determined whether or not the findings from other domains translate and apply to the ATC domain that is the focus of this thesis.

1.2 Simulation Use in ATC

Based on discussions with SMEs early in the research process, simulation-based training was identified as a key issue with regards to training individuals for the future operational environment in ATC. The following definition was used to determine what constitutes a simulation in the domain of enroute ATC within the context of this thesis: any situation where an individual actively practices providing some part or all of the air traffic services provided by controllers in the operational environment. Typically this is achieved through the use of some form and/or combination of tools, objects, or personnel to replicate parts of the real world task environment.

Determining how to most effectively use simulation is a high priority in the current and future training programs for ATC, as it allows trainees to develop and refine the needed KSAs in a controlled and safe replication of the operational environment. A significant challenge with this task, introduced in Section 1.1, is how to incorporate simulation into ATC training programs effectively as there is little objective research within the domain on which to base these decisions; this potentially increases costs due to unnecessary training being conducted, and/or less efficient training processes being used. Research is needed to provide a basis for clear, justified training standards; ensure that training resources are being used efficiently and effectively; and to ensure that the ATC training process is developing the KSAs needed in current and near-term operational environments.

A key challenge for ANSPs in this effort is categorizing the fidelity of existing simulations, and developing a better understanding of how to match the fidelity of a simulation with training objectives. With a multitude of options available simply by changing the fidelity of a simulation, the challenge then becomes determining the level of simulation fidelity or combination of different levels of simulation fidelity that is most effective and efficient at training individuals to proficiency for different skills.

Previous research has shown that higher fidelity simulation is not always necessary or desirable, and that the optimal level of fidelity depends on the skills being learned (Dahlstrom et al, 2009). In order to compare simulation of varying degrees of fidelity, an objective definition of what fidelity means in the context of ATC is required. This will enable researchers to be able to critically and objectively compare the different simulations available, making it possible to objectively determine which simulations are best at training what types of skills as well as what parts of the ATC operational environment affect the learning of those different types of skills.

To limit the scope of this thesis, the domain of focus will be the domain specialty of enroute ATC, defined here as any type of radar-based ATC such as terminal, low enroute and high enroute depending on where the aircraft is during its flight. This means there is no direct line of sight with the aircraft and the controller relies on radar feedback to establish safe separation between aircraft. This is in contrast to tower control, which relies primarily on visual contact with aircraft from an elevated tower at an aerodrome to ensure safe separation (e.g. aircraft landing, departing, as well as ground taxiing). Enroute ATC is specified for this research as it offers a domain where limited research on simulation fidelity has been conducted and where the results of such work can have a significant impact on simulation use.

However, as demonstrated in Chapter 2, there has been little research conducted in the field of simulation fidelity within the domain of enroute ATC. Establishing a simulation fidelity definition specific to enroute ATC would provide a formal reference point when discussing simulation fidelity, thus introducing more clarity to a naturally ambiguous concept and eliminating the use of vague labels such as 'high, medium, and low.' This would also allow for the creation of a standardized simulation environment categorization system, similar to those used by the FAA for flight simulators, which could improve the consistency of how simulation of varying degrees of fidelity are used across the industry. While the main impetus for the work in this thesis arose from simulation use within a training context, establishing a definition would be a valuable tool for a range of applications beyond just training such as determining cost-effective ways of assessing new operational concepts, tools and procedures or identifying the effects simulation fidelity has on the applicability of research findings.

1.3 Research Objectives

The primary goal of this thesis is to develop and validate an objective enroute ATC simulation fidelity definition and categorization system, key tools in performing the aforementioned research. To achieve this goal, this thesis has the following four objectives.

Objective 1 – Identify how people within the domain of ATC currently make simulation fidelity determinations.

To better understand the concept of simulation fidelity within the domain of ATC, it is necessary to understand how people in the industry currently view the concept and make determinations regarding the fidelity of a simulation. Accomplishing this objective will help identify the specific environmental components individuals are considering when making fidelity determinations. This will allow for the creation of a definition capable of making objective comparisons with regards to the fidelity of various simulations.

In order to accomplish this objective, input from professionals who use simulation within the industry was required. Two methods were used to gather data from these professionals: (1) interviews were conducted with training professionals at an ANSP who use simulation regularly, and (2) an industry-wide survey of perceptions on the topic of simulation fidelity was distributed online.

Objective 2 – Develop an objective enroute ATC simulation fidelity definition.

A key tool needed to perform research on the effects of simulation fidelity on training outcomes is an objective definition of simulation fidelity for the domain of enroute ATC. This definition would identify the components of the operational environment that can affect simulation fidelity and allow for objective comparisons between simulations along those components.

In order to accomplish this objective, an initial definition was developed based on personal experiences with the domain of enroute ATC and literature describing the operational environment. This definition was presented to SMEs for initial feedback and was subsequently refined. It was then presented to interviewees during the site visits at the ANSP for an preliminary validation.

Objective 3 – Develop an enroute ATC simulation environment categorization system.

The definition itself does not compare simulations but provides the points of comparison. What is needed is an operationalized version of the definition in the form of a simulation categorization system that provides an easy tool for differentiating between the fidelity of simulations. As a first step, and in keeping with traditions found in other domains, a categorization system focused on simulation environments is developed.

In order to accomplish this objective, questions were posed to interviewees regarding the potential for a categorization system similar to the one used by the FAA for categorizing flight simulators but for enroute ATC simulations and what a categorization such as this might look like. This began to inform the development of the categorization system, which was refined after the site visits were complete.

Objective 4 – Validate the enroute ATC simulation fidelity definition and categorization system within a diverse sample of the industry.

In order to ensure the two fidelity constructs are not organization-specific to the ANSP, a validation exercise is used to conduct an initial assessment of whether the constructs would be useful to a wide variety of potential users. To accomplish this objective, a survey was created that was disseminated across the industry to ANSPs in different countries and researchers who specialized in the domain of ATC. The survey presented the two fidelity constructs to participants in order to receive critical feedback from SMEs and validate the constructs. Participants for the survey were those who used simulation for training, testing new operational tools or concepts, or future ATC environment research; these participants were targeted to gain the widest possible acceptance of the two fidelity constructs.

Completing each of these objectives would result in an objective and widely-accepted definition of simulation fidelity for enroute ATC, as well as a framework capable of objectively categorizing simulations that already exist into separate, well-defined levels of fidelity rather than the use of vague fidelity terms such as low, medium and high. These fidelity constructs then allow for more in depth study of the effects of simulation fidelity on training in the domain of enroute ATC, most specifically in trying to determine what parts of the operational environment are required at what level of fidelity in order to train certain skills. These constructs form the foundational elements of this larger scale research as they offer an objective representation of the different components of the operational environment that affect fidelity, and how these components vary across the different levels of simulation fidelity most commonly used in the ATC industry.

1.4 Thesis Organization

The remainder of the thesis is organized as follows:

- **Chapter 2: Background** contains a review of the relevant literature regarding the concept of simulation fidelity, more specifically the process required to define it objectively for a given domain and simulation fidelity's effects on training outcomes.
- **Chapter 3: Approach** contains a detailed description of the approach taken to achieve the research objectives enumerated in section *1.3 Research Objectives.*
- **Chapter 4: Understanding Simulation Fidelity in Enroute ATC Based on Simulation Use for Controller Training** seeks to develop a better understanding of the concept of simulation fidelity within the domain of study for this thesis. A case study approach is used, focusing on simulation used for training at an ANSP. Limitations of the highest fidelity simulation and comparisons between the simulation and the operational

environment are used to determine how individuals are making determinations about the fidelity of simulations. This chapter contributes to accomplishing the first objective as stated in section *1.3 Research Objectives*.

- **Chapter 5: ATC Industry Perceptions Regarding Simulation Fidelity** presents the different perceptions of the concept of simulation fidelity within the ATC industry, the confusion that arises from people's different perceptions of simulation fidelity and the appropriateness of a categorization system for the enroute ATC domain. This chapter provides the primary motivation as to why the fidelity constructs are required and also serves to accomplish the first objective.
- **Chapter 6: Enroute ATC Simulation Fidelity Definition** consists of four sections: the development process of the simulation fidelity definition for enroute ATC, a detailed explanation of the definition itself, the validation of the simulation fidelity definition, and finally a discussion regarding choices made while developing the definition and how it will most likely be used. This chapter and Chapter 7 accomplish objectives 2 and 4 from Section 1.3.
- **Chapter 7: Enroute ATC Simulation Environment Categorization System** follows the same format as Chapter 6 but focuses on the simulation environment categorization system. This chapter accomplishes objectives 3 and 4 from Section 1.3.
- **Chapter 8: Other Findings from the Research** presents current industry perceptions regarding the choice of a simulation of a particular level of fidelity to accomplish a particular task.
- **Chapter 9: Conclusions and Future Work** summarizes the final conclusions drawn from the research conducted in the completion of this thesis, as well as detailing potential avenues of further research using the foundational work laid out here.

Chapter 2

Background and Literature Review

This chapter presents the background material that motivates the need for a simulation fidelity definition and categorization system in the enroute ATC domain. It is divided into four sections: Approach to the Literature Review, Understanding Simulation Fidelity, Examples of Classifying Simulation Environments in the Literature, and Impact of Simulation Fidelity on Training.

The first section explains the process used to conduct the literature review, which includes search methods and criteria for including or using a particular piece of literature. The second section presents and discusses the most pertinent literature to understanding the concept of simulation fidelity and highlights some important considerations for developing a simulation fidelity definition. The third section offers a brief examination of simulation environment categorization systems in different industries and how they can offer potential examples to developing a categorization system for enroute ATC. The final section offers a summary of the available literature pertaining to the impact of simulation fidelity on training.

2.1 Approach to the Literature Review

The main goals of this literature review were to develop an understanding of the effects of simulation fidelity on training, how simulation fidelity has been defined in other domains and to identify examples of simulation environment categorization systems on which to base a categorization system of enroute ATC simulation environments. Sources were reviewed from a variety of domains including aviation, medicine, military, process control systems, business and others. Considering the variety of domains researching simulation fidelity, it is believed that analyzing literature from across these domains rather focusing on one specific domain, such as flight simulation, provides a more robust understanding of simulation fidelity and thereby a potentially stronger simulation fidelity definition for enroute ATC. The process of gathering the relevant literature focused on searching for documents using the terms ‘simulation fidelity’ or ‘simulator fidelity’ across several research databases which included the FAA Human Factors library, Sage Publications (sample journal: Human Factors), Taylor and Francis Online (sample journal: International Journal of Aviation Psychology), IEEE and Google Scholar. These databases taken as a whole offer the largest possible pool of human factors research periodicals, and therefore ought to provide the greatest amount of literature on the topic of simulation fidelity.

Once the key works were identified from this initial search, the pool of literature was broadened even further by tracking back through the works cited of the most relevant articles and reports as well as searching forward as to works that had subsequently referenced these key works. A clear effort was made to use the most recent work on the topic, but much of the most relevant work on simulation fidelity was conducted during the period of 1988 to 2002 and thus is relied upon throughout the literature review. A variety of sources were included such as journal articles, conference papers, technical reports and book chapters. Journals sourced for this review include Theoretical Issues in Ergonomics Science, Quality and Safety in Health Care, The International Journal of Aviation Psychology, Human Factors, Medical Education, and others.

2.2 Understanding Simulation Fidelity

This section presents a more general understanding of fidelity, exploring the myriad of terms used to describe fidelity, the different environmental components that affect a simulation's fidelity, and how one arrives at a formal, objective definition of fidelity for a particular work environment. This section relies primarily on the review of simulation fidelity performed by Hays and Singer in their book *Simulation fidelity in training system design: Bridging the gap between reality and training* (1989). This book is a seminal piece of research on the topic of simulation fidelity and much of the research that has occurred since has either referenced its findings or has restated its conclusions (i.e. Feinstein et al, 2001). Due to the lack of general simulation fidelity research that occurred after 1990, Hays and Singer's book offers an in depth review and analysis of the concept of simulation fidelity and its history that is still relevant today, and should therefore be consulted if a more in depth historical review of the concept is desired.

The general term of simulation fidelity can be best understood from the definition posited by Hays and Singer in 1989 (p. 50): "Simulation fidelity is the degree of similarity between the training situation and the operational situation which is simulated." This is the simplest definition of fidelity and one which is now widely accepted (Liu et al., 2008). Fidelity, however, is not a singular concept, but can be sub-divided into high-level components that correspond to different parts of the operational environment. Recent work has seen a convergence on the use of three main components of simulation fidelity: physical, functional, and psychological fidelity (van Merrienboer and Kirschner, 2013; Estock et al., 2006; Alexander et al., 2005; Hays and Singer, 1989; Allen et al. 1986). The definition of each of these high-level components is provided in Table 2.1.

Table 2.1 Currently accepted high-level components of simulation fidelity.

Fidelity Component	Description
Physical fidelity	"The degree to which real-world operational equipment is reproduced in a simulated task environment (e.g. looks like, smells like, and feels like)." (Van Merrienboer & Kirschner, 2013)
Functional fidelity	"The degree to which a simulated task environment behaves in a way similar to the real task environment in reaction to the tasks executed by the learner." (Van Merrienboer & Kirschner, 2013)
Psychological fidelity	"The degree to which training tasks reproduce actual behaviours or behavioural processes required in real-life tasks." (Van Merrienboer & Kirschner, 2013)

From these three components, researchers can then identify the more specific fidelity components of a given work environment that fall under each high-level component described in Table 2.1. Each particular work environment will yield different environmental fidelity components, meaning a cockpit, a nuclear power control station, an operating room, or an air traffic control workstation will all have different, domain-specific fidelity components.

While the three categories or dimensions of fidelity detailed above do bring some consistency to the current understanding of fidelity, the psychological component has not been included in the work presented in this thesis. Physical and functional fidelity should be seen as essential components of fidelity measurement given their relevance to how workers perform tasks whatever the environmental context. Most importantly, these two categories of components of the operational environment are capable of being directly manipulated or controlled by simulation designers and their level of fidelity measured objectively.

Psychological fidelity, on the other hand, is a category of fidelity inherent to the user of the simulation. It is best described as the cognitive reaction from the user of the simulation induced by the manipulation of the physical and functional fidelity characteristics of the simulation, a second-order effect that is not directly controlled by simulation designers. This raises questions as to its appropriateness when discussing simulation fidelity and it is put forth here that it should be removed from the fidelity lexicon, as others have previously noted:

> "The term 'fidelity' should be restricted to descriptions of the required configuration of the training situation and not be used when discussing behaviors… The issue of training fidelity only becomes muddled if we attempt to use the same term to cover all of the interactive variables in the training situation. This is not to say we should throw out behavioral concepts, rather, we should use the standard labels for these concepts and not confuse them with fidelity." (Hays and Singer, 1989, p.49)

This is an important distinction to make and one that has been lost in the current body of literature on fidelity (e.g. both Van Merrienboer & Kirschner, 2013 and Estock et al, 2006 include this component without discussion of its merit). Psychological fidelity touches upon issues of user acceptance, simulation immersiveness, physiological reactions, and cognitive load among others; these should be regarded as important factors when researching simulation, but they should be also be treated as separate from the concept of simulation fidelity. Physical and functional fidelity represent elements of the operational environment that can be directly manipulated by those designing simulations; with psychological fidelity, there is no ability to directly control a user's cognitive reaction to a simulation as it is a secondary effect of the simulation itself. It is for this reason that in *Chapter 6 The Enroute ATC Simulation Fidelity Definition*, these psychological considerations have been excluded in the creation of the ATC simulation fidelity definition.

Therefore, the following high-level definition of simulation fidelity, as presented in Section 1.1, is used as the foundation for a proposed domain-specific ATC simulation fidelity definition:

> Simulation fidelity is the degree of similarity between the training situation and the operational situation which is simulated. It is a two dimensional measurement of this similarity in terms of: (1) the physical characteristics,

> for example, visual, spatial, kinesthetic, [auditory], etc.; and (2) the functional characteristics, for example the informational, and stimulus and response options of the training situation. (Hays and Singer, 1989, p.50)

Fidelity for a given simulation is based upon the components of the operational environment subject matter experts believe are relevant to performing the job, and therefore relevant in the simulation (Hays and Singer, 1989). The definition above identifies the scope of what will be considered when determining which components or parts of the live operations environment affect the fidelity of a simulation of those live operations.

Another approach to defining fidelity is to attempt to quantify it, essentially reducing it to a formula where levels of fidelity can be objectively and numerically calculated. While the notion of an objective formula to measure fidelity is an enticing one, there are several reasons why this has not yet been achieved and is very challenging. Firstly, creating a mathematical model for simulation fidelity is complex; there are a large number of potential influences and variables that need to be accounted for (Liu et al., 2008). Cognitive complexity, a popular area of focus in ATC research, suffers from similar issues in attempts to develop a quantified formula in order to measure the changing complexity of an ATC scenario (Hilburn, 2004).

Secondly, even if it were possible to generate a simpler mathematical model, would it be useful to those making decisions regarding what level of fidelity is needed in training? This process is not just about defining simulation fidelity, but making the concept easily accessible to those who use simulation and for them to be able to discuss the topic with clarity. Fidelity definitions must be structured in a way anyone in the domain can understand, and the primary focus of any who seek to define fidelity for a given work environment should have the target users of the definition in mind.

While the high level working fidelity definition helps to give scope as to what ought to be considered when discussing simulation fidelity, in its current form it does not specify what elements within a given work environment affect the fidelity of a simulation of that work environment. An example of this process of developing a domain-specific definition of simulation fidelity can be found in Estock et al. (2006), where they sought to identify and refine specific environmental components (Estock et al. refer to these as dimensions rather than components) that they believe affected the fidelity of a simulation of an F-16 cockpit:

> "Researchers have further divided these types of fidelity [the three dimensions identified in Table 2.1] into distinct dimensions. For example, Lee (2005) decomposed physical fidelity into the dimensions of visual scene simulation, sound effects and communication simulation, whole body motion, and handling qualities and control loading. Furthermore, several researchers have identified specific subcategories within these fidelity dimensions. For example, Heintzman, Middendorf, and Basinger (1999) separated motion cues into maneuver cues and disturbance cues (p. 4)."

Some of the components Estock et al. have identified are unique to their work environment, such as the visual scene simulation and whole body motion, while others such as communication simulation are important in a variety of work environments. This demonstrates the contextual nature of simulation fidelity definitions, as many of these components are appropriate for an F-16 cockpit but not an ATC workstation.

An important aspect of the process specified by Estock et al. is that once they had identified their fidelity components, they were verified by consulting with flight simulation experts to determine their validity. As identified by Hays and Singer (1989), receiving feedback from subject matter experts is an important step in defining simulation fidelity for a particular domain. Since they are experts within the domain being studied, their experience with the operational environment will be able ensure that no components have been overlooked. This process of narrowing the focus of a fidelity definition to be highly contextual is necessary for researchers to be able to study how fidelity is perceived in a given work environment.

This exercise of defining the components of a work environment that can affect the fidelity of a simulation has rarely been done, and never for enroute ATC. Literature from databases such as the FAA Human Factors library, Sage Publications, Taylor and Francis Online, and IEEE offered no examples of papers that have used simulation in enroute ATC experiments explicitly reporting or discussing the implications of the fidelity of the simulation used, particularly if it is assumed to be a ‘high-fidelity’ simulation. If domain-specific components are specified there is little description or explanation regarding the process of identifying and validating them. For example, Loft et al. (2004) describes an enroute ATC simulation they created for the purposes of research within the ATC domain. They explore their simulation’s usefulness as a research tool but use the vague ‘low, medium and high’ fidelity terms to describe the fidelity of the simulation.

2.3 Examples of Classifying Simulation Environments in the Literature

The section discusses the development and use of categorization systems for simulation environments based on level of fidelity in various domains. Defining simulation fidelity for a particular domain’s work environment is very often not the final step in simulation fidelity work, as the definition does not in itself differentiate between simulation environments. What is needed is a classification or categorization system that clearly delineates between the levels of fidelity of various simulation environments. While the fidelity terms low, medium and high appear to offer a clear delineation from one level of fidelity to the next, too often these terms are not based on a set definition or objective criteria but are subjectively determined based on an individual’s own perceptions. Reliance on a categorization system such as this is insufficient for conducting the type of research that is required to identify connections between simulation fidelity and outcomes in simulation use for training, testing new operational concepts and research on the future ATC operational environment.

Several simulation environment categorization systems were identified in industries such as aviation, driving and marine transportation. The current maritime simulation classification system is provided by DNV GL, the world's leading ship and offshore classification society (DNV GL, 2014a). Their classification system provides the levels of full mission (Level A), multi task (Level B), limited task (Level C), and special task (S) (DNV GL, 2014b). However, the full system they have developed, a 95 page document, is complex and not particularly user friendly. While it is acknowledged that the marine work environment is complex, there is likely a more effective way to communicate these levels (such as the original, condensed version of this categorization system put forth in Drown and Lowry (1993)).

There have been attempts to categorize the fidelity of simulation environments used during driving studies (e.g. Eryilmaz et al, 2014), but this categorization system's uses appear to be primarily academic and do not extend to the driving industry at large. This could be due to the fact that simulation based training is typically not used for training new drivers as they are not performing a task as complex or safety-critical as the operators within the aviation or maritime industry that rely heavily on simulation to replace in-situ training.

Other domains have yet to establish widely accepted, standardized simulation environment categorization systems. For example, the medical domain is beginning to rely more heavily on simulation for the training of health professionals, but there seems to be little standardization with regards to the different levels of fidelity that exist and how they ought to be used (Craighead et al, 2007). Beaubien and Baker (2004) offer a simplified categorization system that provides examples of simulations at varying levels of fidelity (i.e. case studies and role plays, part task trainers, and full mission simulators); however, these represent high level descriptions of a variety of different simulation environments. No attempt is made to better understand how fidelity changes from one environment to the next in a more objective and structured fashion.

One of the clearest examples of a standardized simulation environment categorization system can be found in the domain of flight simulators, where the FAA has created two categorization systems for the various flight simulation environments. The FAA's Full Flight Simulator (FFS) categorization system is presented in Table 2.2 and provides four levels of simulation fidelity with regards to flight simulations that replicate the full experience of flight. The FAA's other categorization system, which they refer to as Flight Training Devices (FTD), differentiates between lower fidelity simulators that are viewed as part task training environments (the distinction being that full flight simulators simulate the motion of flying while the training devices do not).

The descriptions for each level of simulation fidelity attempt to capture in as few words as possible the key elements of the simulation environment being described, as well as the gradual increase in fidelity from one level to the next. There is a clear effort to use general terminology within the descriptions of each level of fidelity which is likely to avoid a particular level being associated with a particular simulator and to make the comparisons as simple as possible.

Table 2.2 The FAA's FFS Categorization System (adapted from FAA, 2013)

Level	Description
FFS Level A	A motion system is required with at least three degrees of freedom. Airplanes only.
FFS Level B	Requires three axis motion and a higher-fidelity aerodynamic model than does Level A. The lowest level of helicopter flight simulator.
FFS Level C	Requires a motion platform with all six degrees of freedom. Also lower transport delay (latency) over levels A & B. The visual system must have an outside-world horizontal field of view of at least 75 degrees for each pilot.
FFS Level D	The highest level of FFS qualification currently available. Requirements are for Level C with additions. The motion platform must have all six degrees of freedom, and the visual system must have an outside-world horizontal field of view of at least 150 degrees, with aCollimated (distant focus) display. Realistic sounds in the cockpit are required, as well as a number of special motion and visual effects.

A key challenge with developing a categorization system such as this is how much detail to provide within the descriptions. Too much detail and it becomes overly complicated and not very user friendly, whereas too little information and the levels become confusing and ambiguous. The FAA's categorization system tends towards the latter, a point made explicitly by an article posted on the Aircraft Owners and Pilots Association website, the world's largest general aviation association, which describes the FAA's categorization system as follows:

> "Simulators are classified by what the industry and the government agree is a confusing system—a trail of good intentions followed by the FAA that left behind an indecipherable classification system (Marsh, 2011)."

While the FAA's simulation environment categorization system may have some flaws, it remains the most prominent example of a categorization system of simulation environments and is therefore likely to be a template for those attempting to create similar systems in other domains. For example, Craighead et al. (2007) used the FAA flight simulation environment categorization system as the basis for their own five-level categorization system of computer-based robots which is presented in Table 2.3. While there is a significant amount of detail for each level of fidelity described in Table 2.3, there is still a certain amount of ambiguity in the descriptions. For example, there is a requirement of simulators of Class C to provide simulation of 'some' forces, control surfaces and effectors. This leaves open to interpretation what 'some' means in this context and could lead to people developing different representations of what that level of simulation actually looks like.

Table 2.3 Robot simulator classification chart adapted from Craighead et al. (adapted from 2007).

Class	Description
Class A	Class A encompasses all simulators that do not meet the requirements for Class B or higher. Class A simulators approximate the motion of a robot, operation of effects and sensors can be assumed. Class A simulators are not required to provide any visual input.
Class B	Class B simulators approximate the motion of a robot and effects, simulation of physical forces and control surfaces is not necessary. Class B simulators run in a basic 3D environment. Sensor simulations must be rough approximations of the real output, high fidelity is not required. Additionally Class B simulators must support all relevant features of Class A simulators.
Class C	Class C simulators simulate some forces, control surfaces, and effectors of a robot so that operation of the simulated device is equivalent to the real device. Sensor simulations, must approximate real output to the fullest ability of the simulation platform. Additionally Class C simulators must support all relevant features of Class B simulators.
Class D	Class D simulators provide simulation of all forces, control surfaces, and effectors of a robot so that operation of the simulated device is equivalent to the real device. Sensor simulations must approximate real output to the fullest ability of the simulation platform. Additionally Class D simulators must support all relevant features of Class C simulators.
Class E	Class E simulators provide full scale mock up of the control unit of the simulated robot. Additionally, Class E simulators must support all relevant features of Class D simulators.

Categorization systems can certainly prove a useful tool, especially in the context of a regulatory body such as the FAA trying to track training standards and requirements, but developing a categorization system that satisfies all of the potential users of such a system remains a challenging endeavour given how much information is trying to be communicated with a simple table or figure. Given the lack of examples to follow and discussion of developmental processes, developing a simulation environment categorization system for enroute ATC simulations will require careful consideration of the strengths and weaknesses of the systems that do exist in order to develop a categorization system for enroute ATC that is as useful and clear as possible.

2.4 Impact of Simulation Fidelity on Training

One of the primary focuses of research on simulation fidelity is in trying to determine what impact different levels of simulation fidelity have on training specific knowledge, skills and/or abilities (KSAs). This requires an understanding of *transfer of training*, which is how "a new skill, or a skill in a new environment, capitalizes on what has been learned before (Wickens and Hollands, 2000)." To determine transfer of training, experimental research is usually required to test different training paradigms against each other to determine which provides the best performance on the actual task. While findings from this type of research would provide the best insight to determine how to use simulation for training controllers, no specific transfer of training research for enroute ATC simulations could be identified based on a review of publications typically used to publish ATC studies (e.g. International Journal of Aviation Psychology, Air Traffic Control Quarterly, Human Factors, etc.). This lack of research could be due to the highly cognitive nature of the ATC task environment. This makes it challenging to measure transfer of KSAs from the training environment to live operations thereby making it less attractive for researchers to pursue given the abstract nature of the subject matter.

A common notion regarding higher fidelity simulation is that transfer of training is maximized when the training environment is as much alike to the work environment as possible (a notion identified by Noble, 2002; Caird, 1996; Hays and Singer, 1989). Hays and Singer (1989) traced this 'common sense' notion back to the *Handbook of Experimental Psychology* from 1951, where a chapter on training by Wolfe stated that essentially the more alike a training situation is to the subsequent working situation, the more positive transfer will occur. Current research shows, however, that lower fidelity simulation can offer just as much if not more in terms of training benefits (Dahlstrom et al., 2009; Beaubien and Baker, 2004; Noble, 2002; Wickens and Hollands, 2000; Salas et al., 1998; Caird, 1996). In fact, Caird (1996) noted that, "…there is some evidence from flight simulation that higher levels of fidelity have little or no effect on skill transfer and reductions in fidelity actually improve training (p. 128; also noted by Wickens and Hollands, 2000)." This finding that deliberate departures from high fidelity can increase the training benefits of a simulation, making it a more effective training tool for students, can be traced as far back as the early 1970s (Blaiwes et al., 1973; Weitz and Abler, 1973; Smode, 1971). The fact that this finding can be traced back decades indicates it is not a new revelation, yet the previously stated 'common sense' notion of reliance on high fidelity simulation seems to persist.

In addition to the added training benefits, a key feature of lower fidelity simulation is the cost effectiveness and availability of these types of simulations (Dahlstrom et al., 2009; Thomas, 2003). While all levels of simulation offer a cost savings to some degree over training conducted in the operational environment, the higher the fidelity of a simulator the more expensive it becomes to not only build, but also to operate and maintain. The size of high-fidelity simulators along with the cost associated with building them typically mean they are not abundantly available for trainees. These factors serve to increase the cost of using high fidelity simulation for organizations as the training tools themselves are scarcer. It was noted by Jackson (1993) that those designing training programs should only use simulation with the fidelity required to achieve transfer of training, but not to overly exceed that threshold as this would result in spending money with no increased training benefit.

There is some evidence to support the use of lower fidelity simulation options for training more generalizable skills. Findings from research on simulation training for general skills and competencies can be generalized given the fact that those types of skills are not domain-specific. One of the best examples of this is illustrated by Dahlstrom et al. (2009), where they used a medium fidelity marine ship simulation to train individuals to handle underspecified problems, essentially problems that are almost impossible to predict (e.g. United Airlines flight 232 in 1989 flying into Sioux City, Iowa). Their question was: how do you train for these events if it is almost impossible to foresee their occurrence? They determined that it is actually the more generalizable skills, such as communication, coordination, problem solving, and team management, that results in successful outcomes in these cases, and that these are perhaps better trained in lower fidelity simulation. They noted that:

> "In spite, or perhaps because, of its lack of fidelity to photorealistic representation and feedback, the engagement and level of intensity of communication, cooperation and decision making observed in groups normally surprise the participants themselves as well as instructors… The lack of [high-fidelity] features leaves groups with no option but to focus on use of general competencies as tools to manage the situations they encounter." (Dahlstrom et al., 2009, p. 310)

This indicates the potential of lower fidelity simulation being used for training these generic competencies and improving a trainee's ability to cope with unforeseen events in the operational environment. It is noted that current training programs in general seem to be lacking this element for training these types of general competencies (Thomas, 2003).

This finding is echoed by Beaubien and Baker (2004) in their analysis of medical simulators and what lower-fidelity simulation options can potentially provide. In fact, they went so far as to state that "when a training programme is properly designed, the level of simulation fidelity becomes somewhat less important (Beaubien and Baker, 2004, p. i51)." This indicates the importance of training program design as a whole rather than simulation fidelity being a sole driving principle. In fact it was noted by Hays that, "It becomes increasingly clear that we cannot productively deal with the concept of fidelity in isolation but rather as a function of total training context--which includes the training tasks, the stage of learning of the trainees, and the instructional system designed to train the tasks. (1980, p. 11)" Simulation is one component of the training paradigm, and even if the appropriate level is used at the right time, that does not guarantee effective training will take place.

While this is helpful advice in a general context, the literature falls short in offering specific conclusions on how different fidelity levels directly impact training. One of the biggest challenges is trying to generalize the conclusions drawn by research on simulation fidelity to different domains. A significant portion of that research has been conducted in the aviation industry analyzing flight simulators (e.g. Hess and Marchesi, 2009; Longridge et al., 2001; Hays et al., 1992). Comparing simulation use in one domain to another is complicated, as operators in different domains rely on different sets of KSAs depending on their particular job. This can preclude the generalization of findings from simulation fidelity research to all domains.

Maran and Glavin (2003) highlighted the differences between simulators in aviation and medicine. In aviation, simulation can accurately replicate the cockpit, the visual field of view, the motion of an aircraft, the aircraft aerodynamics, and the consistency of aircraft (i.e. a particular 747-400 would look and behave similarly to any other). In medicine, however, they work with individual human beings with all their differences and variability. It is much more difficult to create a simulation of humans that is not only accurate but generalizable to large populations. While medical simulation tools can be helpful for training professionals, it is not feasible to use them as qualification tools as the aviation industry is able to do.

The domain of ATC offers similar challenges in terms of comparison with flying, as ATC primarily works with the human element just like the medical domain. Simulating the variability inherent in how all the different aircraft in the sky are flown is a significant challenge to training designers, and the skills required for a controller to be successful are different from those of a pilot. This demonstrates why the domain of ATC cannot simply rely on simulation fidelity work done in the context of flight simulation given the differences in the work environments, and why further ATC-specific simulation fidelity research is required.

Summary While there are some conclusions to be drawn from the current array of literature on the links between simulation fidelity and training, it is clear that work still needs to be done as even these general concepts have yet to be ubiquitously integrated into training programs relying on simulation. Specifically, there is a need to further pursue identifying how specific uses of simulation within the areas of training, testing of new operational constructs or research can be tied to specific levels of simulation fidelity. Given what literature does exist on the subject is tied primarily to other work domains, more focused research on the links between simulation fidelity and simulation use within the domain of enroute ATC could provide significant benefits to simulation users within that industry.

Since there is evidence supporting the idea that fidelity does have an effect on training, there is a need to measure what fidelity is within the domain of ATC. In order to measure fidelity, an objective definition of simulation fidelity is required. This creates the ability to clearly and objectively measure how a simulation changes across the fidelity spectrum and what parts of the operational environment affect the level of fidelity of a simulation, thus making it possible to identify how training a particular skill can be most effectively accomplished in a particular level of simulation fidelity.

2.5 Chapter Summary

The literature provides a foundation in how to approach defining simulation fidelity for ATC simulations. The psychological impact of fidelity that has become a prevalent component in current fidelity considerations was identified as being not directly related to the simulation fidelity concept and was therefore removed from consideration in this work. The literature also offers high-level advice on how to approach simulation use for training purposes, but it is less definitive on mapping the learning of certain skills to specific levels of fidelity. Certain work has indicated that lower fidelity simulation is more appropriate for learning general competencies and higher fidelity is best used at developing highly specified, domain-specific skills. It is clear that more research is required in these areas, especially in the domain of enroute ATC which simulation fidelity research has yet to substantially address, and in order to begin such work, an objective definition of fidelity must first be developed. A standardized definition of simulation fidelity can offer the necessary objective structure in performing comparisons of ATC simulation environments of varying degrees of fidelity. The following chapter outlines and explains the methods used to develop this definition of simulation fidelity for enroute ATC simulations.

Chapter 3

Approach

This chapter details the approach taken to accomplish the objectives outlined in Section *1.3 Research Objectives*. The first section of this chapter presents a general overview of the process used to conduct this research project and provides reasoning regarding the choice of research methods employed to accomplish the objectives. Included here is a sub-section that more specifically outlines the specific methodology used to develop the enroute ATC simulation fidelity definition. Finally, a detailed description of each research methodology used during the course of this research project is provided that explains how each method was used and the scope of the results each method provided to this thesis.

The methods are presented and explained in detail in this chapter rather than paired with their results in subsequent chapters due to the fact that results presented in subsequent chapters draw upon multiple methods at a time, relying on data from various methods and then consolidating into presentable results. Details about how each method contributes to the results presented in the following chapters can be found at the end of the sub-sections in *Section 3.2 Details of Methods Used*.

3.1 Overview of the Approach to the Research Problem

In order to illustrate the overall approach to accomplishing the research objectives, Table 3.1 offers a sequential representation of the high-level research activities conducted in the completion of this thesis.

Table 3.1 Overview and timeline of research activities.

Research Phase	Activities	Contributors			Thesis Objectives			
		Researcher	ANSP	Industry	1	2	3	4
Phase 1	• Initial fidelity definition generated • Initial definition presented for SME review and feedback, definition refined	✓	✓			✓		
Phase 2	• **Semi structured interviews** and **site observation** data gathering activities • Interviewee validation of refined fidelity definition	✓	✓		✓		✓	✓
Phase 3	• Development of categorization system based on interview data and systems from other domains	✓					✓	
Phase 4	• Industry-wide validation exercise via **online survey** • Industry-wide simulation fidelity perceptions data gathering activity via **online survey**			✓	✓			✓

Phase 1 took place during Summer 2013, Phase 2 during Summer and Fall 2013, Phase 3 during Winter 2014, and Phase 4 during Spring and Summer 2014. The "Activities" column in Table 3.1 represents high level summaries of the primary research activities undertaken during the completion of this thesis. The bolded terms within this column are the specific methods used to accomplish the research objectives from Section 1.3, and are presented and discussed in detail later in chapter. The "Contributors" column reflects the contributions from the primary parties involved in the completion of the research activities: the researcher, an ANSP, and a sample of the industry. The final column, "Thesis Objectives", represents how each phase serves to accomplish the four research objectives identified in the introduction to this chapter.

As illustrated in Table 3.1, the first steps to accomplishing these objectives was to build a partnership with an air navigation service provider (ANSP) that created opportunities for hands-on experience with enroute ATC simulation and operational environments, as well as subject-matter expert (SME) feedback. Phase 1 relied heavily on input from the ANSP as SMEs who were part of the project team helped refine the initial iteration of the simulation fidelity definition. Phase 2 then presented this definition to a separate set of SMEs at various facilities within the ANSP during formal **semi structured interviews** in order to provide a preliminary validation of the construct. In addition, the **semi structured interviews** produced a significant amount of data that would influence the structure of follow-on research activities such as the development of the categorization system and provided initial insight into how individuals within the industry view the concept of simulation fidelity. Conducting the research with access to these resources increases the validity of the fidelity constructs and would likely increase industry acceptance of them once they are completed. Phase 3, the development of the categorization system, was influenced by questions posed to interviewees about the appropriateness of a categorization system for enroute ATC, **site observations** of various simulations that were used at the ANSP, and a review of pre-existing simulation environment categorization system in other domains.

In order to extend and broaden the findings from the interviews conducted at the ANSP during Phase 2 and validate the fidelity definition developed during Phase 1 and the categorization system developed during Phase 3 on a larger scale, the approach also included reaching out to the ATC industry at large as illustrated by Phase 4 in Table 3.1. This consisted of an industry-wide validation exercise and data gathering activity that sought to include a larger number of participants then the interviews from a wider variety of backgrounds. These follow-on activities were accomplished by way of an **online survey**.

3.1.1 Specific Approach Used to Define Simulation Fidelity for Enroute ATC

The following sub-section details the specific approach developed and used to define simulation fidelity for enroute ATC, the primary output of this thesis, as illustrated by Phases 1 and 2 in Table 3.1. The results of this approach are presented in *Chapter 6 Enroute ATC Simulation Fidelity Definition*.

Based on the literature reviewed in Section *2.3 Understanding Simulation Fidelity*, an approach was developed that enables the creation of an objective definition of simulation fidelity for the enroute ATC domain. This approach consists of first determining the scope of the fidelity definition, then identifying the specific parts or components of the operational environment that fall under the scope of the definition, and finally validating those components via SME feedback. This methodology was formed based on several sources, but two of the key contributors were Hays and Singer (1989) and Estock et al (2006). Once completed, this process would yield an objective, industry-validated definition of what influences simulation fidelity in the domain of enroute ATC.

The initial step was determining the scope of the fidelity definition that would be developed. The word scope is used here to describe what is being taken into consideration in terms of the different elements that can affect the fidelity of a simulation. As was discussed in Section 2.2, it was determined that excluding the psychological considerations from the definition, as Hays and Singer (1989) did but appears in some current fidelity work (e.g. van Merrienboer, 2013; Estock et al, 2006; Alexander et al, 2005), was important in developing an objective, comprehensive definition of fidelity that captured the relevant parts of the operational environment. This means excluding user-oriented factors, such as user acceptance and simulation immersiveness, that contribute ambiguity and confusion to the concept of simulation fidelity. For a more thorough explanation of why these factors were eliminated, please refer back to Section 2.2. Therefore only elements of the operational environment that could be directly manipulated or controlled by simulation designers were to be considered in the creation of a simulation fidelity definition.

After establishing the scope of the definition, it is possible to begin identifying the components of the operational environment that can affect the fidelity of a simulation. A qualitative approach was followed in order to accomplish this, an approach inspired by Estock at al.'s (2006) work defining fidelity for an F-16 flight simulator. Their approach consisted of identifying specific environmental components that could be grouped under broad themes or categories via observation and SME feedback. For this thesis, the process undertaken to define fidelity is represented in Phase 1 in Figure 3.1 by the initial development of the definition and its preliminary review by SMEs. This preliminary review of the definition by SMEs included a high-level training program manager, learning quality specialist and a training program specialist with operational experience. The feedback that was provided during the meeting was used to refine the definition into the version that was presented for validation during the semi structured interview sessions.

Returning to the structure of the definition, a category is meant to capture the general theme or topic of the components which fall under its scope, whereas a component is much more specific in terms of a specific part of the operational environment being captured. An example of a high level category with a more specific underlying component would be "Shape and layout of Cockpit Controls and Displays" underneath the main category of "Cockpit" (Estock et al, 2006). An example from the definition developed for enroute ATC within this thesis is the component of "Control Interfaces" within the category of "Physical Environment".

The final step is then validating these components by receiving critical feedback from SMEs. Given that they are most familiar with the operational environment the simulation is attempting to replicate, they are in a strong position to determine whether or not a component truly does affect the fidelity of a simulation or should be excluded from consideration (Estock et al, 2006; Hays & Singer, 1989). In Figure 3.1, this is accomplished in Phase 2 and 4 by first validating the definition within the ANSP and then undergoing a larger-scale validation through the survey.

To summarize, the methodology for defining simulation fidelity for enroute ATC consists of first determining the scope of that definition, then identifying the relevant components of the operational environment that can affect the fidelity of a simulation of that environment, and finally validating these components via SME review. This methodology, consistent with successful attempts in defining fidelity in the flight simulation domain, represents a strong and comprehensive approach for developing an objective definition of simulation fidelity for enroute ATC.

A similar approach was used in order to develop the simulation environment categorization system. Observations of simulations and SME input regarding the structure of a categorization system were used to generate an initial simulation environment categorization system, which was then presented to a broader pool of SMEs via the online survey in order to validate the simulation environment categorization system.

3.2 Details of Methods Used

The following three sub-sections provide detailed explanations of the three research methods which contributed to this thesis: semi-structured interviews, site observations, and a simulation fidelity survey. The approach to the literature review was detailed in section 2.1 and will not be covered again here. Table 3.1 summarizes which methods were used to accomplish each of the research objectives of this thesis.

Table 3.2 Mapping of research methods to research objectives.

Objective	Methods
1 – Identify how people within the domain of ATC currently make simulation fidelity determinations.	• Semi structured interviews • Site observations • Literature review • Survey
2 – Develop an objective enroute ATC simulation fidelity definition.	• Semi structured interviews • Literature review
3 – Develop an enroute ATC categorization system.	• Semi structured interviews • Literature review
4 – Validate the enroute ATC simulation fidelity definition and categorization system within a diverse sample of the industry.	• Survey • Semi structured interviews

3.2.1 Semi Structured Interviews

Interviews were conducted over the course of two site visits to two separate ATC facilities at the ANSP that were spaced two weeks apart during the Fall of 2013. A semi-structured interview format was used. The following definition of semi-structured interviews is used as the template for the interviews conducted during this research project: “In a semi structured interview, the researcher has a general plan for the topic to be discussed but does not follow a fixed order of questions or word these questions in a specific way… The interview is generally audio-recorded and transcribed, and… analysis typically involves comparison, coding, and summarization (Packer, 2011).”

The prepared questions covered the following topic areas: progress through the training program, simulation capabilities, simulation selection, testing/training for new tools and procedures, on-the-job training (OJT), simulation scenarios, fidelity and the researchers’ proposed fidelity definition. There were 13 interview participants consisting of training managers (2), en-route program specialists (2), learning quality specialists (1), operational training specialists (4), simulation specialists (2) and temporary duty instructors (2). The full set of prepared questions that were used during the interview sessions can be found in Appendix A. For 7 of the interviews there were two researchers present, whereas the other 6 where conducted by just one researcher.

While there were pre-arranged questions, the researchers used their discretion as to which questions were most suitable to be asked during each interview depending on the background of the participant. Since it was a semi-structured interview, lines of questioning that do not appear as part of the pre-arranged questions were also pursued when topics of interest arose naturally through the process. Data collection consisted of audio recordings of each interview that were supplemented with notes taken by the primary researcher during the interview. There was over 17 hours of audio recorded across the 13 interview sessions. The recordings of the interviews were then transcribed.

The interview transcripts were analyzed by searching for major themes that over-lapped between participant responses. Where appropriate and possible, responses were coded and analyzed by calculating response frequencies across the interviewees. This was done for questions where interviewees were asked to list items in their responses, thus providing the potential for response patterns to emerge across multiple interviewees.

The term coded is used here to describe the act of abstraction and generalization of the interview data (Packer, 2011). This process is best described as separating out specific responses or parts of responses that become the units of analysis, identifying the high-level category that these units fall under and labelling them as such, and finally using these categories to describe similar responses from other interviewees. No pre-existing coding scheme was used, meaning the codes were being generated based on the responses provided by interviewees. The coding was performed by a single researcher.

As an example of this, the question, "In your opinion, what skills or parts of the job are most difficult for students to learn?" yielded responses from interviewees where they then listed all of the skills and/or parts of the job that they felt were most difficult for ATC students to learn. The data from this question is presented in Table 8.2 in Section 8.1. A sample response from one interviewee provided the statement "Priorities, difficult because it changes all the time" which was subsequently coded as *Prioritization*, while another interviewee provided the response "Take a rule or regulation that you are familiar with and apply it somewhere else, it's still the same rule it's just a completely different application or area" which was coded as *Adapting knowledge to new situations*. Other examples of the results from these types of question are presented in Tables 4.5, 4.6, and 5.3.

The results from these interviews are used as the primary foundational material for the majority of *Chapter 4 Understanding Simulation Fidelity in Enroute ATC Based on Simulation Use for Controller Training*, but also made important contributions to Sections 5.1 and 5.2, the initial validation of the simulation fidelity definition detailed in Section 6.3.1, and as part of the development of the simulation environment categorization system presented in Section 7.1.

3.2.2 Site Observations

Site observations were conducted on three separate occasions at three separate facilities of the ANSP during the Summer and Fall of 2013. The observation sessions consisted of observing the following three different activities: a four day early-phase training course building and testing session for a PC-based simulator, two separate 30 minute sessions using their workstation simulation that consisted of validating a training scenario and providing refresher training to a controller, as well as two 30 minute sessions observing work habits and listening to live air traffic control in the operational environment. Data was gathered via researchers taking notes during each of these activities.

These site observations took place within the first few months of the research project commencing. While the findings from these observation sessions do not explicitly appear in this thesis, these were important activities to observe as they helped provide context to what was being reported during the interviews. This allowed for greater depth of analysis when drawing conclusions about the findings from the interviews by providing first-person experience with the simulations that were being discussed by the interviewees as well as the operational environment that was being simulated.

3.2.3 Simulation Fidelity Survey

The final research activity was developing and distributing a simulation fidelity survey to the enroute ATC industry. The survey was developed using the online survey website FluidSurveys, a Canadian based company. The survey was used to validate the two simulation fidelity constructs and to gather industry perceptions on a wide range of issues related to simulation fidelity in the ATC domain. While there was a preliminary validation of the simulation fidelity definition performed during interviews at the ANSP, it was felt that feedback from a wider range of SMEs was required to accurately determine how applicable the definition would be across the global industry of ATC.

Survey Design Survey questions consisted of a mix of Yes/No, Likert scale ratings, and short and long answer questions. Topics covered in the survey include participant perceptions regarding the concept of simulation fidelity in the domain of air traffic control, what level of simulation fidelity is required to train for a certain skill or test/evaluate a particular concept, and acceptability and accuracy of the simulation fidelity definition and categorization system. A printout version of the survey is available in Appendix B.

Survey Distribution The survey was first distributed to personal contacts within various ANSPs and researchers around the world who met the participant criteria of the survey. This was done to try to ensure that the survey participants were coming from as reliable a source as possible. The target population was anyone who had experience developing or using ATC simulations, which included the following examples of potential participants:

- Active air traffic controllers who have used simulation for training / participated in simulation studies
- Controller training designers / developers
- Air traffic control instructors
- Researchers who have used simulators for human-in-the-loop studies
- Operational concept developers and controller tool developers who have used the results of simulation studies

The personal contact was provided with a brief description of the study's purpose and asked if they would participate. At the end of the email a request was made regarding forwarding the survey information to anyone in their network who fit the target population and who they felt would participate in the survey.

Approximately 4 weeks after exclusively personal contact recruitment of survey participants and having received 58 completed responses via this approach, the survey was then made publicly available on aviation public domain websites (e.g. liveatc.net, pprune.org) and through air traffic control publications (e.g. ATC Network and Air Traffic Management) where the target population for this research typically frequent. According to a demographic question that asked the survey participant how they heard about the survey, only 6 of the 86 completed responses were submitted by those recruited from the public domain advertisements. Another safety measure used to ensure the quality of survey participants during the personal contact phase of distribution were the security settings provided by FluidSurveys. These security features meant the survey would not show up in any search results on a search engine, ensuring that the only way to complete the survey during the personal contact distribution phase would be to click on the link provided in the email. The survey was available for a total of approximately 10 weeks.

Survey Demographics The total number of completed surveys was 86. There are certain demographic results that increase confidence in the conclusions being drawn from the survey data. For instance, 60% of all respondents had over 10 years of experience working with enroute ATC simulations. In addition, 60% of all respondents indicated they had operational experience. In terms of survey participants' areas of professional experience, 65% had experience with terminal operations, 78% had experience with low enroute operations and 74% had experience with high enroute operations, indicating a relatively even distribution amongst the three types of enroute ATC[1]. Finally, survey participants were predominantly North American with 40% coming from the United States and 35% from Canada, but with a significant contribution from the International community at 25%. The full results from the demographic questions of the survey are presented in Appendix C.

Survey Analysis Where open questions were used to gather further feedback, the responses were coded and analyzed using response frequencies. Certain responses could have more than one code attached to it depending on the topics the survey participant discussed. An example of this is when survey participants were asked the question, "In your opinion, what are 3 key differences between the highest-fidelity simulation environment you have worked with and the enroute ATC work environment?" and were then provided three open answer text boxes to provide their responses. Some sample responses from this question include "In a human piloted simulation, the fact that all the aircraft pilots have the same voice" which was coded as "Communications", or "Unusual situations are hard to simulate such as in flight emergencies or pilot requests" which was coded as "Operational uncertainty". Data from this question is presented in Table 4.7.

In certain figures or tables, there were response frequencies reported that do not match the total number of completed responses for that question. There are two possible reasons for this: (1) due to ethical considerations, participants were not compelled to answer every question and therefore left that particular question blank, or (2) certain responses were excluded by the researcher as their responses indicated they did not understand the question

---

[1] Each figure reported in this sentence is a percentage of total respondents as survey participants were able to select multiple answers for this particular question.

and therefore provided an answer which was not pertinent to the question. For the latter situation, this was only done where long answer responses were provided and it was clear to the researcher that the participant did not understand the purpose of the question creating reasonable doubt regarding the validity of their response. An example of this is in Figure 5.1 where 63% of responses are in the Yes/No responses, 27% were coded as Non-pertinent, and 9% offered No explanation. The question asked to participants was "Do you believe that simulation fidelity is a well-defined concept in the ATC domain?" where the participant would respond either yes or no and then proceed to explain their answer. Two examples of those who were judged to have clearly understood the question are provided below:

- "I think it's defined and conceived just fine, but, in my opinion, it's not implemented very well."
- "I believe simulation fidelity means different things to different people. I believe current controllers are not involved enough in validating the fidelity of a simulation before it is used in the field."

Three examples of responses from those who it is believed did not understand the question:

- "I have been working in ATC for 23 years, and simulation has been in use all of this time."
- "Simulation is a training tool. You can't simulate what experience teaches you."
- "Today's sim environment is much improved, and keeps getting better as a function of computing power and continuous learning."

It was felt that including responses such as the latter examples would unnecessarily cloud the results making it more difficult to develop insights into direct responses to the question asked. Therefore these responses were coded as Non-pertinent and are included in figures, but are not considered when drawing conclusions about the data presented in the figure. Anywhere this was done, there is an explicit note stating that this was done and then a cross-reference provided to this section.

Based on the categorical nature of the data, the chi square goodness of fit statistical test was identified as the most appropriate statistic to determine any differences between the observed results and a potentially non-significant result (Howell, 2013). Subsequent chi squared statistical analyses were performed where demographical information was available to compare the response rates of different demographic sub-groups. The results from the surveys appear primarily in *Chapter 5 Industry Perceptions Regarding Simulation Fidelity in ATC*, as well as the primary validation efforts for both fidelity constructs as covered in Sections 6.3.2 and 7.3.1.

3.3 Chapter Summary

This chapter outlines the overall process taken to accomplish the research objectives of this thesis, along with detailing the specific methodologies used during that process to gather data. This approach relied on tapping into the resources provided by an ANSP, as well as the industry at large, in order to better understand the different simulation tools that exist and how they are used based on their simulation fidelity. In addition this allowed for the creation of a robust definition of simulation fidelity for enroute ATC given the input from industry experts throughout the process. The following chapters present the results of the methods described in the present chapter.

Chapter 4

Understanding Simulation Fidelity in Enroute ATC Based on Simulation Use for Controller Training

The first objective of this research project was to identify how individuals within the domain of enroute ATC currently make fidelity determinations regarding various simulations. To achieve this objective, this chapter identifies the different types of simulations that are primarily relied upon at an ANSP, what they are used for and their limitations and differences from the operational environment. Analyzing these different aspects of enroute ATC simulation and how they are used highlights some of the key areas where individuals are making fidelity judgments and provides preliminary identification of some of the environmental components that affect simulation fidelity in enroute ATC.

4.1 Simulation Capabilities for Enroute Controller Training at an ANSP

The following three sub-sections present findings regarding the different uses, strengths and weaknesses of each type of simulation that the ANSP employs in its training programs based on the site observations and interviews conducted at the ANSP as described in Chapter 3.

The following definition, as previously stated in Section 1.2, was used to understand what constitutes a simulation of enroute ATC within the context of this thesis: any situation where an individual actively practices providing some part or all of the air traffic services provided by controllers in the operational environment. Typically this is accomplished using some form and/or combination of tools, objects, or personnel to replicate some part of the real world task environment. Examples of different types of simulation within this context include an instructor moving plastic airplanes around on a tabletop sector, a personal computer part-task program, and a complete reproduction of a radar controller's workstation and work environment.

The structure of each sub-section consists of first introducing the simulation by describing its components and providing examples of similar simulations used in other contexts. Then a description of how the simulation is used within the ANSP is provided based on comments made during the interviews. Finally, a brief discussion of the simulation's strengths and weaknesses is provided. This forms a clear picture of the qualities of these simulations, how they differ from each other, and what their typical uses are.

4.1.1 Classroom-based Simulation

This specific form of simulation can best be described as role-playing situations or case studies. Examples of this type of simulation include drawing a scenario on a whiteboard and working through a conflict conversationally with trainees or using static pictures of radar screens to test a controller's ability to recognize potential conflicts. This type of simulation requires few resources to operate and possesses an inherent flexibility in running through scenarios and situations given the lack of resources required to use this simulation. The comments in Table 4.1, taken from interviews at the ANSP, provide insight into how SMEs believe this simulation should be used and what its strengths are.

Table 4.1 Interviewee comments regarding classroom-based simulation.

Comments
• "Good for problem solving, conflict solving. Really the theory behind what we are looking at when solving conflicts." • "Good for visualization, understanding and application of the rules. Without the radarscope you do have to visualize where the AC are in their tracks. Also cost effective." • "It's convenient, don't need to open up the simulator, it's kind of like simulating the simulation." • "Really helps with the analysis, what to look for, what are the most important factors that will affect your solution, for preparing alternate plans, and to give them confidence on their knowledge." • "It takes advantage of resources and efficiency, it's that bridge before you plug in. We've talked about it, you know it in theory and have been given examples, … more opportunity for coaching in there."

Based on these comments, interviewees believe that classroom-based simulation's current role in training involves developing the foundational skills of ATC in a simplified setting, and provides the capability for analyzing the more complex skills and situations in greater detail after they are presented in higher fidelity simulation. Regarding the latter role, it is clearly believed to be an important technique for making the transition from discussing theory in the classroom to using higher fidelity simulation, as is made evident by comments such as, "… it's the bridge before you plug in" or "Really helps with the analysis, what to look for, what are the most important factors that will affect your solution…" Interviewees believed it helps to prepare the trainees in terms of knowledge required to complete the exercises in the higher fidelity simulation, what to expect in those simulations, and to develop confidence in the trainees' abilities. It is also used for debriefing after exercises performed in higher fidelity simulation for more in-depth analysis of the more complex scenarios encountered in higher fidelity simulation and a deeper understanding of a controller's thought process.

A key strength of this type of simulation highlighted by interviewees is the convenience and cost effectiveness it provides to those using it. Scenario design is much simpler, consisting of verbally presenting a situation to trainees, handing out a static picture of a radar screen, or simply drawing the problem on a white board. The time-consuming and expensive process of designing scenarios associated with the higher fidelity simulation is reduced significantly with classroom-based simulation. This type of simulation also eliminates the potential for trainees to use the automated system tools as they are not available without a computer; they must develop a controller's thought process to solve conflicts in this type of simulation rather than rely on automation to guide them which is believed to be an important developmental step for trainees.

It is clear, however, that interviewees feel this simulation should be used in conjunction with the higher fidelity simulation. It cannot function on its own as it lacks the dynamic nature that is inherently part of ATC. It also lacks the system tools which trainees must eventually become highly familiar with as the automated systems are very much integrated into how controllers perform their job in the modern ATC operational environment.

Based on these findings, interviewees feel classroom-based simulation is most useful for going more in-depth into analyzing the controller's thought process and providing more opportunity to develop automaticity in some of the base foundational skills that experienced controllers rely on. It is also a significantly more cost-effective and convenient tool to use than the dynamic simulation tools, yet it is believed by interviewees to be most effective when paired with simulation of a higher level of fidelity.

Given interviewees' views on this simulation's limitations, it is clear interviewees are making fidelity judgments for this simulation based on its lack of equipment and timing realism.

This provides two potentially important environmental components that can affect the fidelity of a simulation if interviewees believe they are the main reasons why classroom-based simulation ought to be considered lower-fidelity than the other environments used at the ANSP.

4.1.2 PC-based Simulation

The second type of simulation used by the ANSP is a PC-based simulation. This type of simulation involves replicating a radar screen on a computer monitor where the user is presented with a particular airspace and aircraft moving in a realistic fashion through this airspace. These aircraft can be actively controlled by the user or operator of the simulation by using voice recognition software or a text-based control input. This type of simulation is particularly popular in many research studies that use ATC as a domain for analysis given it offers a dynamic experimental environment while still remaining relatively simple to operate (e.g. Weber, Oberheid and Papenfuss, 2013; Loft, Finnerty and Remington, 2011; Jha et al, 2011; Sethumadhavan, 2009).

At the ANSP, the PC-based simulator is primarily used during the early-phase training in order to build and master the fundamental, basic skills required to perform ATC. Table 4.2 provides interviewee comments that illustrate this point. Some of the skills that interviewees stated that they look for trainees to demonstrate include: phraseology, managing simple and isolated tasks, demonstrating some proficiency in handling multi-tasking but at a low level of complexity, prioritization, and basic concepts of ATC (i.e. issuing a clearance or radar identifying an aircraft). The PC-based simulator, while interviewees acknowledged its limitations in terms of fidelity and functionality, is well suited to building these basic skills.

Table 4.2 Interviewee comments regarding the use of a PC-based simulation for building basic ATC skills.

Comments
• "It's very effective for basic skills because it's a simulated … environment where you have situations that help them recognize and understand the basic concepts/skills of ATC." • "Basic introduction to the skills they are going to need, the fidelity of it is there are some basic things that you can make a pretty good determination on that simulator as to whether someone has the potential to move forward or not."

As is demonstrated in interviewee comments provided in Table 4.3, a key strength of this type of simulation is its ability to provide self-practice opportunities to trainees. This can be an important feature for trainees, as they are allowed the opportunity to become more familiar with the job and the dynamics of the system they will eventually be controlling. Interviewees noted that there is the potential for bad habits to form during their self-practice as it can be considered unstructured learning time, but they also noted that with the appropriate in-class support and ensuring only appropriate scenarios are available to trainees at the right time, the potential for bad habits to form is reduced.

Table 4.3 Interviewee comments regarding the PC-based simulation's ability for self-practice.

Comments
• "One of the biggest benefits to the students is they can go play on it on their own." • "Main advantage of desktop sim is they can practice on their own time. Disadvantage is that the fidelity is quite limited." • "Students don't get as much time to practice and master skills in higher levels of fidelity of simulation, and self-practice provides a lot of potential for students to get better."

The voice recognition software is what creates this self-practice ability as there is no need for instructors or simulation specialists to play the role of pilots. Voice recognition software acts as the pilots that controllers coordinate with in their airspace, allowing the user of the simulation to issue commands via a headset connected to the computer with the software understanding and implementing the actions requested by the user for the appropriate aircraft. This also offers cost savings in terms of running the simulation, either during class time or by trainees on their own. There are still limitations with the voice recognition technology, as its actions are restricted by the programmed responses available and some interviewees noted it can be temperamental.

Based on these findings, it is clear that interviewees believe that the PC-based simulator is most suited to developing the basic skills required to perform ATC in the real world. In comparing with the classroom-based simulation, there are clear differences in the fidelity of the physical environment (operated on a personal computer or laptop in a classroom vs. static scenario), communications (uses voice recognition software or text based data input vs. real people) and overall functionality of the simulation (provides a basic radar display with a few of the functions available in the real world operating system vs. no functionality). The differences illustrated here highlight further dimensions which can affect the fidelity of a simulation in enroute ATC (e.g. communications). This being said, interviewees believe it is at the appropriate level of fidelity for developing those basic ATC skills in early-phase trainees. It also provides the opportunity of self-practice to trainees, not requiring the same amount of resources to operate as a higher fidelity simulation.

4.1.3 Workstation Simulation

The final simulator used by the ANSP is a workstation simulation. This simulator replicates many of the aspects of the operational environment with the physical components such as the flight strip organization panel, communications touch screen, weather information screen, and airspace map all present. The communications are executed via simulated radio communications with a single simulation specialist in a separate room acting as all the pilots and other controllers in the scenario. The workstation simulator is the highest level of simulation fidelity available at the ANSP. At this point the airspace being replicated is almost always a real airspace, and the aircraft performance modelling is at its highest. It is the simulator used during the final phase of training before a trainee transitions to on-the-job training in the operational environment. This type of simulator is also used to conduct research by research organizations building their own workstation simulation (i.e. NASA's Airspace Operations Laboratory) or via a collaboration between the research community and an ANSP (i.e. Hannah and Neal, 2014 working with Airservices Australia).

The main strength of the workstation simulation that can be identified from interviewees' comments regarding the workstation simulator, provided in Table 4.4, is that it provides the opportunity to develop a high-degree of familiarity with the sector or sub-unit a trainee will eventually end up working. Trainees learn about the traffic flows and patterns, how the structure of their airspace affects those patterns, and develop a comfort level with the operating system used in the operational environment. This simulation also allows for continued refinement of the basic skills first learned using the PC-based simulator, along with certain skills or knowledge that tend to be inherent to specific sectors. The workstation simulation is also physically consistent with the operational environment, meaning new trainees begin to develop a sense of comfort with the human-machine interface that is present in the operational environment and ensures that active controllers have little to no learning curve when doing recurrent training in the workstation simulator.

Table 4.4 Interviewee comments regarding the workstation simulation.

Comments
• "It's very customized and very current, much closer to what occurs on the floor. Every specialty has a very specific pattern of traffic, very predictable and repetitive. The [workstation] sim helps develop that familiarity."
• "Going really specific for things to learn, we have more control because we are designing the simulations specifically for that sub-unit. Very aware of what the students have seen and what they need to deal with now."

Users of the workstation simulator communicate with a real person, a key difference from the PC-based simulator, as the simulation scenarios in this simulator are operated by simulation specialists. This begins to include a human element in the overall system dynamic which is not present in the PC-based simulator. While this addition potentially increases the fidelity of the simulation, it also increases the operating costs. This also serves to identify a key environmental component that can affect the fidelity of a simulation. The human component, both in communications and the unpredictability of operations, can significantly affect the realism of a simulation and ought to be considered when developing a definition of simulation fidelity.

Based on these findings, the workstation simulation is used to refine the basic skills learned in simulations with lower levels of fidelity as well as to become familiar with the characteristics of the sector a trainee will eventually be working. The physical characteristics of this simulation begin to closely resemble those of a controller's actual workstation, and the user of the simulation is interacting with a real person instead of an automated voice recognition system. Investigation into the limitations of the workstation simulation is deferred to the next section.

4.1.4 Summary

Within section *4.1 Simulation Capabilities for Enroute Controller Training at an ANSP*, three simulations with distinctly different levels of fidelity were described and discussed in terms of how they are used throughout the training process at this particular ANSP. Classroom-based simulation, the PC-based simulator and the workstation simulation offer three examples of simulations of varying degrees of fidelity in the domain of ATC, as simply by including or removing certain components of the operational environment several different simulations become available. It also highlights what those who use these types of simulation on a regular basis believe they are best used to accomplish in terms of training objectives. The next step in this process is to identify the limitations and differences of the workstations simulation as compared to the operational environment. This activity sought to elicit specific environmental components from interviewees that have a significant impact on the realism of a simulation and therefore are important when considering the fidelity of the simulation.

4.2 Comparing the Operational Environment and Simulation

To further understand the differences between the operational environment and simulation beyond the discussion initiated in the previous section, questions focusing directly on comparing the two environments were included during the interviews at the ANSP as described in Section 3.2.1. More explicitly, these questions focused on identifying the key limitations of the workstation simulation used at the ANSP, comparing how the workstation simulation and the operational environment differ from each other both at the ANSP and across the industry, and finally identifying additional limitations of the workstation simulation based on the dynamics of the transition by trainees from simulation to live operations at the ANSP. In addition to the questions posed to interviewees, a follow-on question was included in the online survey, as described in Section 3.2.3, to explore industry-wide opinions regarding the key differences between high fidelity enroute ATC simulation and live operations. This section serves to further identify the particular components of the operational environment that people consider when making fidelity determinations. The workstation simulation is specified for this analysis as it is the final simulation the trainees work with before heading to the operational environment, and because it is the most realistic simulation that the ANSP uses.

4.2.1 Limitations of a Workstation Simulation

This sub-section presents the main limitations of the workstation simulation used at the ANSP. Identifying the limitations of this particular simulation will provide several key components individuals are considering when making fidelity determinations. It will also serve to identify some of the more challenging aspects of the enroute ATC operational environment to simulate with high degree of realism.

Interviewees were asked, "What are the main limitations of the simulation used at your ANSP?" The interview data generated from this question was analyzed and then synthesized into Table 4.5 using the methods described in section 3.2.1. The first column represents the keyword or main theme as found by the researchers, the second column represents the percentage of participants who included or mentioned this as one of their simulator limitations, and the final column offers some sample quotes from the interviews that correspond to the topic of that row.

Table 4.5 Limitations of the workstation simulation used for training at the ANSP.

Simulation Limitations	Percentage (N=13)	Sample Quotes
Aircraft performance	77%	• "You hear things like 'Well that aircraft would never have climbed that fast in the real world' very often, you try to avoid the development of false expectations of how aircraft behave." • "Aircraft performance is another big limitation because you do learn how fast that guy is going to slow down and if what you see in the simulator is different, that's an issue." • "Can lead to bad habits, they end up playing the simulator and know how the machine works. They get surprised on the floor and get burned trying to do things they got away with in the sim."
Single person driving sim	46%	• "One person driving the simulation cannot be the equivalent of 6 different pilots all wanting your attention at one time." • "The skilled student will quickly learn that if they drive the pace of the exercise and control the calls then requests can't come."
Equipment/ software lag (not present or lacks same functionality)	38%	• "Can present challenges especially if there are instructors who are used to having those tools and they come up to teach and they are not available." • "Not the top notch equipment that you see on the floor, doesn't quite work to that level." • "Simulation isn't far behind, but there is a gap there." • "There are certainly some tools though that are on the floor and used every day that we don't have, which is a bit of a pain. Need to work around that."
Unexpected events/anomalies	15%	• "When it's real, anything can happen technically. When it's simulated, it's managed and controlled."
Weather/ turbulence	8%	• "Winds remain overly constant in spots, same with turbulence."

There were three strongly agreed upon main limitations with each receiving over 35% mention rates: aircraft performance, a single person driving the simulation, and equipment or software that was not as up to date as the floor.

Aircraft Performance Characteristics Aircraft performance was a highly-mentioned limitation, being brought up by 77% of the participants interviewed. Interviewees highlighted the effects it has on expectations that are developed within trainees.

One of the key issues in terms of aircraft performance characteristics noted by interviewees was the predictability of the aircraft, as aircraft always behaving a certain way is not congruent with the dynamics of the real world. This has a direct impact on how trainees learn to apply their skills, as they eventually learn to play the simulator rather than learn to control. "In the sim, you can perform an action and close your eyes and know it will be fine. In the real world, you can't because you just don't know what will happen."

Several interviewees believed that the predictability was helpful early in a trainee's development as it made it easier for them to pick up the skills, but the closer they got to switching to live operations the more that predictability became an issue. One participant stressed that the variability inherent to live operations should:

> "… all be [learnt] on the floor. Simulation needs to be built on normal operations. Trainees should learn the skills on how things should work in an idealistic environment before introducing any anomalies or complex situations."

In fact, the consistency in how the aircraft perform is what allows the scenario designers to produce the conflicts when and where they want them. This, however, doesn't stop instructors and trainers from pointing out to trainees the lack of realism in how the aircraft behave. Certain interviewees noted that they would or have seen instructors explain to trainees the differences in aircraft and airline behaviour in the simulation as compared to the real world.

Participants stated that aircraft performance is directly tied to how controllers perform certain skills, such as putting aircraft in trail, and that if the performance characteristics were off or were overly predictable, the trainees would not be able to develop those skills to an appropriate level. This is one of the main reasons why accurate aircraft performance characteristics are so highly desired. "[Some] seem to think we want the aircraft performance package just so we can replicate reality, this isn't the case. If we don't have better performance, we won't be able to target specific skills in our training."

Limited Communication Realism The limitation of having a single person driving the simulation runs can be tied into the concepts of multi-tasking and dealing with live communications. It was repeatedly pointed out by interviewees that in the real environment, quite often controllers have multiple people talking to them at once with different demands. This element of multi-tasking and then having to prioritize the demands of those you are dealing with is something that is hard to simulate given that it is only one person on the other side of the scenario controlling all the aircraft. It also means that a trainee is not able to acclimatize to this type of multi-actor communication until they get to on-the-job training (OJT), where they have a significant adaptation period they have to go through to handle that difference. However there is obviously a cost issue involved with this as it is not feasible to pay four or five more people to be involved in the simulation run for one trainee.

Equipment/Software Lag The lag refers to the difference in the equipment or software that the simulation uses versus those used in the operational environment. It is not uncommon for the operational environment to receive the latest equipment and software yet the training environment has to wait to receive similar upgrades. The average lag estimated by one interviewee was approximately 8-12 months. It was regarded by interviewees as both a challenge to overcome but also an accepted practice of the training paradigm given the costs associated with much of the new technology being continuously added to the operational environment. One participant noted that it “can be hard to expect the trainee to pick up on these if the equipment is not the same, having to explain to [the trainees] that what they are seeing now will be different on the floor.” If a trainee is successful in the simulation, but are being told that what they see now is not exactly what they will be doing with the equipment and software present on the floor, it could cause some uncertainty to enter a trainees mind regarding the usefulness of what they are learning in the simulation.

This is where interviewees believe instructor confidence in the simulation tools is an important dynamic, as it then provides the trainee with confidence that even though things are different from the operational environment, it is still a valuable learning tool. This can be difficult for instructors as they are typically active controllers who would like to teach trainees the way things are done in the operational environment, but are restricted by the equipment present in the simulation. It was even noted that the training designers could not create certain training scenarios or situations due to the lagging equipment, software or procedures.

While interviewees acknowledged that the equipment/software lag was ‘the nature of the beast’, it was noted by one interviewee that many pieces of equipment and software in the operational environment have become more than just an extra tool. They have become integrated into how a controller does his/her job; it is part of doing ATC now. It is these types of tools that interviewees were most frustrated at not having available as it can produce adaptation issues when transitioning from the simulation to the operational environment.

Summary The limitations identified here indicate some of the gaps in fidelity between the operational environment and simulation. These limitations directly affect how these simulations can be relied on to provide training, as at this point they can carry a trainee only so far and a significant amount of time is still required in the operational environment to complete their training. To further explore the gap in fidelity between simulation and live operations, questions were also asked to identify key differences between the workstation simulation and the operational environment.

4.2.2 Key Differences between Operational Environment and Simulation

In trying to identify the key differences between the workstation simulation and the operational environment, questions such as, “Are there any cases you can think of where simulation is not able to reproduce critical aspects of the real world environment?”, “In your opinion, why is OJT required after high-fidelity simulation?”, and “What are the challenges experienced in OJT that are not experienced in simulation exercises?” were posed to interviewees. Responses to these questions would provide insight regarding how interviewees believe activities and training conducted in the real world differ from those conducted in the highest fidelity simulation in use at the ANSP, the workstation simulation; responses to these questions could then be abstracted to identify key differences between the operational environment and simulation. There are a lower number of overall responses to these questions as only those 10 interviewees familiar with the operational environment were asked these questions. Table 4.6 summarizes the responses to these questions.

There are five differences that 40% or more of the interviewees identified. These differences are: “Dealing with real people”, the “Unpredictability” of the real world, the “Stress/nerves” that come with working in the operational environment, “Large-system dynamic”, and “Simulation is trainee-centric”. The results identify several aspects of the ATC environment, such as “Dealing with real people” and “Unpredictability”, that are typically not included in simulation at any level of fidelity at the ANSP.

Table 4.6 Key differences between the operational environment and workstation simulation as indicated by interviewees.

Differences	Percentage (N=10)	Sample Quotes
Dealing with real people (personalities, language issues, etc.)	70%	• "When you're controlling your dealing with many different people, don't have the ability to recreate that in the sim because it's usually 1 or two people driving the run." • "… dealing with the people around them and the communication. Some [trainees] just go with the flow and are okay but others are intimidated because it's happening fast and how to the point other controllers are."
Unpredictability	60%	• "When it's real, anything can happen (technically). When it's simulated, it's managed and controlled." • "The unexpected things, simulation is just not realistic enough to handle those things consistently."
Stress/Nerves	50%	• "When they hit the ops. floor, they are dealing with nerves, confidence, a new environment, real people with real demands." • "There is a stress factor, talking to real people and real airplanes."
Large-system dynamic	40%	• "What isn't simulated well is the dynamic of the entire system; people are making decisions above, below, beside, all around you that can affect you, changing what is going on in your airspace as well. You are part of a bigger system."
Sim is trainee-centric (safe environment)	40%	• "In simulation, everyone beat's to the student's drum, instructors are there to facilitate their training and accommodate the student." • "When you are in the sim it's a safe environment because you have the instructor behind you, you can pause the clock and you know it's just created planes."
Seasonal Differences	10%	• "OJT needs to be long enough to cover the different seasons as [ATC] can be very different depending on the season."
Sim is by the book, live environment is more fluid	10%	• "Training is by the book, no corner cutting. In the real world, controllers use shortcuts and accepted workarounds and it's so far to that end of the scale that if someone does it by the book, it's considered inappropriate."
Shift-work	10%	• "Transition to shift-work."
Aircraft performance	10%	• "AC performance is attached to the training of specific skills to mastery before they get to the floor." • "Solutions in the sim would work only 50% of the time in the real world."

Higher fidelity simulation that incorporates some of these elements has been used before (e.g. the simulation used in Lee and Prevot, 2012, or Van de Merwe et al, 2012), but interviewees felt that simulating items such as a "Large-system dynamic", "Unpredictability" and "Dealing with real people" would increase the cost of training without significant added benefit. In fact, these key differences would likely prove significantly more expensive to simulate given the personnel working hours that would be required to not only design but to operate simulation incorporating some of these current differences.

The responses provided in Table 4.6 provide support for the components identified earlier in this chapter as well as additional components of the operational environment people are considering when making judgments about the fidelity of a simulation. "Dealing with real people" can be connected to the "Communications" component that was identified in section 4.1.3 given they both pertain to talking to other people. Some of these components, for example "Unpredictability", are more difficult to measure in terms of fidelity than parts of the physical environment, but it is important to understand that they still have an impact on the overall fidelity of a simulation and need to be accounted for when producing a definition of simulation fidelity.

The item of "Stress/nerves" is an important difference as replicating this feeling in a simulation is very challenging to accomplish. This being said, it is not considered a component that affects the fidelity of a simulation as it is a psychological component inherent to the user. As discussed in Section 2.2, the scope of the fidelity definition being created for enroute ATC does not include any psychological aspects.

It should be noted that aircraft performance is only mentioned by one participant in Table 4.6, while in sub-section 4.2.1 investigating the limitations of the simulation it was ranked as the limitation with the highest response frequency. Given interviewees were very adamant about the importance aircraft performance plays in the realism of a simulation, it would make sense for it to be higher up on the list in Table 4.6. This discrepancy is likely due to the researcher asking the question about the limitations before any of the questions investigating the key differences between the operational environment and simulation, as well as asking them relatively close together. It is possible that having just responded with aircraft performance to a similar type of question, they felt it necessary to focus on other aspects of the simulation as their beliefs regarding aircraft performance as a key difference between the two environments had already been registered.

While the differences identified above could potentially be generalized to various simulations used across the industry, it could also be the case that these differences are ANSP-specific. Therefore the following question was included on the industry-wide survey (as described in sub-section 3.2.3) in order to generate a set of differences that reflected simulation used across the industry: "In your opinion, what are 3 key differences between the highest fidelity simulation you have worked with and the enroute ATC work environment?" Table 4.7 presents the top 10 coded response frequencies and sample comments from this question.

Table 4.7 Key differences between a high-fidelity ATC simulation and the operational environment as identified by survey participants.

Code	Percentage of participants who identified code (N=84)	Sample Comments
Communications	52	• "Realism of the communication exchanges." • "Coordination between sectors and units." • "Phraseology."
Aircraft performance	41	• "Aircraft flight performance." • "Lack of accurate aircraft performance characteristics"
Equipment	24	• "Technology at the fingertips." • "Ergonomic layout."
Operational uncertainty	21	• "Route changes." • "Pilots do not always operate aircraft exactly the same way."
Safe environment	17	• "Level of controllers' stress: during simulation participants know that it is 'only' a simulation." • "Realism of the safety aspects missing in the simulation"
Weather	16	• "Realistic weather effects." • "Actual weather."
Physical Environment	13	• "No environmental noise and distractions."
Operator capabilities	12	• "Pilot proficiency."
Traffic	10	• "Background air traffic."
Large system dynamic	8	• "Multi-tasking with multiple aircraft and agencies competing for ATC attention."

Once again, not one key difference was identified by all of the participants, and only one difference was mentioned by over half the survey participants. Communications, which has strong similarities to the "Dealing with real people" difference identified by interviewees in Table 4.6, is the number one difference. As indicated by an interviewee, this reflects one of the hardest aspects of the operational environment to simulate as it is difficult to replicate the dynamic nature of real people with computer systems as well as how operators within the ATC system interact with each other. Communications is also a key part of a controller's job, so the fact that this is not replicated to a high degree of fidelity is one clear indication why simulation can only take trainees so far before they require time in the operational environment to develop more realistic expectations of communications between system operators. Also, supporting the hypothesis regarding the placement of the "Aircraft Performance" difference being unusually low in Table 4.6, here it is ranked number two and is therefore more consistent with what would be expected given interviewees' perception of that component as a main limitation. The items Operational uncertainty and Safe environment in Table 4.7 reflect similar items and ranks put forth in Table 4.6 by interviewees, except they are labeled “Unpredictability” and “Sim is student-centric”.

4.2.3 Dynamics of Trainee Transition from Simulation to Live Operations

A concept tied directly to the key differences enumerated in the previous sub-section is how a trainee handles the transition from the workstation simulation to the operational environment. At some point, each trainee has to make the transition if they are to continue on with their training and become qualified controllers. It is at this point where the differences between the two environments will be at their most pronounced as the trainee is only familiar with the simulation which has become their reality in a way. While there are evaluations that help make the determination of whether or not a trainee is ready to move on, further investigation was done during the interviews to determine whether or not interviewees had their own criteria to determine whether they felt a trainee was ready to perform in the operational environment. The following question was posed to interviewees: “How do you know when a trainee has reached the point where they’ve gotten all they can out of the simulation and are now ready to move on?”

From the responses provided to this question, three key indicators of trainee readiness were identified: (1) trainees displayed confidence in their abilities, (2) that their abilities are consistent with what they know to be good practice in the operational environment, and (3) that they demonstrate an understanding of the limits of their knowledge and skills. The third point is particularly important in a safety context, as it can be dangerous if a trainee tries to overreach their abilities in complex situations. Sample comments from the responses are provided in Table 4.8.

Table 4.8 Interviewee comments regarding when a trainee is ready to switch from simulated ATC to live operations.

Sample Comments
• "Conflict spotting on the board, trainees should be catching 95% of conflicts on the board and 5% on the radar. Trainees should be able to make every conflict work, whether it's using the instructor's method or a different approach that's safe. Not 'picking up the phone', hoping that situations go away by moving on to other tasks. Trainees are making things work even if it's not the best solution." • "From specialty to OJT, I would like to say that when we move them they are not going to have their confidence shattered, they are safe, an OJI [on-the-job training instructor] is going to feel comfortable putting that person on their license. They have a pretty good handle on working independently in the sim on a basic routine level of traffic that they would have down on the floor." • "Important attribute for instructors to see is that trainees know their limitations, know when to ask for help. They need to show they understand the job and their abilities within it."

Since it is "impossible to simulate the finesse of solving a conflict", a consequence of the simulation lacking realism in the areas identified in Tables 4.6 and 4.7 in the previous section, it is believed by interviewees that simulation is best used in a training context to develop mastery in the simpler aspects of the job and the basic concepts of conflict solving. This resonates with a comment made in Table 4.5 about Aircraft Performance, in which it is stated that half of solutions trainees use in simulation would not work for real operations. In fact, simulation limitations are identified as a trigger for the need to transition to the operational environment.

> "When they have learned whatever skill they are supposed to have learned from the course, and what they have to learn next they can't learn in the simulator. **There's a blockage there when you tell the students that this is what you have to do in certain situation but then they answer, 'Well, that doesn't happen.'** When you reach those limits, those blockages between the students and the trainers, you don't really want to fight with them cause it's true in their world. That's sort of when you know, the next thing we have to teach you is the live environment."

As noted by interviewees, when a student reaches the limits of a simulation's capabilities they will begin practicing bad habits and developing false expectations. Based on the responses from interviewees, the differences and limitations identified in the preceding sub-sections are most likely the underlying reasons why those being trained using simulation cannot simply start doing the job unsupervised after they have completed the simulation portion of their training. There is still too large a gap between simulation and the operational environment that requires further training time to overcome, demonstrating the limits of the fidelity of the simulation used at this particular ANSP.

4.3 Identifying Specific Environmental Components that Affect Fidelity

When interviewees and survey participants identified the simulation limitations and key differences from the operational environment within this chapter, they were also providing specific components of the operational environment that they believe affect the fidelity of a simulation. This section will discuss some of the components provided and the next steps that must be taken in order to form an objective definition of simulation fidelity.

Components from Simulation Limitations The limitations identified in sub-section 4.2.1 in Table 4.5 begin to provide a deeper understanding as to the aspects of the simulation interviewees are considering when making fidelity determinations. For instance, many of the interviewees believe that “Aircraft Performance Characteristics” are a key limitation with the workstation simulation at the ANSP. When identifying this environmental component as a limitation, it also indicates that interviewees are identifying it as a component that they believe has an impact on the fidelity of a simulation for enroute ATC. The other two key limitations identified, “Communications” and “Equipment”, have also been identified as environmental components that can affect the fidelity of a simulation in section 4.1.2, further supporting their importance in determining the level of fidelity of a simulation. That is not to say that only the components included in Table 4.5 are relevant to the fidelity discussion, but it begins to identify certain components individuals are considering when making fidelity judgments and thus ought to be considered when developing a formal definition of simulation fidelity for enroute ATC.

Components from Simulation Differences The key differences between the operational environment and simulation provided by survey participants in Table 4.7 offers several more potential fidelity components to consider. There are similar components to those mentioned in the limitations, such as “Communications” and “Aircraft Performance”, but also other components such as “Equipment”, “Operational Uncertainty”, “Weather”, and the physical environment. The fact that “Communications” and “Aircraft Performance” appear in so many different locations only serves to reinforce their potential importance when discussing what components affect the fidelity of a simulation for enroute ATC. This does not mean that components with a low mention rate in Table 4.7, such as “Equipment”, are not important to the fidelity discussion. That table presents results regarding the differences between the operational environment and simulation. The items provided in that table reflect components that are typically not simulated to a great degree of accuracy. As was noted earlier, the components provided here only begin to illustrate *some* of the components individuals are considering when making a fidelity determination. More work is needed to explicitly identify the components individuals are considering and if a consensus can be formed. The results of this work are presented and discussed in Section 5.2.

Summary The differences identified in Tables 4.6 and 4.7, along with the limitations of the workstation simulation identified in Table 4.5, not only indicate what components of live operations a simulation is not able to replicate very well, but also what environmental components people are considering when making determinations about the fidelity of simulations. These differences begin to highlight specific environmental components, such as "Communications", "Aircraft Performance" or "Equipment", that people are using as comparative points to determine the differences in fidelity between various simulations.

4.4 Chapter Summary

This chapter has provided a case study of how three distinctly different simulations are used at an ANSP for training, and more specifically has identified the key limitations with a higher fidelity simulation and the main differences between that level of fidelity and the operational environment. This chapter primarily serves to accomplish the first objective, which is to identify how individuals are currently making fidelity determinations with regards to simulation.

The differences between the simulation and operational environment identified in sub-sections 4.2.1 and 4.2.2 illustrate some of the specific components of the operational environment that individuals within the industry believe affect the fidelity of a simulation, thereby potentially meriting inclusion in a standardized definition of simulation fidelity for enroute ATC. The response frequencies regarding the differences between high fidelity simulation and live operations provided by survey participants, as presented in Table 4.7, begins to show why there is a need for this standardized definition. The fact that not one of the differences identified in Table 4.7 is close to being provided by all interviewees and survey participants begins to indicate that people may not share the same representations of simulation fidelity within the domain of enroute ATC. These inconsistent representations of simulation fidelity are further explored in *Chapter 5 Industry Perceptions Regarding Simulation Fidelity in ATC*.

Chapter 5

Industry Perceptions Regarding Simulation Fidelity in ATC

This chapter further investigates current ATC industry perceptions regarding the concept of simulation fidelity, following on with the potentially incongruous representations of simulation fidelity identified in the previous chapter. Findings presented within this chapter are separated into three sections. The first section, Inconsistency in Individuals' Simulation Fidelity Representations, illustrates the different perceptions that exist regarding the concept of simulation fidelity within the domain of ATC and the confusion which arises from these different perceptions. The second, Source of Inconsistent Fidelity Representations, section provides potential evidence regarding the probable source of these incongruous representations of simulation fidelity. The final section, Simulation Categorization Systems Based on Fidelity, addresses the use of vague fidelity terminology such as low, medium, and high for describing the fidelity of simulations and how these terms are insufficient for objectively assessing the fidelity of a simulation. Together these sections serve to develop a clearer picture regarding how the concept of simulation fidelity is perceived within the industry, thus contributing to achievement of the first objective of this thesis, and also motivating why the fidelity constructs are required for enroute ATC. This chapter is based upon a combination of data gathered from the site visits to the ANSP and the online survey.

5.1 Inconsistency in Individuals' Simulation Fidelity Representations

Is Simulation Fidelity Well-Defined for ATC? An important first step in investigating industry perceptions regarding simulation fidelity is to determine if people within the industry believe that simulation fidelity is already a well-defined concept in the domain of ATC. In order to determine to what extent simulation fidelity is already perceived to be well-defined, a question was included in the survey asking participants whether or not they believe simulation fidelity was a well-defined concept in the domain of ATC. Figure 5.1 presents the results from responses to this question.

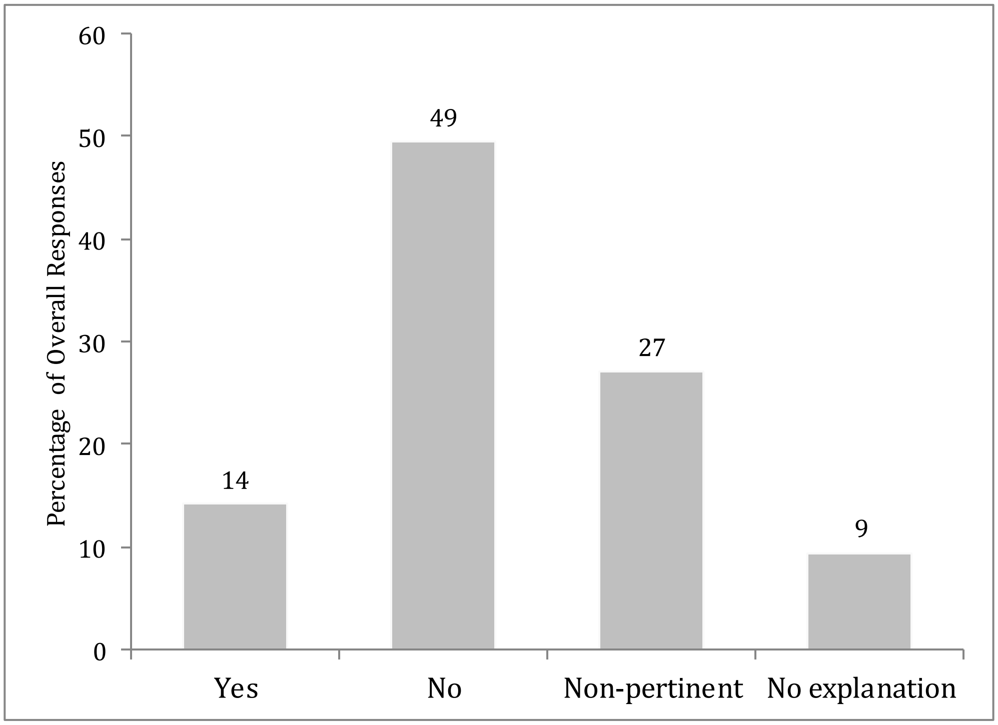

Figure 5.1 Survey participant responses to the question: "Do you believe that simulation fidelity is a well-defined concept in the ATC domain?" N=85 (Question 11 in Appendix B)

The 'Yes' and 'No' columns in Figure 5.1 represent responses where participants demonstrated a clear understanding of the question based on their follow-on explanation of why they answered Yes or No. Survey participant explanations that clearly indicated they did not understand the question were categorized as 'Non-pertinent', a process described in section *3.2.3 Simulation Fidelity Survey*. The final column, 'No explanation', represents the percentage of responses where survey participants provided no explanation to their Yes/No answer and therefore an assessment of their understanding of the question could not be established.

Results of a chi-square goodness of fit test show that the observed Yes/No response frequencies are statistically significant. They are lower than what would be expected if half of the participants believed that simulation fidelity is well defined for the ATC domain (χ^2 (1, N=54)=16.67, p<0.001). Demographic data collected as part of the survey was used to investigate whether the perception that fidelity is not well defined is wide-spread across gender, nationality, experience and primary use of simulation. A comparison between the Yes/No response rates for these four demographics is presented in Figure 5.2. As seen in the figure, the proportion of Yes/No responses, while varied, shows a strong and consistent pattern of a belief that simulation fidelity is not well defined. A chi square analysis was performed to determine if there were any differences within the demographic groups. It was found that there were no differences with regards to the belief that simulation is **not** well defined for ATC when comparing within the demographic groups of gender, (χ^2 (1, N=54)=0.04, p=.851), nationality, (χ^2 (2, N=54)=2.06, p=.385), years working with simulation, (χ^2 (3, N=54)=3.78, p=.287), or survey participant's use of simulation, (χ^2 (2, N=54)=1.83, p=.400).

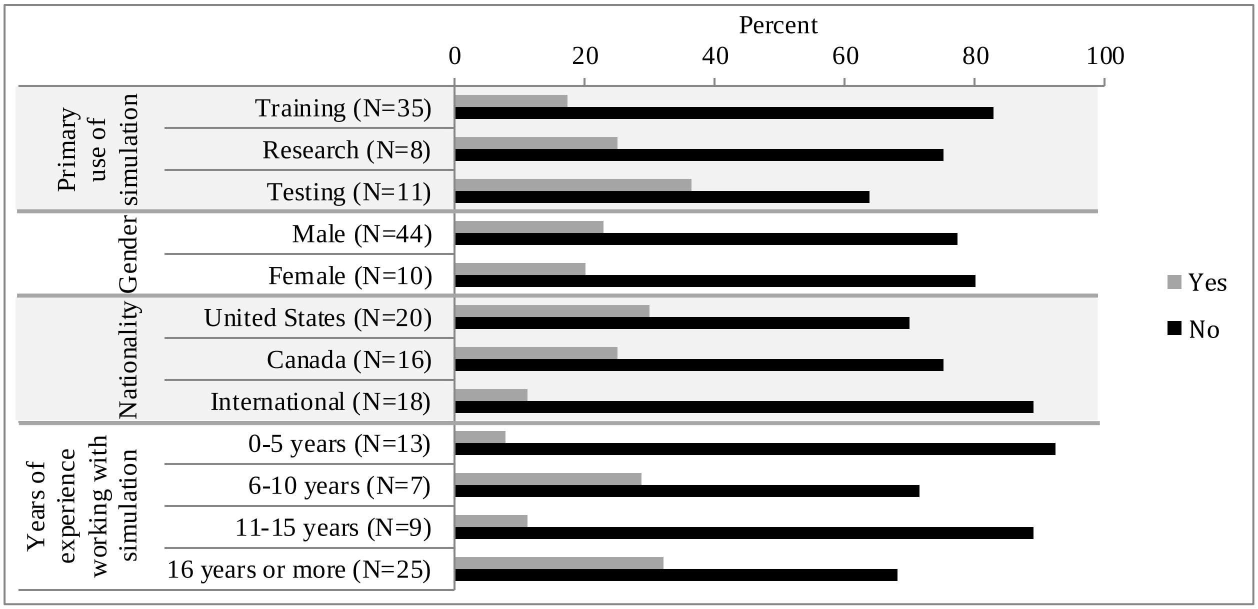

Figure 5.2 Demographic sub-group comparison of responses to the question: "Do you believe that simulation fidelity is a well-defined concept in the ATC domain?"

In order to further explore why survey participants feel simulation fidelity is not well defined for ATC, Table 5.1 presents sample comments from both the pertinent "Yes" and "No" responses to the follow-up question asking if they could explain their answer in more detail.

The sample comments from those who responded "Yes" are representative of a recurring belief that simulation fidelity is a well-defined concept, but is not put into practice or referenced enough with regards to the many uses of simulation within the industry. However, what is clearly demonstrated in the sample comments from those who responded "No" is that the problem is not with an individual's definition in isolation, but rather when discussing the issue as a collective and not sharing the same definition with those they interact with. Comments such as "I believe simulation fidelity means different things to different people", "On the contrary, many times the term "high fidelity" is interpreted in various ways", or "I don't believe this [his/her interpretation of fidelity] to be a universally shared interpretation and that there are varying degrees of separation from my idea", all indicate an awareness of the impact of a lack of standardization with regards to simulation fidelity in the ATC domain.

Table 5.1 Sample comments from survey participants' explanations of their responses to the question in Figure 5.1.

Sample Comments from 'Yes' Responses	Sample Comments from 'No' Responses
• I think that it is well-defined, but in reality, it is under-utilized. • We all know what fidelity means. Realistic. Realistic aircraft, realistic routes, realistic responses. Responses that are dynamic in nature, changing depending on what the student is doing. • Though I'm not aware of a quantitative definition, fidelity is something researchers and trainers know when we see it, and it is easy to ordinally rank different simulators or simulations in terms of their fidelity. I have created and used an informal table that lists the different levels of fidelity and their characteristics. • I think it's defined and conceived just fine, but, in my opinion, it's not implemented very well.	• I think that "simulation fidelity" is one of those concepts that "everyone knows what it means" but that formal, valid definitions are lacking. • I believe simulation fidelity means different things to different people. I believe current controllers are not involved enough in validating the fidelity of a simulation before it is used in the field. • I have not come across such a concept definition so far. On the contrary, many times the term "high fidelity" is interpreted in various ways. • I've met a lot of people in my business who have a significantly different perception of what is high and what is low fidelity simulation. • My interpretation of high fidelity simulation is the recreation of the real live ATC environment in as much detail as possible. I don't believe this to be a universally shared interpretation and that there are varying degrees of separation from my idea.

Similar High-Level Understanding of Fidelity Even though survey participants believe that simulation fidelity is not a well-defined concept within the domain of enroute ATC, both interviewees at the ANSP and survey participants demonstrated a consistent understanding of the high-level concept of simulation fidelity.

Many interviewees at the ANSP offered a definition of simulation fidelity similar to the first half of the definition provided in section 1.1 as the high-level fidelity definition used as the foundation for this research project: "Simulation fidelity is the degree of similarity between the training situation and the operational situation which is simulated". While not all interviewees provided the same exact definition, they were consistent with the general understanding of the concept.

Survey participants demonstrated a similar view of the general concept of simulation fidelity. They were asked what keywords or phrases come to mind when they think of simulation fidelity (see Question 10 in Appendix B). Table 5.2 presents the frequencies of the top 8 coded responses along with sample responses for each code. Even though participants were asked to give keywords thereby potentially eliminating the need to code responses, coding was still necessary as some keywords provided by survey participants were quite similar and could be grouped under one term (e.g. 'accuracy' and 'accurate' being coded as "Accuracy") or the response was longer than just one word and a singular code describing that phrase was required (e.g. 'accurately replicating aircraft performance' being coded as "Aircraft performance").

The two most common codes were "Realism" and "Accuracy". These terms are consistent with the high-level definition of simulation fidelity noted above in that simulation fidelity is the degree of *realism* or *accuracy* between the training situation and the operational situation which is simulated. Other coded responses that survey participants provided, such as "Aircraft Performance", "Traffic", "Scenario", and "Equipment", offer insight into the specific components of the enroute ATC environment participants believe affect the fidelity of a simulation. These components make up the more specific definition of fidelity and are discussed in further detail in Section 5.2.

Table 5.2 Frequency of coded keywords and phrases regarding survey participants' understanding of simulation fidelity. (N=84)

Code	Percentage of participants who identified code (% of N)	Sample Comments
Realism	54.7	• Realistic • Realism • Degree of realism
Accuracy	17.4	• Accuracy • Accurate
Training	12.8	• Learning • Practice • Prepare better for controlling real aircraft
Fidelity	10.5	• Process fidelity • Psychological fidelity • Low fidelity
Aircraft performance	10.5	• Aircraft performance accuracy • Similar air traffic behaviour to real world • Accurately replicating aircraft performance
Scenario	10.5	• Real life conflict situations • Scenarios • Real world scenarios
Traffic	9.3	• Accurate replication of live traffic • Realism in traffic flow • Traffic
Equipment	9.3	• Real appearance of working position • Display • Hardware/software look and feel

Inconsistent Specific Definitions of Simulation Fidelity Even with this same starting point, however, interviewees than diverged into very different interpretations of what influenced the fidelity of an ATC simulation. This divergence was vividly illustrated in the comment of one interviewee provided below.

> "One thing I find in talking with all sorts of people is the interpretation of the fidelity in everybody's head is very different. Sometimes I think that there is a lot of frustration with perceptions on what fidelity means. High fidelity is not defined and everybody has their own definition of what it could be in their head. This results in frequent miscommunications."

Evidence of this lack of congruity can be seen in responses to a question asking interviewees to compare perceptions of two different simulations used at the ANSP: "On a scale of 1 to 10, with 1 being not similar in any way and 10 being an identical replication, how close is the workstation simulation to the actual working environment? Same question, but for the PC-based simulation?" The results of this question are summarized in Figure 5.3. The question asking interviewees to rate the PC-based simulator was not asked to all interviewees, as some were not as familiar with it.

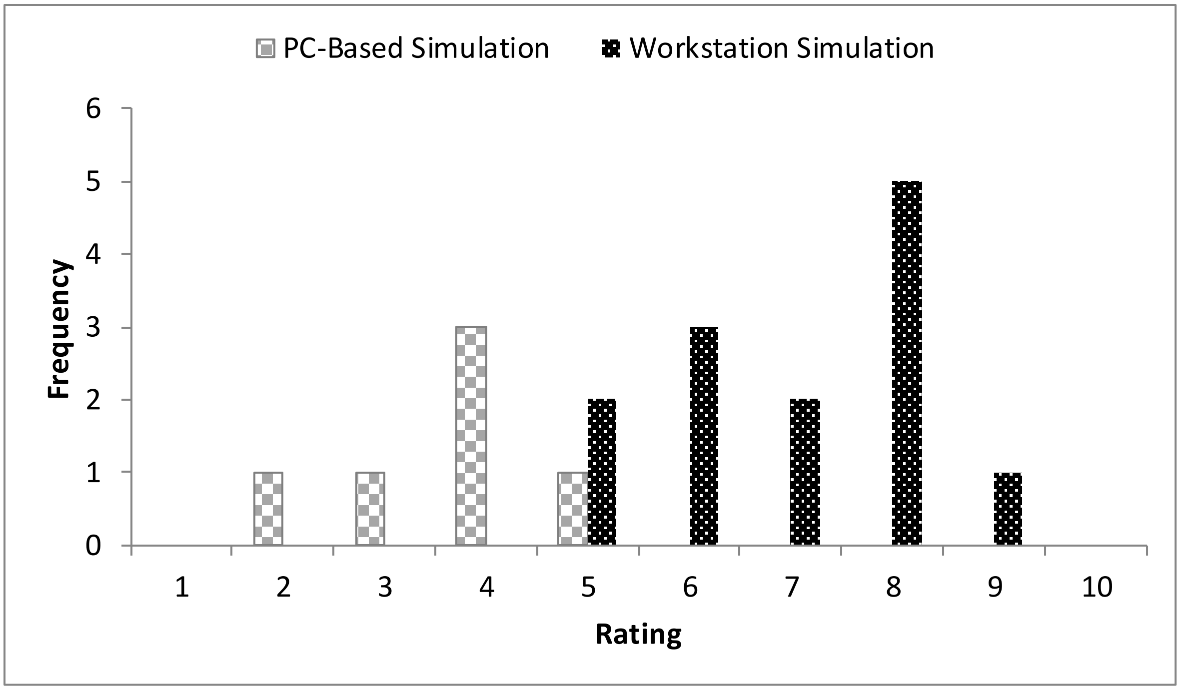

Figure 5.3 Comparison of interviewee fidelity ratings of the PC-based and workstation simulators at the ANSP. (Workstation N=13, PC-based N=6)

Figure 5.3 indicates that participant responses demonstrated both the expected difference in average ratings of the two different simulations as well as a large spread in the ratings for each individual simulation. There was also an overlap in ratings between the PC-based and workstation simulations. This overlap indicates that different participants rated these two distinct simulations at the same level of fidelity. This demonstrates the difference in people's perceptions of simulations, and is an indicator of how miscommunications can occur when discussing simulation.

In order to further investigate this finding on a larger scale, a similar question was included in the online survey. Survey participants were asked to rate, on the same 1-10 scale described for Figure 5.3, the fidelity of the highest fidelity simulation they have ever worked with. Once they had rated it, they were asked to provide a brief description of the simulation. In order to compare the findings from this question with those from Figure 5.3, only survey responses from those who worked at the same ANSP as interviewees were considered[2]. The results are presented in Figure 5.4[3].

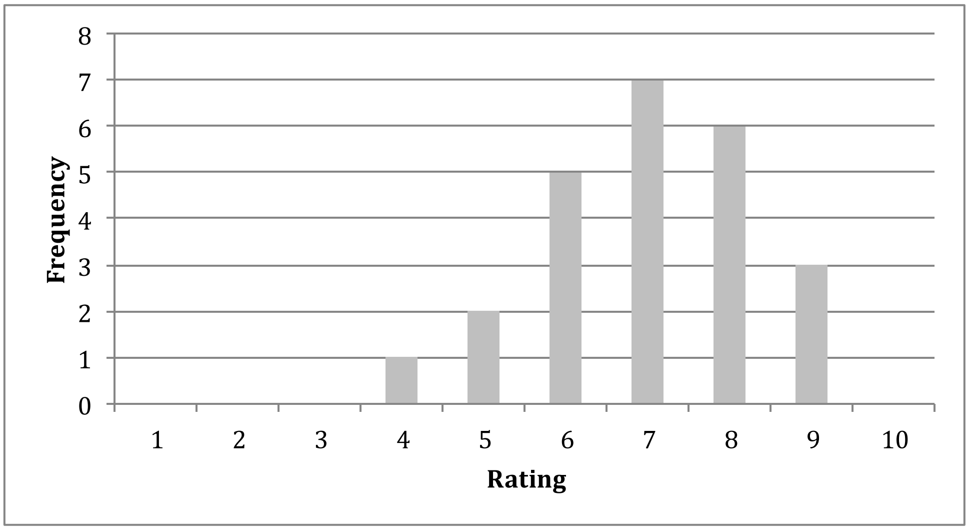

Figure 5.4 Survey participant fidelity ratings of the highest fidelity simulation environment they have worked with. (N=20-30)

---

[2] It is believed that the highest fidelity simulation for survey participants from this particular country is the same as the workstation simulation considered by interviewees in Figure 5.3, thereby making the results presented in Figure 5.4 comparable to those in Figure 5.3

[3] A range is provided for the number of survey participants to ensure anonymity of respondents.

A similar pattern to that of the workstation simulation from Figure 5.3 emerges as a wide spread of ratings is provided with the majority falling around 7 or 8. Once again, individuals are viewing the same simulation at varying degrees of fidelity based on the different representations of simulation fidelity they likely possess.

It is hypothesized that the underlying cause of this spread in perceptions is that individuals value components of the simulation differently, which leads to the different interpretations of enroute ATC simulation fidelity. Without a standardized and objective definition of simulation fidelity, people operate under their own beliefs and assumptions as to what influences fidelity and make decisions based on how important they feel that component is. Combined with a lack of knowledge regarding how best to use simulation of varying degrees of fidelity, this will create a variety of different approaches to using simulations and potentially create inconsistent and inefficient training.

5.2 Source of Inconsistent Fidelity Representations

As noted when discussing Figures 5.3 and 5.4 in the previous section, the source of the inconsistent representations of simulation fidelity is hypothesized to be a difference in individuals' inclusion of certain environmental components within their own definition of simulation fidelity along with how they value those components.

To explore this hypothesis, the first step was to ask interviewees at the ANSP what components of the operational environment they felt affected the fidelity of an enroute ATC simulation in order to identify any differences in individuals' sets of components. This question was posed to interviewees before being shown the set of components developed for this project in order to avoid biasing their responses. The responses to this question were then coded using the approach discussed in section *3.2.1 Semi structured Interviews*, and response frequencies were subsequently calculated. Table 5.3 presents the results of the analysis of this question.

Table 5.3 Coded response frequencies and sample comments from interviewee identified components that affect simulation fidelity for enroute ATC. (N=13)

Fidelity Components	Response frequency (% of N)	Sample Comments
Equipment	77	• The equipment difference, all the latest features within 12 months are not available within the simulators • Equipment
Operating System Functionality	54	• How one piece affects another piece (if I can click on a target and change his altitude and then a strip is printed with that new altitude, that is exactly what happens on the ops floor) • Simulation functionality
Aircraft Performance	54	• Aircraft performance • Realism of aircraft performance
Communications	54	• Communications with pilots and controllers • Coordination between their position they are working at the time and outside units
Environment	31	• Physical similarity of environment • The work environment
Scenario	23	• The traffic situations reflecting real-life eventually, can stage it to learn a skill set, but at some point it would be nice if the scenarios mimicked real situations
Airspace	8	• Sector boundaries should always be the same, airspace should be 100%
Unpredictability	8	• Realism of the unpredictability

The component with the highest mention rate, "Equipment", is to be expected given that a simulation user's initial fidelity impressions of a simulation are most likely tied to the fidelity of the equipment as that is what they perceive first. Only three other components were mentioned by at least half of the interviewees. There are similarities between components noted above and some of the components mentioned throughout Chapter 4 and discussed in *Section 4.3 Identifying Specific Environmental Components that Affect Fidelity*, such as the inclusion of the "Equipment", "Aircraft performance", and "Communications" components.

Taken as a whole the components listed above could offer a reasonable consensus of the components which can affect the fidelity of an enroute ATC simulation; however, it is clear that these components are not unanimously agreed upon given the response frequencies or the fact that there are so few largely shared components amongst the interviewees. This begins to illustrate where peoples' different representations of fidelity, as demonstrated in Figures 5.3 and 5.4, likely come from. This being said, a key limitation with these findings is that they are based off feedback from a relatively small group of interviewees from a single ANSP, limiting the ability to generalize these findings to what is a large industry.

In order to account for this limitation, a question was included in the online survey that asked survey participants to list the components which they felt could affect the fidelity of a simulation of the enroute ATC work environment (Question 16 in Appendix B). Eight text boxes were provided to survey participants in the response area in order to offer participants the ability to provide as many different environmental components as possible. The top 8 components identified from the responses to this question are presented in Table 5.4, using the same approach to analysis as the results presented in Table 5.3.

As with Table 5.3, there is a lack of agreement amongst survey participants on a clear set of components that affect the fidelity of an enroute ATC simulation with "Communications" being the only component mentioned by over half of survey participants. There is agreement, however, between responses from interviewees and survey participants in terms of the most important components based on response frequency, with “Communications”, “Equipment”, “Aircraft Performance” and “Environment” appearing high in both Tables 5.3 and 5.4. This potentially suggests there are key components both interviewees and survey participants believe affect the fidelity of a simulation more significantly than others.

Table 5.4 Coded response frequencies and sample comments from survey participant identified components that affect simulation fidelity for enroute ATC. (N=73)

Fidelity Components	Response frequency (% of N)	Sample Comments
Communications	62	• Communications both controller/pilot and between controllers • Communication with other facilities
Equipment	42	• Equipment usability • Changes of operational equipment that don't get brought into the simulator environment
Environment	32	• Background noise/distractions that occur in the operational environment that don't occur in the simulator. • Successfully replicating the atmosphere of a control room
Aircraft performance	30	• Atypical aircraft performance • Realistic aircraft performance modelling
System participants	29	• Experience of the pseudo pilots providing supporting traffic to the simulation • Pilot actions
Unpredictability	29	• Off-nominal situations • Unusual but realistic 'odd' requests (extend downwind, stay high, early descent, etc.)
Scenario	23	• Complexity of traffic flow • Traffic volume changes are more dramatic in real life, both increase and decrease
Weather	21	• Realistic turbulence and weather scenarios • Weather simulations -- changing conditions, moving thunderstorms, varying winds and visibility.

One component that is presented in Table 5.4 that does not appear in Table 5.3 is "System participants". "System participants" refers to the simulation's replication of the other operators within the system aside from the primary operator of the enroute ATC simulation, such as pilots and controllers in other sectors. This component captures how consistent the actions these system participants take within the simulation reflect those of real world operators.

In addition to the overall response frequencies for survey participants presented in Table 5.4, the response frequencies for the demographic groups of nationality, gender, survey participant's primary use of simulation and survey participant's years of experience with simulation were calculated. The results for the nationality demographic group are presented in Table 5.5.

Table 5.5 Top ten coded fidelity component response frequencies for all survey participants with nationality demographic group comparison.

Fidelity Components	Response frequency (% of N)			
	Overall (N=73)	United States (N=31)	Canada (N=24)	International (N=18)
Communications	62	55	71	56
Equipment	42	35	46	44
Environment	32	32	42	17
Aircraft performance	30	16	46	28
System participants	29	23	38	22
Unpredictability	29	19	42	28
Traffic	23	19	25	28
Weather	21	10	42	11
Automation	19	16	17	28
Operational stress	11	6	13	11

Tables similar to Table 5.5 were also prepared for the demographic groups of gender, simulation use, and years of experience with simulation, and are presented in Appendix D. Across all demographic groups, the "Communications" component received the highest response frequency for each sub-group, indicating its high overall rank was the result of a widespread and shared perception of its importance for a definition of fidelity for enroute ATC simulations. Not all components appear to be perceived equally across the different nationality groups, though statistical tests of significance have not been completed. For instance, Canada had a much higher response frequency for "Weather," while the United States had lower response frequencies for "Aircraft performance", and the International group had lower response frequencies for the "Environment" component but higher response frequencies for "Automation". From the tables shown in Appendix D, the researchers demographic group overwhelmingly identified "Communications" (73%) and "Equipment" (45%) components, while all other components were at less than 27%. The demographic group of testing new procedures had almost no (< 7%) mentions of "Unpredictability", "Weather", "Automation", and "Operational Stress". Table 5.5 also illustrates that there were differences in how many components each nationality was providing, with Canadian survey participants providing more components then the other two groups.

The responses from both interviewees in Table 5.3 and survey participants in Table 5.4 and 5.5 offer tangible evidence to support the hypothesis that individuals possess different mental representations of simulation fidelity due to their inclusion of different components within their own personal definitions. Further work is likely necessary to better identify how each component is valued by individuals but it is clear that there is a disparate view of what components ought to be considered when identifying the fidelity of an enroute ATC simulation. The fact that there is a clear lack of a standardized definition of simulation fidelity for ATC, as indicated by the analysis of Figure 5.1 and Table 5.1, likely contributes to peoples' varying mental models of the components that affect the fidelity of a simulation. Developing a standardized, objective definition of simulation fidelity for enroute ATC will most likely reduce this variability and bring more people onto the same page with regards to this concept within the domain of enroute ATC.

5.3 Simulation Categorization Systems Based on Fidelity

Even if a simulation fidelity definition is developed and accepted by the industry, there is a need for a categorization system that operationalizes the definition. A simulation fidelity definition would provide the points of comparison for simulations whereas a categorization identifies the broader levels of fidelity that exist within the domain. This allows for the ability to easily and directly compare the fidelity of a variety of simulations.

There are three important questions that must be considered when discussing categorization systems: (1) is it useful to categorize simulations within a given domain, (2) how should the categories be communicated or presented, and (3) how are the categories determined? It can be easily argued that many believe it is useful to categorize simulations as most people do so sub-consciously when they use the popular fidelity terminology of low, medium and high.

While these terms do offer an indication of a simulation’s level of fidelity, there are no objective criteria attached to these terms in order to define what makes a simulation low, medium or high, and is often the result of an individual’s subjective assessment. This could result in different definitions of each term as one individual’s definition of low, medium, and high will not always be the same as another’s. In order to investigate how the ATC industry views these fidelity-describing terms, a question was included in the survey asking participants whether or not they felt these terms were useful (Question 22 in Appendix B). Figure 5.5

presents the findings from the Yes/No portion of the question using the same analysis format as Figure 5.1.

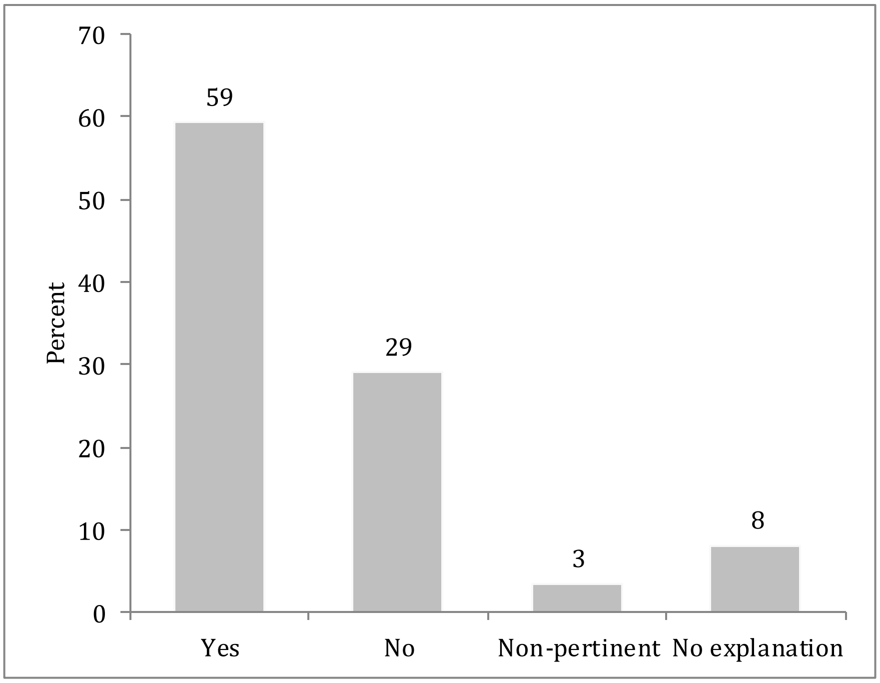

Figure 5.5 Survey participant responses to the question, "Do you feel that the terms low, medium and high for describing fidelity are useful?" (N=86)

Results of a chi-square goodness of fit test show that the observed Yes/No response frequencies are statistically significant. They are significantly higher than what would be expected if only half of the ATC industry believed that terms 'low, medium, and high' were useful (χ^2 (1, N=76)=8.89, p=0.003). Demographic data collected as part of the survey was used to investigate whether the perception that these terms are useful is wide-spread across gender, nationality, experience and primary use of simulation. The proportion of Yes/No responses, while varied, indicate a strong and consistent pattern of a belief that these terms are useful, as indicated by Figure 5.6. A chi square analysis was performed to determine if there were any differences within the demographic groups. It was found that there were no differences when comparing within the demographic groups of gender, (χ^2 (1, N=76)=0.04, p=.834), nationality,

(χ^2 (2, N=76)=5.13, p=.077), years working with simulation, (χ^2 (3, N=76)=1.29, p=.731), or survey participant's use of simulation, (χ^2 (2, N=76)=1.83, p=.780). This finding could be interpreted as indicating that these terms are sufficient in differentiating between the fidelity of simulations and, based on the third question raised at the beginning of this chapter, are acceptable as the different categories of fidelity for enroute ATC.

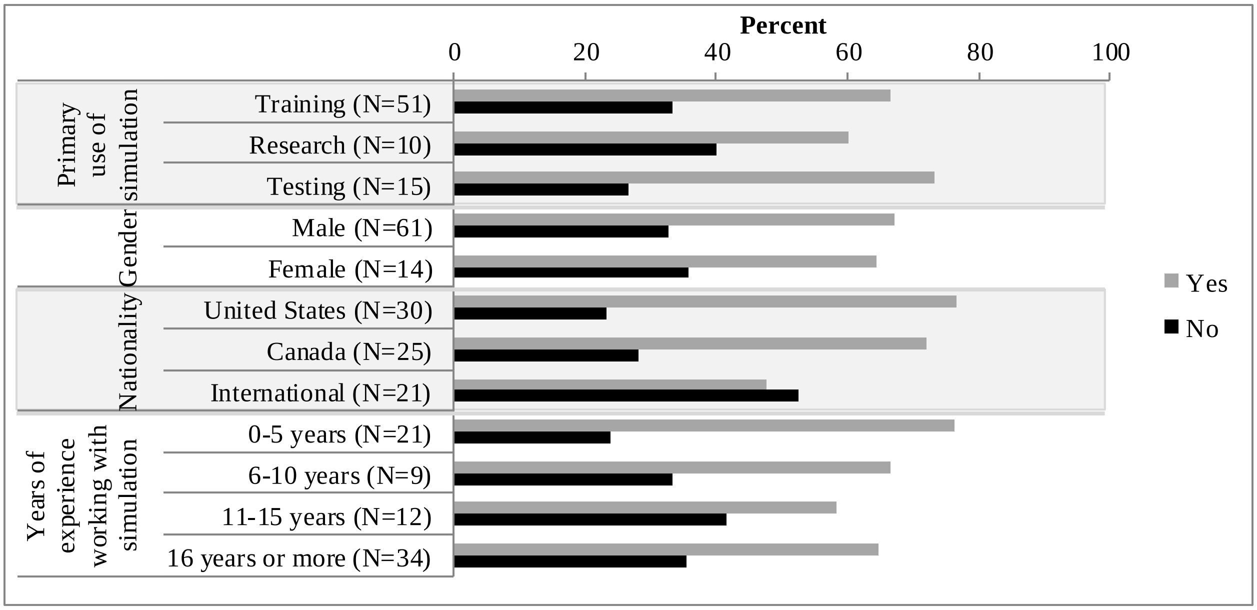

Figure 5.6 Demographic sub-group comparison regarding the usefulness of the fidelity terms "low, medium, and high".

From Figure 5.5, however, it cannot be determined if these terms are sufficient to make objective comparisons between simulations, or if more work needs to be done. The question asked if those terms were useful, and they are useful in the sense that they are better than having no means of differentiating between the fidelity of simulations at all, thus helping to answer the first question at the outset of this section of whether or not a categorization system would be useful for this domain. In this respect and with hindsight, the question should have been worded differently as this format did not provide the most relevant data in determining if the fidelity terms low, medium, and high are sufficient for comparing simulations in an objective manner.

However, a follow-on portion of the question asking survey participants to explain their Yes or No answers was provided. Sample comments from those who understood the question are provided in Table 5.6. These comments provide a clearer indication of the insufficiency of the fidelity terms low, medium and high and why further work is indeed necessary.

The clear and pervading theme is that the 'low, medium, and high' terminology offers an easily-understood, broad categorization of the simulation being used. What is seen in a few of the 'Yes' sample comments, however, is that there is a desire for more specificity in the categories of simulations available even though they felt low, medium and high were useful. This begins to respond to the question of how the categories of simulation fidelity for enroute ATC are determined, the third question noted at the outset of this section. One comment even mentions that a categorization system's usefulness "is dependent upon well-defined criteria and common understanding of these criteria for the gradations," something which is lacking from the fidelity terminology of low, medium, and high.

This desire for increased specificity and standardization is echoed strongly in the sample comments from survey participants who responded 'No' in Figure 5.5. They believe the low, medium and high system of terms suffers from a lack of well-defined criteria and is therefore not useful but rather misleading, with one comment indicating that the vague fidelity terminology is "abused". As discussed at the outset of this section, one of the most important questions when developing a categorization system is how the categories are determined. There is, therefore, a strong belief that a categorization system ought to be developed based on clearly and objectively defined criteria, such as an objective and validated simulation fidelity definition.

Table 5.6 Sample comments from survey participants' explanation of their answers to the question presented in Figure 5.5.

Sample Comments from 'Yes' Responses	Sample Comments from 'No' Responses
• It does point the brain right away to what someone would be working with. Is it a basic, entry-level tool just to get you started, or is it a full-blown mock-up of the real thing? • Yes, they are useful to broadly describe the simulated environment but it is not enough detailed. Practitioners should be able to describe first why it is low/medium or high, what type of elements are low, medium and how it can affect/impact measures/assessments. • Gradations of fidelity are very useful to evaluate options and weigh requirements against costs for options. I would add though that their benefit is dependent upon well-defined criteria and common understanding of these criteria for the gradations. • Allows a rough categorization of fidelity level without being too complex. Having said that, in some cases, a slightly more complex definition might be required, e.g. a simulation environment might have different degrees of fidelity on your various dimensions.	• Because the terms do not specify where the simulation is lacking fidelity. There are multiple types of fidelity in ATC simulation, traffic flow, equipment, environmental, and scenario etc... • It does not indicate what aspects of fidelity are low medium or high. If that were defined - it would be more useful. • You need to have more choices in a rating criterion if you are really going to flesh out a concept. • If you don't define further what is what this classification can be and is abused. • There is no benchmark for what constitutes these levels of fidelity. It is too subjective. • Because there isn't a common standard stating what those terms result in, it's all relative and to a degree based on personal experience. It would be similar to asking what low, medium, and high levels of air traffic are - there would be some similarities, but the spectrum of responses would vary significantly.

This theme of increased standardization and objectivity also appears when participants were asked explicitly about the usefulness of a more standardized approach to differentiating between simulations within ATC. Survey participants were asked whether they believed a simulation categorization system, similar to the full flight simulator categorization system developed by the FAA presented in section *2.3 Examples of Classifying Simulation Environments in the Literature* (which was presented to survey participants when asked this question), was required for the ATC industry. This line of inquiry seeks to address the second question presented at the outset of this section, in essence how should the different categories be presented. Results from this question are presented in Figure 5.7 using the same analysis format as Figure 5.1 and Figure 5.5.

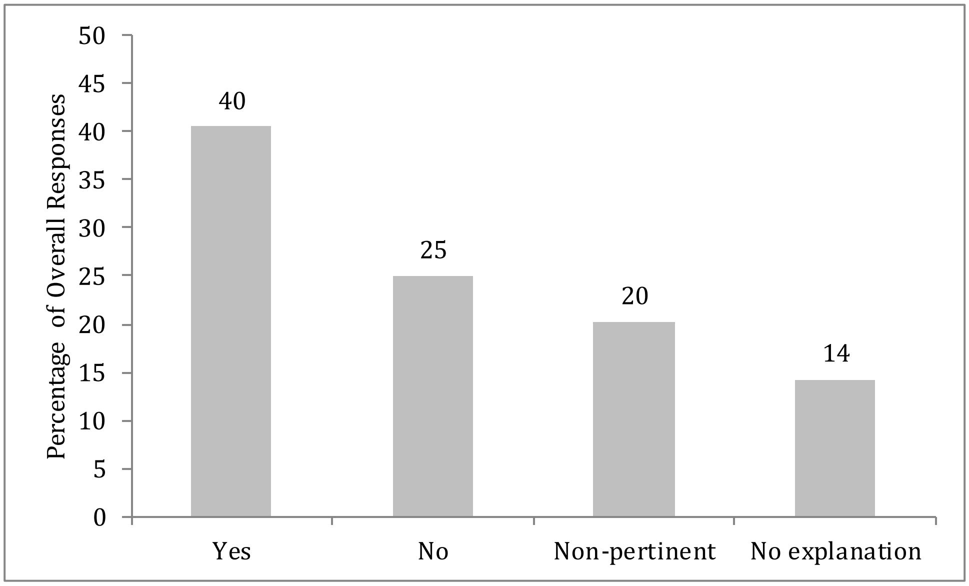

Figure 5.7 Survey participant responses to the question, "Do you feel that a standardized simulation categorization system similar to the FAA flight simulation categorization system but adapted for enroute ATC simulations is required?" (N=86)

Results of a chi-square goodness of fit test show that the observed Yes/No response frequencies were not statistically significant. They are not different than what would be expected if only half of the ATC industry believed that a categorization system was required (χ^2 (1, N=55)=2.62, p=.106).

The fact that Yes responses were not significant but trending in that direction is likely due to the question asking if the categorization system ought to be 'required', as opposed to asking if it would be a useful tool for those who use simulation. An absolute term such as 'required' may have caused survey participants to be a bit more conservative in making their judgment as there are potentially consequences for making a categorization required. To further explore this issue, sample comments from survey participants' explanations of their responses in Figure 5.7 are provided in Table 5.7 for more in depth analysis.

Table 5.7 Sample comments from survey participants' explanation of their answers to the question presented in Figure 5.6.

Sample Comments from 'Yes' Responses	Sample Comments from 'No' Responses
• I think that more description is needed than just low, medium, or high and this criteria adapted to ATC would allow for consistency in describing fidelity. • It would be useful in deciding what simulation is best suited for the training that is to be accomplished. • It would finally allow ATC to speak in concrete terms about simulation, rather than abstract terms of high or low fidelity simulation. • This would provide operational grounding and consistent categories for use across all studies. • A formal categorization would aid in judging the value and benefit of a given simulation study or effort. • At the moment there is a variety of expectations when people talk about ATC high fidelity or low fidelity simulation. Some standard is required to clarify this situation.	• Not sure it needs to be required but would be helpful in planning out resources during the design phase of training curriculum development. • Some standards while valuable, inhibit innovation and are not forward looking or responsive to changes in technology. Perhaps evolving guidelines is better? • Required, no; potentially useful, yes. If adapted the lower levels would be akin to the plastic airplanes, strip writing exercises with the highest level being the complete reproduction of the ATC radar environment • Not unless the regulations are adapted to allow for a reduction of minimum operational training days for ATC based on type of simulator. • The answer is maybe. If the ATC simulator needed to be certified and hours in the simulator counted as OJT hours, similar to flight simulators, then a standardized categorization system is necessary. In the USA, that is not the case and simulation cannot substitute for training hours. Therefore a categorization system is not necessary.

It is clear from the sample comments provide by those who responded ‘No’ that there was an issue with the term ‘required’. Several responses indicated that they believed the categorization should not be required, but that it would still be a potentially useful tool for objectively differentiating between simulations and even potentially planning resource allocation during training design. Survey participants who responded ‘Yes’ were responsive to the issues identified in the analysis of Table 5.6 in terms of the lack of well-defined criteria, believing that a standardized categorization system would eliminate the need to rely on the terms low, medium, and high and allow people to speak in more “concrete terms”.

Given the feedback gathered from industry experts, there is clearly a desire for increased standardization when it comes to classifying and differentiating between simulations and that the use of terminology such as low, medium and high is likely insufficient to accomplish those needs. Responding to the second question from the outset of this section, how should the categories be presented, a formalized enroute ATC simulation environment categorization system, such as the FAA’s Full Flight Simulator categorization system, would fill this need and provide a significant improvement over the reliance on such vague fidelity terminology.

In response to that crucial third question posed at the outset of this section, a more objective approach is required to determine the different categories of enroute ATC simulation which is where the simulation fidelity definition could provide a greater amount of structure and objectiveness to the categories.

5.4 Chapter Summary

This chapter has identified the existence of differing representations of simulation fidelity between individuals within the ATC industry due to the consideration of different environmental components when making fidelity determinations. While vague terms used to describe the fidelity of simulations such as low, medium and high provide some indication of a simulation's level of fidelity, they lack well-defined criteria and are often determined subjectively making them insufficient for use as a classification tool. As noted in section 5.3, there is a strong desire for a more objective and standardized approach to differentiating between simulations. Taken together, these findings motivate the development of an objective definition of simulation fidelity for the domain of enroute ATC, as well as a standardized simulation environment categorization system based on the definition. *Chapter 6 An Enroute ATC Simulation Fidelity Definition* and *Chapter 7 An Enroute ATC Simulation Environment Categorization System*, present and discuss the development and application of these two constructs respectively.

Chapter 6

The Enroute ATC Simulation Fidelity Definition

This chapter presents the simulation fidelity definition, the primary research output of this thesis. The chapter is divided into four main sections: Development, The Enroute ATC Simulation Fidelity Definition, Validation, and Discussion. The first section outlines and explains the process and activities undertaken to create the definition from initial idea to final product. The definition is then presented with a detailed explanation of its different components. This is followed by the validation of the definition which is described and detailed in order to demonstrate how it was received by industry SMEs and whether or not any further work is required to refine it. Finally, a discussion section is included that elaborates on choices made during development of the construct and how the definition could be used.

6.1 Development

Based on the literature reviewed in Section *2.2 Understanding Simulation Fidelity*, an approach was identified that enabled the creation of an objective definition of simulation fidelity for the enroute ATC domain. This approach was initially presented and discussed in Section *3.1.1 Specific Approach Used to Define Fidelity for Enroute ATC*. The methodology consists of first determining the scope of the fidelity definition, then identifying the specific parts or components of the operational environment that fall under the scope, and finally validating those components via SME feedback. The validation process may require changes be made to components based on the SME feedback. Once completed, this process would yield an objective definition of what influences simulation fidelity in the domain of enroute ATC, validated by a group of industry professionals from a variety of backgrounds.

An initial simulation fidelity definition for enroute ATC was developed based on documentation regarding the current enroute ATC operational environment at ANSPs such as the FAA, Eurocontrol and Nav Canada and previous research experience in the enroute ATC domain. The structure of this initial definition was similar to the final version of the definition presented in the next section in that it identified a set of high-level component categories along with more specific environmental components. The work described in Chapter 4 where individuals were considering different elements of the operational environment when identifying the differences between simulation and the operational environment contributed significantly to this process of identifying the relevant environmental components. The components and their high level categories at this point required testing and refinement, and the definition was therefore presented for review to a small group of SMEs from the ANSP who were part of the project team as was noted in Chapter 3. The definition was then refined based on the feedback provided which produced the final iteration that is presented in the subsequent section.

A key decision when developing the simulation fidelity definition was determining how specific to make the components. For example, the component of "Control interfaces" can be further refined to keyboard, mouse, voice switch panel and headset. While this makes the components more specific, it would also increase the overall number of components thereby increasing the complexity of the definition itself. It also makes the definition less versatile as other ANSPs may have different versions of these items or not include them in their operational environment at all.

6.2 The Enroute ATC Simulation Fidelity Definition

The final version of the enroute ATC simulation fidelity definition is presented in Figure 6.1.

	Simulation Environment Components			
Category	**Physical Environment**	**Inter-Personal Communications**	**Simulation Functionality**	**Simulation Scenario**
Component	• Control interfaces • Main visual and auditory displays • Other information displays and/or tools • Physical environment	• Communication participants • Types of communication • Communication dynamics	• Time element • Aircraft performance • Operational uncertainty • Operating system functionality • Weather/ turbulence • Scenario capacity	• Scenario complexity • Part-task vs. whole-task • Working method implementation

Figure 6.1 The enroute ATC simulation fidelity definition.

The definition has three main categories of components that are inherent to the simulation environment: “Physical environment”, “Inter-personal communications”, and “Simulation functionality”. First, there is the “Physical environment” which represents all the physical characteristics of an enroute controllers work environment (e.g. visual, spatial, auditory, etc.). Table 6.1 provides a description of each of the components within this category. Taken together these are all the elements of the operational environment that a controller can physically interact with in some way, whether it be pushing a button, looking at a display or hearing an alarm. The components in this category capture the first half of the definition of simulation fidelity used as the foundation for the work in this thesis put forth in *Section 2.2 Understanding Simulation Fidelity*: “(1) the physical characteristics, for example, visual, spatial, kinesthetic, [auditory], etc.” (Hays and Singer, 1989). When an individual enters a simulation of the enroute ATC work environment, the fidelity of the components within this category are likely what an individual will notice first as most of these components are visible to the individual without having to run a scenario through the simulation.

Table 6.1 Description of the "Physical environment" fidelity components.

Component	Description
Control interfaces	Any interfaces that the controller interacts with that are absolutely necessary for the controller to perform their job (i.e. keyboard, mouse, voice switching interface, flight strips).
Main visual and auditory displays	The main visual and associated audio signals that arise from that display (i.e. main radar screen with subsequent auditory alarms).
Other information displays and tools	This component consists of all the non-essential information displays and tools that are provided to the controller (i.e. weather, sector map, etc.).
Physical surroundings	Any of the surrounding area outside of what is in the immediate vicinity of a controller's workstation (i.e. the large room, other controllers working, ambient noise, etc.).

The second category is "Inter-personal communications", which represents the communication between the controller and all the operators within the ATC system (e.g. pilots, other controllers, flight information specialists, etc.). Table 6.2 provides a description of each of the components in this category. This category is not explicitly represented within the simulation definition posited by Hays and Singer (1989), and while communications in other domains might be considered secondary to the primary task, they represent a very important part of the ATC task which is why it has been isolated in its own category here. Taken together these components capture who a controller might be talking to, what form of communication they are using to talk to that individual, and the dynamic nature of the conversation.

Table 6.2 Description of the "Inter-personal communications" fidelity components.

Component	Description
Communication participants	All of the operators who controllers could potentially communicate with in the ATC system (i.e. pilots, other controllers, flight service specialist, weather service, etc.).
Types of communication	The different modes of communication available to an active controller (i.e. voice over radio, text communication via CPDLC, gestural communication with surrounding controllers, operating system communications).
Communication dynamics	This component captures the fluid, dynamic nature of how controllers communicate with other actors in the ATC system (includes things such as delay in responses, garbled transmissions, communicating with multiple actors simultaneously, tasks of the actors being communicated with, etc.).

The third category is "Simulation functionality", which corresponds to the ability of a simulation to replicate the variety of stimuli or sensory information from the operational environment that an operator needs to monitor (e.g. how aircraft behave on the radar screen or shifting weather patterns) and the response options that are available to controllers to influence the system. Table 6.3 provides a description of each of the components in this category. The components in this category capture the second half of the definition of simulation fidelity used as the foundation for the work in this thesis put forth in *Section 2.2 Understanding Simulation Fidelity*: "(2) the functional characteristics, for example the informational, and stimulus and response options of the training situation" (Hays and Singer, 1989).

Table 6.3 Description of the "Simulation functionality" fidelity components.

Component	Description
Time element	This component represents whether or not the simulation is being run at real-time and if the user is being subjected to the time pressure of live operations.
Aircraft performance	This component captures how closely the aircraft performance characteristics in the simulation mirror those of the real life aircraft (this includes the variability of how different airlines and pilot may fly the aircraft).
Operational uncertainty	This component represents whether or not the simulation is capable of capturing off-nominal events that could happen in the real world given the operational environment's inherent unpredictability (e.g. an aircraft going left when told to go right)
Operating system functionality	This component captures how closely the tools and capabilities of those tools of the operational environment operating system are reproduced in the simulation environment.
Weather/turbulence	This component captures how accurate weather or turbulence phenomena are reproduced in the simulation environment.
Scenario capacity	This component represents the capability of a simulation to present a specific type of scenario to the user (e.g. a classroom-based simulation could only handle simpler scenarios).

The fourth fidelity component, the "Simulation scenario" created and presented to the user of the simulation, can also affect the fidelity of a simulation; however it is not an inherent component of a simulation environment. The scenario can change drastically depending on if it has been designed to focus on a specific skill for a new recruit, for recurrent training of an experienced controller, testing the viability of a new sector traffic flow dynamic, or for a researcher investigating workload and situation awareness. This is why the components within this category fall outside the scope of the simulation environment itself as indicated by the fidelity definition in Figure 6.1. The realism of the scenario can certainly affect the fidelity of the simulation which is why they are included in the definition, but it is important to differentiate between components that are part of the simulation environment itself and components that are modified depending on how the simulation environment is being used. The latter is where components in the "Simulation scenario" category fall.

Table 6.4 Description of the "Simulation scenario" fidelity components.

Component	Description
Scenario complexity	This component captures the inherent difficulty and volume of air traffic present in the scenario.
Part-task vs. whole-task	The scenario may only present certain elements of air traffic situations to the user (i.e. only presenting simple overtaking situations at a set altitude).
Working method implementation	This component captures the degree that a scenario is designed to elicit accepted controller strategies or approaches for handling air traffic situations.

Taken together, these categories and their associated components are thought to capture all key parts of the enroute ATC operational environment that affect the fidelity of a simulation of that environment. The next step in the development of an objective definition of simulation fidelity for enroute ATC as identified in the Section 6.1 is to validate the simulation fidelity definition via a review by industry SMEs.

6.3 Validation

This section details the validation of the enroute ATC simulation fidelity definition, which consisted of a two-step process. First, a preliminary small-scale validation activity was performed at the ANSP in order to determine whether or not the definition required any further refining. Following a successful initial validation, an industry wide validation was performed using the online survey as a delivery tool. The following two sub-sections discuss these two activities and a detailed explanation is provided regarding how they function together as a validation of the simulation fidelity definition.

6.3.1 Initial validation during interview sessions at ANSP

During the interview sessions at the ANSP, each interviewee was presented with the definition at the end of the interview in order to measure their acceptance of the components present in the definition. The definition was presented at the end of the interview session to avoid planting concepts and ideas within interviewees regarding simulation fidelity that could have affected responses to previous questions. The interviewee was asked the following question (as it appears at the end of the semi structured interview provided in Appendix A) after being presented with the definition as it appeared in Figure 6.1 and receiving a detailed explanation similar to the breakdown provide in Tables 6.1 to 6.4:

> "In your opinion, do these components of ATC simulation fidelity accurately represent the components of an ATC simulation that can affect the experienced level of fidelity by the user? Would you change any of these components?"

All 13 of those interviewed at the ANSP believed that this definition captured the relevant components when discussing simulation fidelity in the context of enroute ATC. One interviewee did note that stress was not included here, but operational stress falls under the psychological fidelity created by a simulation meaning it is inherent to the user of the simulation and not the simulation itself. As stated in section *2.2 Understanding Simulation Fidelity*, psychological fidelity components were removed from consideration in the development of this definition as only components inherent to the operational environment were considered in the development of this definition.

The acceptance of the definition by interviewees was positive and indicated that it had merit. However, limitations with this validation activity consisted of three main factors: (1) only 13 individuals were involved in the validation, (2) they all had training-oriented backgrounds, and (3) they all came from the same ANSP. Given these limitations, a larger scale validation effort was undertaken by including questions on the industry-wide survey to test the definition's comprehensiveness across a greater amount of individuals with a variety of professional backgrounds within the industry.

One component that was added shortly after this initial validation process at the ANSP was that of "Operational uncertainty" in the category of "Simulation functionality". This particular component was added when a reviewer of an academic paper based on this work discussed from their own experience in an ATC simulation the fact that controllers were noting the simulation was too "perfect". That is, none of the elements present in the simulation were inaccurate, but lots of little things happen in the real world that did not happen in the simulation. This inherent operational uncertainty is an important part of the operational environment and is one of the reasons why being a controller is so challenging, thus making this component an important addition to the definition. The "Operational uncertainty" component was then added to the definition for the larger scale validation effort that is discussed in the following sub-section.

6.3.2 Survey validation

The survey validation of the enroute ATC simulation fidelity definition consisted of presenting the definition to participants at it appears in Figure 6.1 and explaining its components. The survey participants were provided with a condensed version of Section 6.2 as a description of the fidelity definition. Once they had considered the definition and the accompanying explanation, the participants were asked to rate their agreement on a 7-point Likert-type scale with the following statement:

> "The four main components and their sub-components in the definition above accurately capture all the relevant components of the enroute ATC work environment that can affect the perceived fidelity of an enroute ATC simulation."

The results from the responses to this question are presented in Figure 6.2.

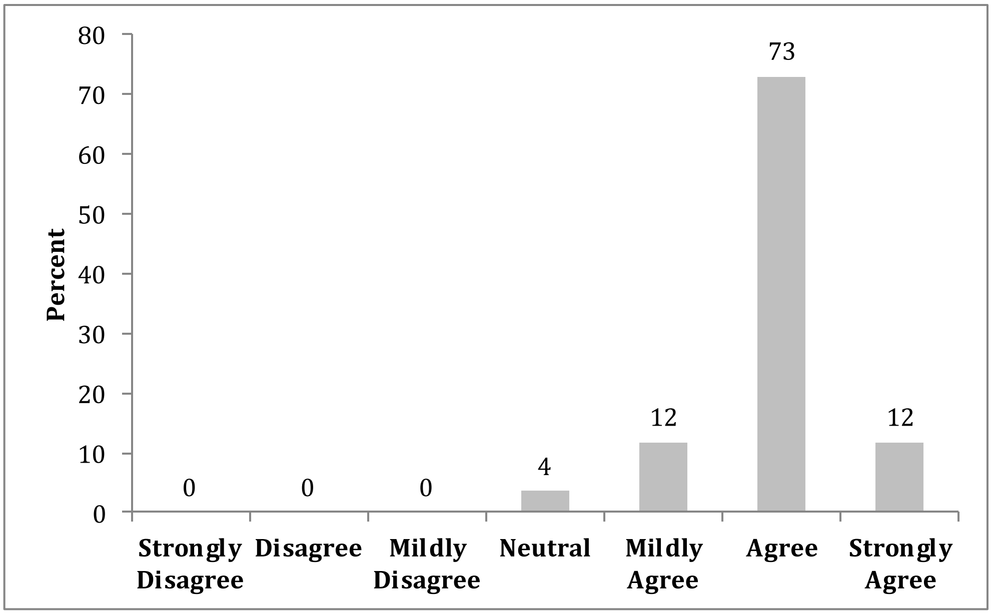

Figure 6.2 Level of survey participant agreement with the simulation fidelity definition as it was presented in Figure 6.1.(N=85)

Given the distribution of the observed response frequencies in Figure 6.2, it is clear that survey participants are in agreement with the simulation fidelity definition presented in Figure 6.1. Results of a chi-square goodness of fit test show that the observed response frequencies observed are statistically significant. They are significantly different than what would be expected if survey participants responded equally between all seven categories indicated in Figure 6.2 (χ^2 (6, N=85)=252.69, p<0.001).

Demographic data collected as part of the survey was used to investigate whether this level of agreement is wide-spread across gender, nationality, experience and primary use of simulation. A comparison between the response rates for these four demographics is presented in Figure 6.3, where the various "Disagree" and "Agree" categories were collapsed into single categories. As seen in the figure, the proportion of "Disagree"/"Neutral"/"Agree" responses shows a strong and consistent agreement that the simulation fidelity definition does capture all of the relevant components that can affect fidelity for enroute ATC. A chi square analysis was performed to determine if there were any differences within the demographic groups. It was found that there were no differences with regards to agreement with the definition when comparing within the demographic groups of gender, (χ^2 (1, N=84)=0.73, p=.392), nationality, (χ^2 (2, N=84)=1.89, p=.388), years working with simulation, (χ^2 (3, N=85)=2.94, p=.402), or survey participant's use of simulation, (χ^2 (2, N=85)=1.62, p=.445)[4].

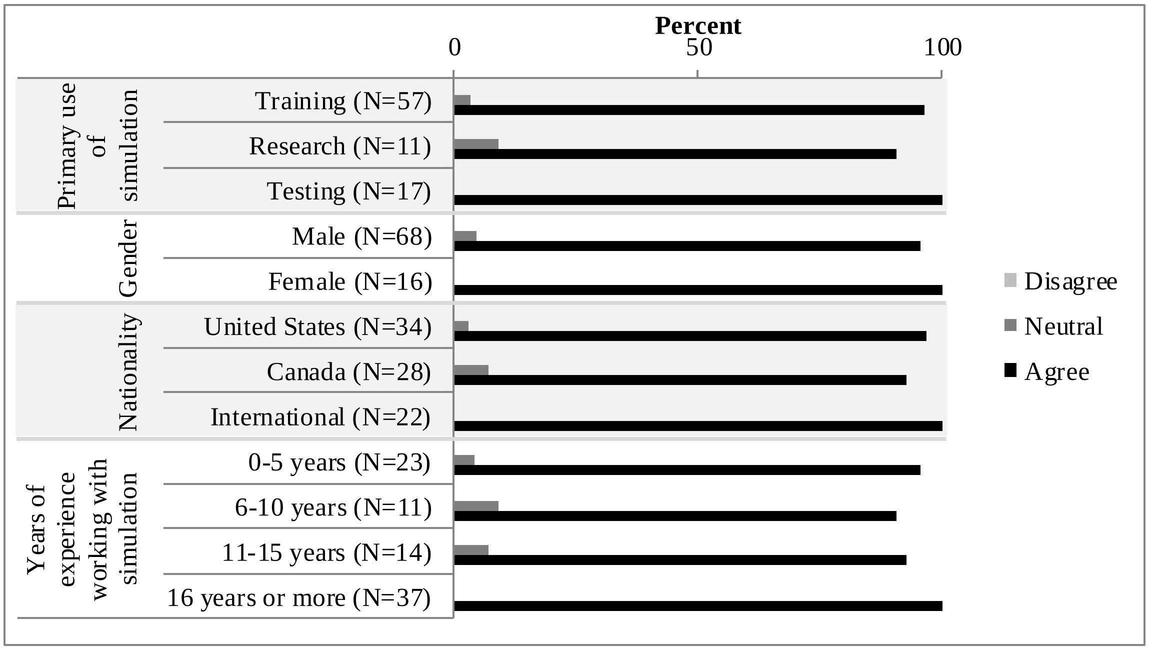

Figure 6.3 Demographic sub-group comparison regarding survey participant agreement with the simulation fidelity definition in Figure 6.1.

---

[4] N values are different for the demographic groups due to certain participants not responding to certain questions, as noted in Section 3.2.3. This discrepancy occurs for any following statistical analyses in this thesis.

This is a strong statement of acceptance by a sample of the ATC industry, as well as a strong indication that the definition is not organization specific but potentially valid to a larger variety of simulation users. The strength of the agreement on this question could cause concern about how the question was presented to participants; however, the consistency between this finding and the 100% acceptance rate when presenting the definition to the 13 SMEs at the ANSP suggests the definition has strong support.

In addition to the rating question, a follow-up question was included which asked participants if they would keep the definition the same as it was presented, or if they would make changes. Figure 6.4 presents the findings from this question.

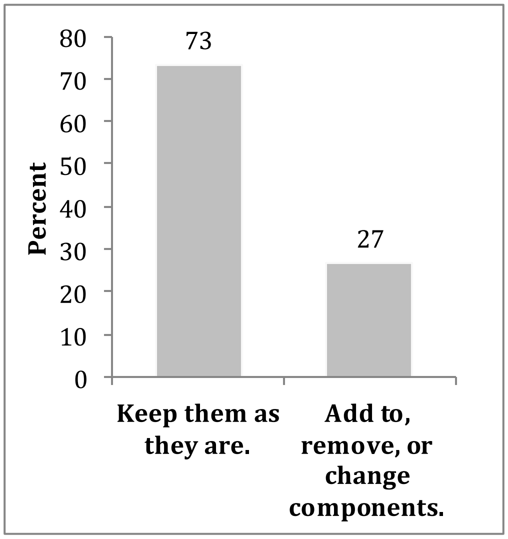

Figure 6.4 Survey participant opinion regarding whether or not they would make changes to the simulation fidelity definition as it is presented in Figure 6.1.

Results of a chi-square goodness of fit test show that the observed response frequencies are statistically significant. The “Keep them as they are” responses are significantly greater than what would be expected if half of survey participants wanted to change the definition in some way (χ^2 (1, N=86)=18.61, p<0.001). Demographic data collected as part of the survey was used to investigate whether this belief is wide-spread across gender, nationality, experience and primary use of simulation. A comparison between the response rates for these four demographics shows a strong and consistent agreement that the simulation fidelity definition does capture all of the relevant components that can affect fidelity for enroute ATC. A chi square analysis was performed to determine if there were any differences within the demographic groups. It was found that there were no differences with regards to agreement with the definition when comparing within the demographic groups of gender, (χ^2 (1, N=85)=0.04, p=.837), nationality, (χ^2 (2, N=85)=2.40, p=.302), years working with simulation, (χ^2 (3, N=86)=1.11, p=.775), or survey participant's use of simulation, (χ^2 (2, N=86)=2.26, p=.324).

Of those who did suggest changes to the definition, the two notable suggestions were: (1) whether certain components contained enough granularity to capture all the potential components of the operational environment that could affect fidelity, and (2) the inclusion of psychological components such as the cognitive fidelity of the simulation. These suggestions were considered, but the definition was not subsequently changed for the following reasons. For the first suggestion, as it was explained in Section 6.1, it is believed that the inclusion of too many components at too fine a degree of specificity would introduce unnecessary complexity into the definition while providing little value, and that the current list of components are at the appropriate level of detail given the acceptance of interviewees at the ANSP and the online survey participants in Figure 6.2.

In regards to the latter suggestion, it was demonstrated in *Section 2.2 Understanding Simulation Fidelity* that while the consideration of psychological elements of simulation is important, it should be considered separately from the concept of simulation fidelity as those are elements inherent to the users of the simulation rather than the simulation itself.

These findings indicate that the proposed simulation fidelity definition has the potential for wide acceptance and appears to be effectively capturing the different components of the operational environment believed to affect the fidelity of an enroute ATC simulation.

6.4 Discussion

Developing a simulation fidelity definition is a process of continual iteration and refinement with regards to identifying the specific fidelity components of the operational environment that are believed to affect fidelity. Outlining the scope of the fidelity definition is an important first step in this process as knowing what will not be included or considered can be just as important as what will be considered. It is also important to understand how the profession itself works as that will help identify some of the broader categories of components. As was noted in the previous section, the categories of "Physical environment" and "Simulation functionality" arose from the Hays and Singer (1989) definition. The category of "Inter-personal communications", however, was not explicitly part of that definition but given how important communication is for a controller it was separated into its own category. It could be argued that it belongs in its own category for most other domains as well as there is rarely a profession focused solely on the actions of a single individual, especially when considering complex, sociotechnical systems.

Once the categories had been determined, identifying the specific environmental components became more of an issue of how specific to make the components, as explained at the end of Section 6.1. As was noted in Section 6.1, the components in Figure 6.1 were identified at this particular level of specificity as it allowed the definition to be usable across a wider variety of potential users. It also kept the definition from becoming overly complex and concerned with the minutiae of the operational environment. While it would not be wrong to include very specific components, it could make the definition less user-friendly and approachable. Given that this exercise is focused on developing a better understanding of simulation fidelity within the industry, it would be counter-intuitive to develop a definition that was not easy to understand and apply by the greatest number of users as possible.

One of the most important lessons learned in the development of this simulation fidelity definition was the SME review process. Buy-in from the ATC industry at large not only in the specific components themselves but the value of the construct itself is crucial to its success as a tool for better understanding how simulation fidelity potentially affects simulation use in training, testing new operational concepts, and future ATC environment research. If it is viewed as incomplete, flawed or just not useful, then the definition becomes superfluous and will have no impact on simulation use. But the process of including those who will use the definition or be affected by its use during its development was key in gaining wide-ranging acceptance and highlights the importance of including individuals in the design process of concepts that will eventually affect their work. This is also an important relationship to foster given the iterative nature of developing a simulation fidelity definition. Gathering feedback from SMEs along the way, such as that described in Section 6.1, helps to streamline and expedite the development process considerably. This also serves to improve the industry's acceptance from those who were not included in the development process, as the fact that it was developed with input from colleagues will help acceptance and belief in its viability.

One of the limitations with the definition itself is that it does not provide weights regarding the value of particular components. This is likely an avenue for further research as initial discussion seems to indicate that the value of a particular component is dependent on how the simulation is going to be used. For instance, if a trainee is learning phraseology and basic communications skills, the value of communications component would be high but perhaps components within the category of "Physical environment" would be less so. There are countless scenarios that could be imagined where various components are valued differently, and while no work has yet taken place to identify the weighting of fidelity components in different situations, one of the most important aspects of the definition is that it at least facilitates and encourages a discussion along these lines rather than working on assumptions.

There are limitations to the approach taken to validate the simulation fidelity definition via the online survey. Presenting such a complex, abstract construct such as the simulation fidelity definition within an online survey and asking individuals to then make a considered judgment regarding its viability as a construct is admittedly not a perfect scenario. It is not possible to establish exactly how deeply each survey participant considered the definition and its explanation as well as how clear the construct actually was to them. This finding, presented in Figure 6.2, was not considered in isolation as survey participants were also asked to explain their responses regarding whether they felt the definition captured the relevant components by indicating in the following question whether or not they would keep the definition the same or make changes, data presented in Figure 6.4. In addition, the results from the survey are an extension of the preliminary validation conducted with interviewees at the ANSP where the researcher was able to more clearly establish understanding of the definition and where response to the definition was similarly strong. While not perfect, the online survey was identified as the most appropriate method of validating the simulation fidelity definition on a larger scale.

Establishing the definition provides a basis for broader agreement on applications such as evaluating controller's previous training for the purposes of facility transfers, inter-ANSP transfers and/or previous exposure to relevant procedures, weather conditions, or traffic levels. Simulation has the potential to replace some or all of a controller's background material, increasing staffing flexibility. Given this, a standardized and objective understanding of what components affect the fidelity of enroute ATC simulation can only improve understanding of how best to apply simulation in these various contexts.

However, the enroute ATC simulation fidelity definition does not directly compare or differentiate the fidelity of simulations. What this definition provides is an objective, SME-validated list of the different components that affect simulation fidelity in the domain of enroute ATC. The components are essentially the points of comparison between simulations. Operationalizing the definition would allow for industry-wide convergence on training standards, simulation development standards, and broader sharing and acceptance of the results of procedure and operational concept testing. The effort to operationalize the definition involves developing a categorization system using the definition as the foundation for making comparisons and establishing differences between the different simulations.

6.5 Chapter Summary

The enroute ATC simulation fidelity definition presented in Figure 6.1 offers an objectively defined set of components that can affect the fidelity of an enroute ATC simulation. With its main categories of "Physical Environment", "Inter-Personal Communications", "Simulation Functionality", and "Simulation Scenario" and each of the more specific fidelity components, this definition represents a consistent understanding of simulation fidelity within the domain. This definition will hopefully help to eliminate the variety of simulation fidelity representations that were discussed in sections 5.1 and 5.2 while providing individuals with a standardized set of components as a reference for simulation fidelity in enroute ATC.

As was noted in the Discussion section of this chapter, however, the definition itself does not differentiate between the fidelity of simulations. The components within the simulation fidelity definition provide the points of comparison when comparing the fidelity of simulations. Chapter 7 presents a categorization system for enroute ATC simulation environments while using the simulation fidelity definition from Figure 6.1 as the underlying framework for the categorization system.

Chapter 7

Enroute ATC Simulation Environment Categorization System

This chapter presents the enroute ATC simulation environment categorization system through four main sections: development, a detailed explanation of the categorization system, the work done to validate the system and a discussion section. The development details the process and activities undertaken to create the categorization system based on the interviews and pre-existing categorization systems in other domains. The categorization system is then presented with a detailed explanation of its different levels, what the purpose of the construct is and how it could be used. Following this, the validation of the categorization system is described and detailed in order to demonstrate how it was received by industry SMEs and whether or not any further work is required to refine it. Finally, a discussion section is provided that elaborates on the challenges with developing the categorization system and what the next steps are in its development.

7.1 Development

The belief that an operationalized version of the fidelity definition developed in the previous chapter would be useful for the ATC domain arose from the analysis of other domains where simulation is relied upon so heavily. Traditionally, the categorization systems used in other domains have focused solely on the fidelity definition components associated with the simulation environment (e.g. "Physical environment", "Inter-personal communications", and "Simulation functionality"), excluding from considerations those components associated with how the simulation is being used. In parallel with this tradition, the focus of the enroute ATC categorization system developed in this chapter was narrowed to reflect only the components of the fidelity definition that are inherent to the simulation environment (see Figure 6.1 in Section 6.2). In addition, to reflect this narrower focus, the term “simulation environment categorization system” is used throughout this chapter to indicate that the categorization does not reflect the “Simulation scenario” components or any other aspect of how a particular simulation is being used.

The use of a categorization system for simulation environments in both the flight and marine simulation domains, as documented in *Section 2.3 Examples of Classifying Simulation Environments in the Literature*, influenced significantly the structure and approach to classifying simulation environments for enroute ATC. In those domains they require clear standards regarding the different levels of simulation fidelity available for training purposes due to the fact that simulation is relied on so heavily to achieve so many different objectives within the training of new and current operators.

These categorization systems, as described in Section 2.3, are not overly specific but offer clear delineations between simulation environments of varying degrees of fidelity and offer a quick glimpse as to what their capabilities are. What is clear from all of these systems is that they are trying to present a picture of the different simulation environments available and thereby help users determine which is best for their needs. However, a challenge when using these categorization systems is that it is not explicitly or objectively stated what each level of fidelity is best suited to accomplish. The decision of how to use the various levels of simulation environment fidelity then typically falls to the user based primarily on subjective reasoning. This is due to the limited amount of research that exists regarding the specific impact of varying degrees of fidelity fidelity on training, as noted in *Section 2.4 Impact of Simulation Fidelity on Training*.

In order to develop a categorization system for enroute ATC, however, a deeper understanding of the current simulation tools available within the industry was required. This understanding was achieved by the exploration of simulations used at an ANSP, as was detailed in *Section 4.1 Simulation Capabilities for Enroute Controller Training at an ANSP*. This process helped to clearly identify different simulation environments that currently existed and the changes in fidelity typical from one simulation to the next. A categorization system that captured the differences between these existing tools was desirable. Therefore, during the interviews at the ANSP, interviewees were presented with the full flight simulator classification system developed by the FAA, presented in Table 2.2 in Section 2.3, and were then asked what a categorization system such as the one presented to them would look like but for enroute ATC simulations. A hierarchy of simulation environments began to form based on interviewee responses to that question. From these responses, a categorization system was developed based on the structure and language of the FAA's flight simulator classification systems, using key components from the simulation fidelity definition from Section 6.2 as the primary points of comparisons from one level to the next.

7.2 The Enroute ATC Simulation Environment Categorization System

The proposed enroute ATC simulation environment categorization system, as it was presented to participants during the online survey, is presented in Table 7.1.

Table 7.1 Propose enroute ATC simulation environment categorization system.

Level	Description
A	A static scenario is presented through description, drawings or pictures, **with no** ability to directly control the system. No physical environment requirements. Can include simple information sources (i.e. flight strips). Communications are executed conversationally.
B	A dynamic scenario is presented through the use of a television, projector or smartboard, **with no** ability to directly control the system. Must have a main display consistent to that of operational environment. Can include simple information sources (i.e. strips). Communications are executed conversationally and/or via pre-recorded transmissions.
C	A dynamic scenario is presented through the use of a personal-computer or laptop program **with** the ability to directly control the system. Must have main display consistent to that of operational environment with some of the OS functionality. Can include simple information sources (i.e. flight strips). Communications are executed by voice recognition software and/or conversationally with the use of radio.
D	A dynamic scenario is presented through the use of a simulated workstation **with** the ability to directly control the system. Must have main display identical to that of operational environment with all of the OS functionality. Should include all information sources necessary to do the job. Communications are executed with a single simulation specialist with the use of radio.
E	A dynamic scenario is presented through the use of a simulated workstation **with** the ability to directly control the system. Scenario is integrated with other users who can also control the system. Must have main display identical to that of operational environment with all of the OS functionality. Should include all information sources necessary to do the job. Communications are executed with between users of the simulation and pseudo-pilots.

The levels progress gradually in degree of fidelity from two different types of classroom-based simulation environment to a PC-based simulation environment to a workstation simulation environment and finally to a highly integrated, multi-participant simulation environment environment. The reason for including two distinct types of classroom-based simulation environments was due to the fact that there were two distinct ways in presenting this type of simulation based on visual stimuli. There was either a static representation of an air traffic scenario presented orally or visually, or a dynamic representation of an air traffic scenario via a recorded radar screen presented on a screen. These different ways of presenting scenarios activate different cognitive processes in a user of the simulation as the presence of a radar screen and the live, moving traffic can cause reliance on the screen information rather than internal visualization techniques and strategies. It is believed that this is a significant enough difference in the realism of a simulation environment to merit two separate simulation categories.

The categorization system in Table 7.1 does not include all of the fidelity components from the simulation fidelity definition in Figure 6.1. The primary goals in terms of the structure of this categorization system were ease of use and accessibility to a wide range of potential users. This necessitated keeping the categorization system as simple as possible and being more judicious about how the components were integrated into the system. Thus the categorization system was developed along a single dimension of changing fidelity, from A to E. This was due to the fact that if each component from the fidelity definition had its own level of fidelity, the system would quickly become combinatorially explosive. This was also done to be consistent with the structure used by categorization systems from other domains. The descriptions of the categories in Table 7.1 summarize the key components from the fidelity definition that determined what makes one environment more realistic than another. The determination regarding which components were included in the description of the categories was based off SME input during the interviews at the ANSP and the researcher’s discretion.

A clear effort was made to use consistent, straightforward language to describe each level of simulation fidelity. This allows those reading these levels to be able to easily compare one to the next and to easily see how the components in the description change from one level to the next. Also, this is a generic categorization system and not one directly attached to any simulators currently in use by any particular ANSP or research institution. This was done due to the fact that simulation technology is often being upgraded or changing and it was important that these levels not be directly attached to a specific existing simulation environment in order to remain as current as possible. It also makes the categorization system more useful to a wider audience of potential users. These levels are meant as categories where simulation environments that display the characteristics of a particular level can subsequently be categorized under that level.

During the development of the categorization system in Table 7.1, a more detailed system was created to help more clearly identify the characteristics of simulation environments at each level of fidelity and how they change from one level to the next. This more detailed system is presented in Table 7.2. This system provides an example of a simulation environment at each level of fidelity, the relative cost of each simulation environment and how each component from the fidelity definition changes from one level to the next across the three categories inherent to the simulation environment from the simulation fidelity definition in Figure 6.1. This highly detailed representation of each level of fidelity was meant primarily as a design aide, as it was believed to be overly complicated for more day-to-day use.

Table 7.2 Highly detailed preliminary framework for the proposed enroute ATC categorization system in Table 7.1.

Category	Component	Level A	Level B	Level C	Level D	Level E
Quick Description		Case study or Role playing	Dynamic case study	PC-Based Trainer	Simulated Workstation	Multi-User Operations Simulation
Relative Estimated Cost		$	$$	$$$	$$$$	$$$$$
Physical Environment	Control Interfaces	None	None	Keyboard, Mouse, Headset	Keyboard, Mouse, Headset, Comms. Panel	Keyboard, Mouse, Headset, Comms. Panel
	Main Visual & Auditory Displays	None	Similar radar display	Similar radar display	Identical display with full functionality	Identical display with full functionality
	Other Information Displays/Tools	Strips, Databoard	Strips, Databoard, Simple tools	Strips, Databoard, Simple tools	Strips, Databoard, essential tools	All extraneous information sources
	Physical Surroundings	Classroom or experiment room	Classroom or experiment room	Classroom or experiment room	Simulation room, single workstation	En-route operations room replica
Inter-Personal Communication	Communication participants (simulated by)	Controllers, pilots (Instructor, other students)	Controllers, pilots (Instructor, other students)	Controllers, pilots (Voice-rec., other students, instructors)	Controllers, pilots (Simulation specialist, pseudo-pilot)	All (Simulation specialist, other students, pseudo-pilots)
	Types of Communication	Conversational	Conversational	Simulated radio with voice recognition software or automated text-based comms.	Simulated radio comms. with 1 person	Simulated radio comms. with multiple people
	Communication dynamics	Capable of a variety of communication scenarios	Capable of a variety of communication scenarios	Limited to automated responses	Limited to one person acting as all participants	Multi-participant communication dynamic

Component	**Sub-Comp.**	**Level A**	**Level B**	**Level C**	**Level D**	**Level E**
Quick Description		Case study or Role playing	Dynamic case study	PC-Based Trainer	Simulated Workstation	Multi-User Operations Simulation
Simulation Functionality	Time Element	Clock off or Clock on	Clock off or Clock on	Clock on (slow-time or pause)	Clock on (slow-time or pause)	Clock-on
	Aircraft (AC) Performance	AC are presented in a static representation	AC performance consistent, variability not required	AC performance consistent, variability possible	AC performance very close, variability possible	AC performance is identical, including variability
	Operational uncertainty	Limited uncertainty capabilities	Limited uncertainty capabilities	Limited uncertainty capabilities	Moderate uncertainty capabilities	Full operational uncertainty capable
	Operating system functionality	None	None	Main functions replicated	Full functionality replicated	Full functionality replicated
	Weather/ Turbulence	Static weather systems	Dynamic weather systems	Static weather systems	Dynamic weather systems	Dynamic weather systems

7.3 Validation

Similar to the fidelity definition developed in Chapter 6, the categorization system underwent two validation activities. First, the categorization system was included in the online survey distributed to the ATC industry at large where a usability exercise was provided to determine its usefulness and whether or not survey participants would propose any changes. The second activity assessed the usability of the simulation fidelity categorization system as a tool by using it to analyze previously published ATC research. Published studies that have used ATC simulations over the previous 10 years were classified using the proposed simulation categories. The results were then examined in order to further understand and document current practices for selecting appropriate simulation fidelity and the usefulness of the categorization system.

7.3.1 Survey Validation

The primary validation activity for the simulation environment categorization system took place via the online survey. This consisted of asking survey participants to use the categorization system to indicate a minimum level of fidelity required to accomplish certain outcomes for training, testing and research.[5] After this exercise, they were then asked how useful they found the categorization system in determining the level of fidelity they felt was required for each particular activity. Figure 7.1 presents the findings regarding the usefulness of the categorization system as rated by survey participants.

---

[5] The results of this question are presented and discussed in the next chapter in *Section 8.2 Industry-wide Perceptions Regarding Simulation Selection*.

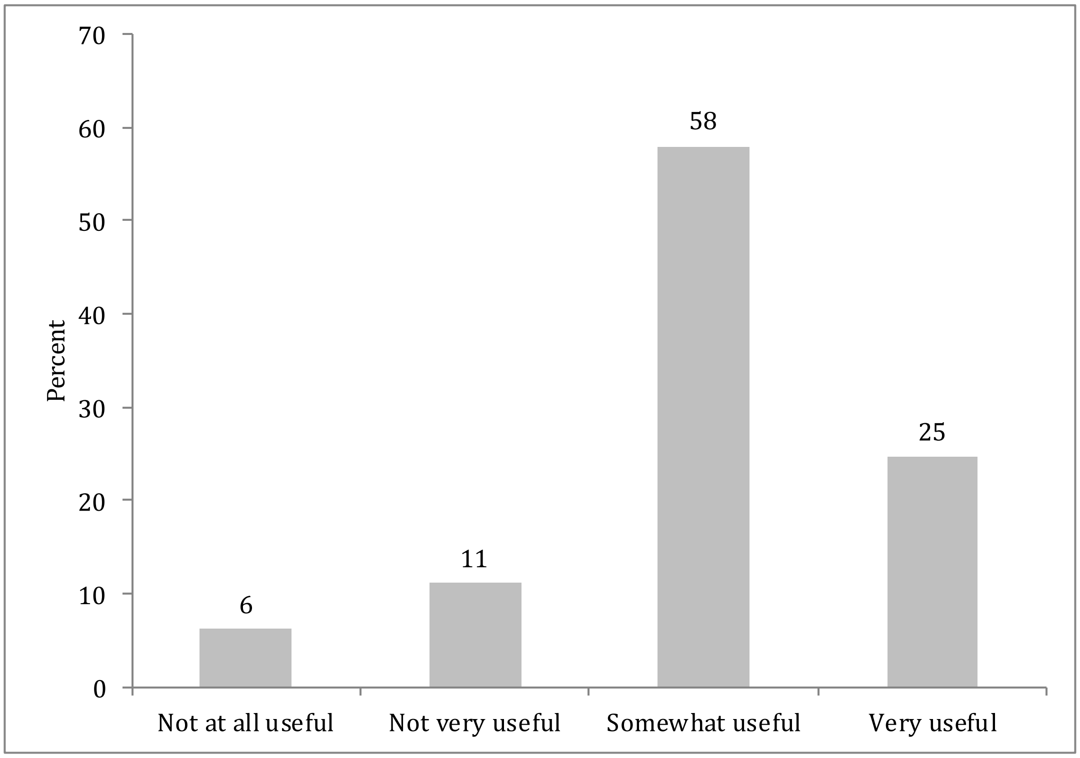

Figure 7.1 Survey participant beliefs regarding the usefulness of the categorization system during the usability exercise. (N=81)

Results of a chi-square goodness of fit test show that the observed response frequencies are statistically significant. They are significantly different than what would be expected if responses were equally distributed between all four categories indicated in Figure 6.2 (χ^2 (3, N=81)=49.49, p<0.001). Based on the distribution of the response frequencies in Figure 7.1, it is clear that most survey participants found the categorization system at least somewhat useful.

Demographic data collected as part of the survey was used to investigate whether this level of agreement is wide-spread across gender, nationality, experience and primary use of simulation. A comparison between the response rates for these four demographics is presented in Figure 7.2, where the two categories of useful responses and the two categories of not useful responses were collapsed into singular categories of 'Useful' and 'Not useful'. As seen in the figure, the proportion of "Useful/"Not useful" responses shows a strong and consistent agreement that the simulation environment categorization system is a useful construct.

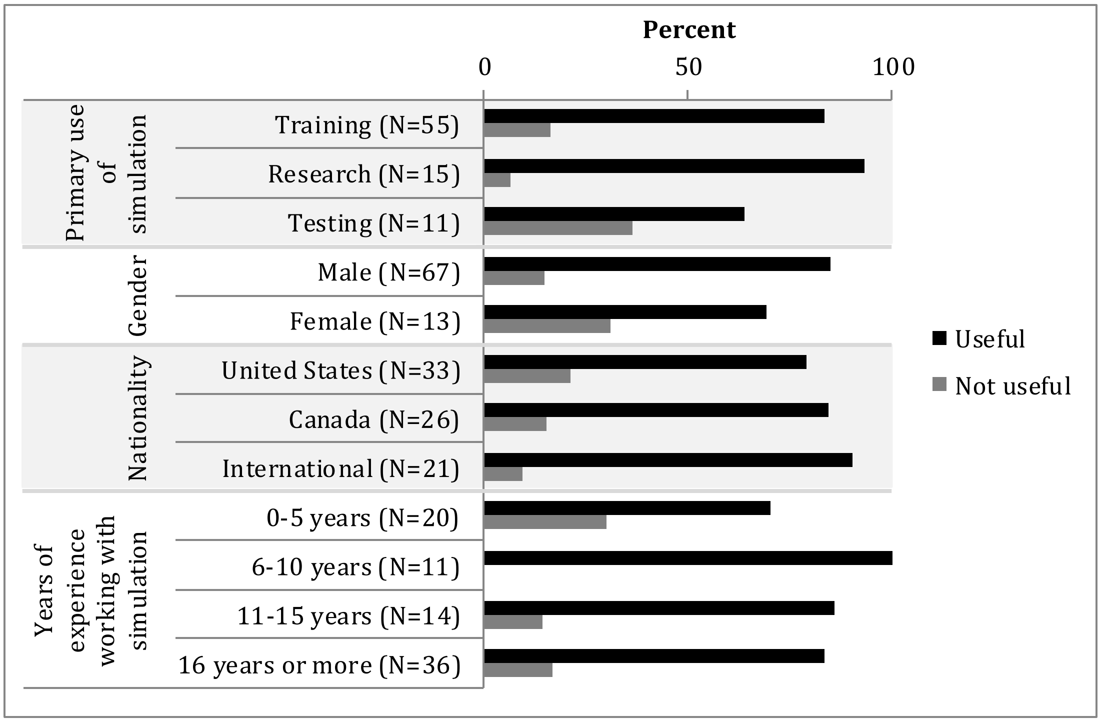

Figure 7.2 Demographic sub-group comparison regarding survey participant agreement with the simulation fidelity definition in Figure 6.1.

A chi square analysis was performed to determine if there were any differences within the demographic groups. It was found that there were no differences with regards survey participant belief regarding the usefulness of the categorization system when comparing within the demographic groups of gender, (χ^2 (1, N=80)=1.89, p=.169), nationality, (χ^2 (2, N=80)=1.31, p=.520), years working with simulation, (χ^2 (3, N=81)=4.66, p=.199), or survey participant's use of simulation, (χ^2 (2, N=81)=4.02, p=.134). This is a significant statement of the categorization system's usefulness by a diverse sample of the ATC industry. Those who felt it was useful believed it provided context to the different levels of simulation available and the potential of standardization they believed it offered. Sample comments from survey participants demonstrating this position are provided in Table 7.3. These comments clearly illustrate the value a system such as the one put forth in Table 7.1 could provide to those who work with simulation on a routine basis.

Table 7.3 Sample comments from survey participants regarding why they feel the simulation environment categorization system in Table 7.1 is useful.

Sample Comments
• "It provided context." • "It helped me to determine what was available so I could compare what I believe should be available to get the best value out of training a student for a given task." • "Assuming that the industry would embrace this concept it lends itself to a standardization that has been lacking in the ATC side of the business since the beginning." • "It helped to organize my thinking about variations in simulation fidelity."

Comments from those who felt it was not useful illustrate some of the potential drawbacks of the categorization system in Table 7.1, with one key issue being that there is still some grey area where simulation environments fall across more than one category. A representative comment was offered by one participant, stating that: "Some of the categories are so similar that the current systems fit into more than one/current systems are not accurately described by the categories." This is a key challenge which a standardized system must address as it should be clear when presented with a simulation environment where it falls on the scale.

Following on to this point, the final part of the survey validation asked whether participants would change the system or keep it as it is. A chi square goodness of fit test performed on the results of this question, presented in Figure 7.3, indicated the response frequencies were significant. The "Keep it the same" responses are significantly greater than what would be expected if half of survey participants wanted to change the categorization system in some way (χ^2 (1, N=78)=11.54, p<0.001).

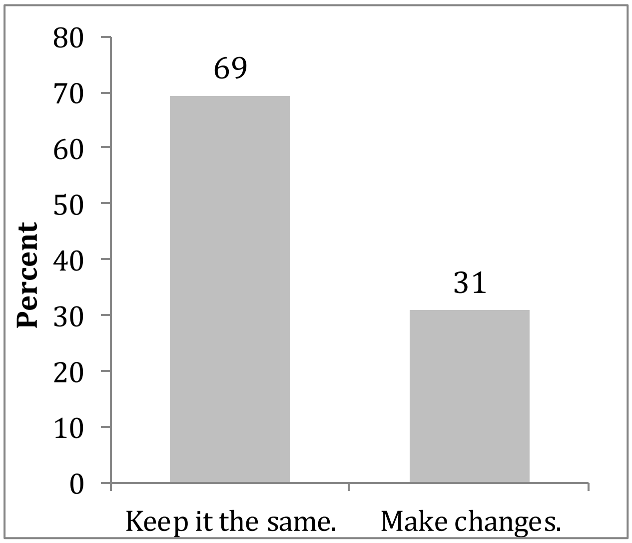

Figure 7.3 Survey participant responses to the question, "Would you keep the categorization system the same or make changes?" (N=78)

Demographic data collected as part of the survey was used to investigate whether this belief is wide-spread across gender, nationality, experience and primary use of simulation. A comparison between the response rates for these four demographics shows a consistent agreement that the categorization system is acceptable in its current format in Table 7.1. A chi square analysis was performed to determine if there were any differences within the demographic groups. It was found that there were no differences with regards to acceptance of the categorization system in its current format when comparing within the demographic groups of gender, (χ^2 (1, N=77)=0.39, p=.533), nationality, (χ^2 (2, N=77)=0.96, p=.620), years working with simulation, (χ^2 (3, N=78)=1.03, p=.795), or survey participant's use of simulation, (χ^2 (2, N=78)=1.52, p=.467).

However, a similar theme to the comments from Figure 7.1 was found within the comments of survey participants who would make changes to the categorization system, with one representative comment indicating that "… in the description [of the levels of fidelity] there are not enough elements to distinguish the fidelity." This once again indicates the possibility of simulation environments falling under more than one category.

The data presented in this sub-section offers potential evidence for the categorization system being acceptable in its current form given survey participant responses in Figure 7.1, 7.2, and 7.3. There was, however, a key concern raised in the comments of Figures 7.1 and 7.3

regarding the suitability of the categories in Table 7.1 being able to accurately provide a category for all enroute ATC simulation environments. To further examine the viability of the enroute ATC categorization system, a follow-on validation exercise was conducted to test the system in a more applied setting.

7.3.2 Analysis of Literature using the Categorization System

The secondary validation exercise for the simulation environment categorization system consisted of using the system as an analysis tool for identifying patterns with regards to simulation use in enroute ATC studies. This served as an application exercise that would help to identify if the categorization was effective at differentiating between simulation environments described in the literature, as well as potentially investigating trends regarding the use of simulation of varying levels of fidelity in enroute ATC research.

The criterion for inclusion in this exercise was that the study or article had to use an enroute ATC simulation in some manner to accomplish its objectives. 78 such articles were identified, with publication dates between 2001 and 2014. These papers were taken from a variety of publication sources, from Human Factors to the Journal of Applied Psychology. The full list of papers used for this exercise can be found in Appendix E.

Once the papers had been assembled, they were then broken down into the following component parts:

- Date,
- Periodical,
- Subject matter keywords,
- Fidelity descriptor of simulation used by author,

- Level of detail in the author's description of the simulation environment,
- The corresponding level of simulation fidelity from the categorization system in Table 7.1, and, if possible,
- The name/origin of the simulation environment.

Once these components had been identified or determined for each paper, the results would be analyzed to identify any patterns or trends regarding simulation use for these enroute ATC studies.

The key part of this validation activity was using the categorization system to assign the simulation environment in the paper a particular level of fidelity. Theoretically this should be a straightforward exercise, as it requires only reading the description of the simulation environment provided by the author and assigning it the appropriate category of fidelity based on Table 7.1. There were two issues that made this a challenging exercise: (1) the lack of a detailed description provided by the author, and (2) there were several instances where the simulation environment described by the author did not fit into one specific category within the categorization system. A detailed discussion of the second issues is provided in Section 7.4.1.

Table 7.4 presents the overall number of times each level of simulation fidelity from Table 7.1 was used in a paper. Based on this table, Level C and D simulation fidelity were clearly relied upon to conduct the bulk of research, with almost 75% of papers relying on a simulation environment at one of these two levels of fidelity. Further, Level C was used by at least half of all papers gathered. This is likely due to Level C simulation fidelity being a middle ground between affordable as well as a dynamic simulation environment in which to conduct research studies.

Table 7.4 Overall percentage of each level of simulation fidelity's use within the gathered papers. (N=78)

Level of simulation fidelity	Quick description of level of fidelity	Percentage of papers using simulation[6]
A	Case study or role playing	8 %
B	Dynamic case study	13 %
C	PC-based Trainer	51 %
D	Simulated Workstation	23 %
E	Multi-User Operations Simulation	9 %

To further explore simulation use within these papers, the top ten areas of research were identified based on the keywords assigned to each paper. This was followed by determining the distribution of the level of simulation fidelity used to research each topic area. The results of this analysis are presented in Table 7.5. The pattern identified in Table 7.4 regarding a reliance on Levels C and D is also apparent within this table. Also, the fact that the highest frequency of keywords was almost 25% and most were well below this mark indicate that there is a great variety of topics being investigated with simulation within the domain of enroute ATC. There is a notable lack of Levels A and B simulation fidelity being used to accomplish research on these top ten areas of research. These levels of simulation fidelity were primarily used to research conflict detection, expertise and complexity.

---

[6] Note: The percentages do not add up to 100%. This is due to several instances where papers reported multiple studies or experiments and where more than one simulation environment was used.

Table 7.5 Top ten keywords identified in paper database with corresponding distribution of levels of simulation fidelity used to research keyword.

Keywords	Percent (N=78)	Level of simulation fidelity used to research keyword (% of keyword N)				
		A	B	C	D	E
Automation (N=19)	24	0	0	68	16	16
Performance (N=19)	24	0	32	47	11	11
Workload(N=16)	21	6	0	63	6	25
Future ATM (N=12)	15	0	0	75	8	17
Training (N=12)	15	0	17	42	42	0
Situation awareness (N=7)	9	0	0	57	14	29
Memory (N=6)	8	0	17	83	0	0
Decision making (N=6)	8	0	33	33	33	0
Assessment (N=6)	8	0	17	0	50	33
Modelling (N=5)	6	0	20	60	20	0
Mixed equipage (N=5)	6	0	0	40	60	0

Finally, the number of times a fidelity descriptor was used to describe a simulation environment, such as "low, medium, or high", was also tracked. Results of this analysis are presented in Table 7.6. These results indicate that only a third of the time did an author use a descriptive term of the fidelity of the simulation environment. It is interesting that the majority of survey participants believe that these terms are useful, as noted by Figure 5.5, yet it appears those who actually use these terms are in the minority. This could be due to the fact that many are aware of the shortcomings with these terms for describing the fidelity of a simulation environment in their current state, as discussed in Section 5.3. It is possible a more standardized approach to categorizing fidelity, such as the one proposed in Table 7.1, may increase the reporting of the specific fidelity of the simulation environment being used.

Table 7.6 Percentage of times a fidelity descriptor was used to describe a simulation environment. (N=80)

Fidelity Descriptor	Percent
No fidelity term used	67
"Low, medium, or high" used	33

7.4 Discussion

The following discussion section is separated into three sub-sections. First, choices made and lessons learned during the development of the categorization are discussed. Following this, a discussion is presented regarding the use of the simulation environment categorization system as described in Section 7.3.2 and a different approach to structuring the categorization system is also discussed. Finally, a sub-section is provided that discusses the potential implications of using a categorization system within the ATC industry.

7.4.1 Developing the Categorization System

A significant challenge associated with designing the categorization system is determining the number of levels, as the researcher must strive to determine what a significant and meaningful difference is between levels of fidelity. As an example, the FAA has a two tiered simulation fidelity categorization system with several levels of simulation fidelity for both lower fidelity flight training devices (FTD) and the higher fidelity full flight simulators (FFS). This approach makes sense for flight simulation categorizing as there is a clear line to delineate between FTDs and FFSs, which is the ability to replicate the motion of flying. Given that enroute ATC simulation environments do not have such a clear delineation point to necessitate a two tiered system such as this, the categorization system presented in Table 7.1 contains the entire spectrum of simulation fidelity levels for the domain of enroute ATC.

The five levels presented in Table 7.1 were based off input from interviewees as well as analyzing the types of simulators used by other ANSPs and in future ATC environment academic research. The fidelity definition was used to differentiate between potential levels of simulation fidelity by identifying differences in some of the key components as described by interviewees and using these components as the primary points of comparison between levels of fidelity (e.g.

Control Interfaces, Main Visual Display, Communication Dynamics, Operating System Functionality).

As was noted in Section 7.3.2, and from survey participant comments in Section 7.3.1, the primary issue that made choosing the appropriate level of fidelity for a particular simulation environment more challenging was the existence of gray areas where a simulation environment did not fit one particular category. There were several instances where the simulation environment described did not fit into one specific category when using the categorization system during the analysis of papers. Based on the description of the simulation environment, certain aspects of the simulation environment could be categorized at one level while other components could be categorized at a different level of fidelity. This introduced a certain amount of subjectivity into determining the level of simulation fidelity that best fit the simulation environment as a decision had to be made based on which particular components were at a different level of simulation fidelity.

As identified in *Section 2.3 Examples of Classifying Simulation Environments in the Literature*, one of the biggest challenges associated with developing a categorization system is how much detail to provide in the description of each category. Too much information and it becomes overly complex and likely less easy to use, whereas too little and the categories become vague and hard to distinguish one level from the next. Based on the validation exercise described in Section 7.3.2, the levels do have enough detail but are too restrictive in their requirements of having components at a particular level of fidelity to be eligible for that category. For instance, a simulation environment could have certain components that perhaps warranted being rated as a Level D, but other components that only merited a Level B.

The proposed categorization system did not include any reference to the “Simulation scenario” components as it is focused solely on the simulation environment. Yet, as evidenced by its inclusion in the fidelity definition from Section 6.2, the scenario used in the simulation does affect the overall fidelity of a simulation and the categorization system could be expanded to reflect these components in some manner.

However, trying to fit the “Simulation scenario” components into the proposed categorization system in Table 7.1 would likely make the system even more difficult to use than it already is. Measuring the three categories of fidelity components inherent to the simulation environment across the five levels of simulation fidelity is already a challenge given the restrictiveness inherent to each level, as was evident in the application exercise described in Section 7.3.2. Adding the “Simulation scenario” components to this calculation would only increase this effect as it adds yet another dimension to measure and therefore makes each level that much more restrictive. Trying to categorize simulation environments into a limited and discrete number of levels of fidelity that are restrictive due to their fidelity component requirements represents a key challenge in developing a usable categorization system. Therefore, the following section discusses potential changes to the structure of the categorization system to address some of the issues identified with the proposed categorization system in Table 7.1.

7.4.2 Potential Modifications to the Categorization System

The fact that there is still a certain amount of subjectivity in the process of categorizing simulation environments for enroute ATC due to the restrictive nature of the proposed categorization system takes away from the usefulness of the fidelity construct and indicates that a different approach to the categorization system may be required. A potential solution to this issue was found in several comments from those who proposed changes to the categorization system in the data of Figure 7.2. The most poignant of these comments offered insight into how a different structure or approach to the categorization system may be the answer:

> “Clearly identify the dimensions which describe 'fidelity' then for each level 'A' through 'E' provide a description of the dimension at that level and some concrete example(s). I can't get back to the four factors in simulation fidelity, but these, coupled with the key terms in your descriptors above, would be the basis of a fidelity estimate. Make it more like a checklist so people can 'objectively' assess a simulation against the categorization system.”

This approach, where instead of the system proposed in Table 7.1 it could potentially resemble something more akin to the underlying framework presented in Figure 7.2, would provide that extra level of detail in order to clearly delineate between two simulation environments. Perhaps even more importantly, this would clearly illustrate the difference in the level of fidelity for each of the main categories within the fidelity definition in Figure 6.1. For example, it is certainly possible for a simulation environment to have a “Physical environment” rated as a Level D, but offering “Inter-Personal communications” and “Simulation functionality” at a Level C.

This also makes it possible to include the category of “Simulation scenario” at a particular level of fidelity as this format allows for a more inclusive categorization process and is not as restrictive as the system proposed in Table 7.1. Further work is required to create a table similar to Table 7.2 for the “Simulation scenario” components identified in Figure 6.1.

Not only is it possible for different levels of fidelity to exist across the different categories of fidelity components within the simulation fidelity definition, but it may actually be more useful to know exactly how each of these categories differs in terms of fidelity and how that then influences the use of the simulation. Being able to see the level of fidelity for each category of fidelity components could potentially offer a clearer picture of a simulation's level of fidelity rather than trying to force a simulation environment into a restrictive category. While this may increase the complexity of the categorization somewhat in that it opens up the possibility of many more different levels of overall simulation fidelity, it may be a necessary trade-off to provide the right amount of information in the categorization system in order to be as useful as possible to those who will be using it.

7.4.3 Potential Implications of Categorization System

Even once a viable categorization system has been established, new issues potentially arise with how it is then used within the ATC industry. One concern that was raised by several survey participants was that it could be adopted as a regulatory tool with regards to training, similar to how the FAA uses its flight simulation environment categorization systems described in Section 2.3. One survey participants comment, in response to the question, "Do you feel that a standardized categorization system similar to the [FAA's full flight simulator categorization system] but adapted for enroute ATC simulation environments is required?", highlighted the potential pitfalls of introducing a categorization system and how it may then be attached to regulation:

> "Provided that a useful link can then be made between the categorization and phases of training where they will be most effective without unnecessary expense. Any categorization carries the risk that it will be incorporated into regulation, therefore the category/use must be carefully defined to ensure that training organizations/operational units are not obliged to invest in simulation equipment that is above and beyond their needs simply because a regulation specifies it."

This is an important issue to consider when developing a categorization system as there can be secondary effects from incorporating the system into the industry's regulatory framework.

One of the issues with incorporating the categorization system into regulation is that it will likely emphasize the use of higher fidelity simulation as there is not enough research to support the use of lower fidelity simulation for achieving specific training outcomes, an issue highlighted in *Section 2.4 Impact of Simulation Fidelity on Training*. It is hoped that the research done in this thesis will encourage research into lower fidelity simulation and how they can be more effectively used to accomplish training outcomes.

7.5 Chapter Summary

What is clear from the work done in this chapter to develop and validate an enroute ATC simulation environment categorization system is that further work is required to create a system that contains the appropriate amount of information while still being user-friendly. There is clearly value in a categorization system, as indicated by the comments from those who thought it was useful in Section 7.3.1 as well as the ATC industry perceptions gathered and analyzed in *Section 5.3 Simulation Categorization Systems Based on Fidelity*. A key issue that will need to be discussed once a categorization system has finally been accepted by the ATC industry is whether or not it will be integrated into the regulatory framework. In order to provide those making this decision with the necessary information, further work is required to investigate how fidelity is linked to specific training outcomes within the enroute ATC domain.

Chapter 8

Other Findings from the Research: Perceptions Regarding the Selection of Simulation Environments for Training and Testing

Even if a clear definition of simulation fidelity is created for enroute ATC, along with an agreed upon categorization system for simulation environments, the challenge of determining what level of simulation fidelity is best used to achieve particular outcomes remains. This is not only a problem for training programs, but also when determining what level of simulation fidelity is going to be used to test the validity of new operational concepts and when conducting research in the ATC domain. This chapter presents a preliminary investigation into industry perceptions regarding what level of simulation fidelity ought to be used to accomplish certain objectives. To be clear, the results presented in this chapter represent perceptions and not necessarily best practices. Questions investigating the appropriateness of different levels of simulation fidelity for accomplishing different tasks were included both during the interview sessions at the ANSP and on the survey distributed to SMEs.

8.1 Trends in Matching Type of Simulation to Training at an ANSP

During the interview sessions at the ANSP, each interviewee was asked the two following questions: (1) “What are the hardest skills or knowledge for trainees to learn?”, and (2) “What are the easiest skills or knowledge for trainees to learn?” The goal of these questions was to identify the types of simulation used to teach the two categories of skills and what that may then imply for simulation use throughout the training program.

Easiest Skills The skills perceived to be the easiest for trainees to learn as posited by interviewees are presented in Table 8.1. The coded knowledge item or skill along with the response rate across the 13 interviewees and sample quotes corresponding to each coded item. This format is repeated for Table 8.2. presenting the results for the hardest skills to train.

Table 8.1 Easiest skills and/or knowledge for enroute ATC trainees to learn as reported by interviewees.

Knowledge/ Skills	Percentage (N=13)	Sample Quotes
Phraseology	62%	• "It's easy because standardized and rule-based." • "Always the same patterns." • "Phraseology can be learned by heart even before they show up, just needs repetition."
Domain-specific knowledge	62%	• "Things that require memorization early on, like phraseology, maps, frequencies, things like that." • "Easy to learn the rules, many students excellent at memorizing all the rules and standards."
Skills in isolation	46%	• "The part tasks, the building of skills in isolation." • "Simple skills and individual tasks seem to be simple for students. It's only when you start combining and coordinating skills they become more challenging."
Secretarial Skills	38%	• "Stuff that you have to do that if nothing ever went wrong or tasks were simple, the students would never get wrong." • "Things that they can repeat at high frequency."

The first two items, phraseology and domain-specific knowledge, are things the trainee is expected to gain proficiency in on their own time, mostly by way of extensive memorization. Much of this information can even be correlated with the introduction package he/she is sent before they arrive for their training. Given that it is mostly provided to the trainee before they arrive for training, this type of information is not something that requires a lot of extra explanation or help from the instructors to learn but rather is information that must be memorized. The secretarial skills (e.g. strip writing, writing as they talk, communication and frequency changes, etc.) and skills in isolation are things which the trainees see in relatively high-volume early in their training and are directly related to the very basic skills needed to perform ATC.

All of the items in Table 8.1 were reported by at least 35% of the interviewees; this indicates a strong consensus regarding what parts of the job are perceived to be easiest for trainees to learn. There was also a strong consensus among participants that these can be considered foundational elements of doing ATC and trainees need to develop a high level of automaticity in these items in order to be successful later in the training program. Based on further conversations with interviewees, it was noted that all of the knowledge and/or skill items in Table 8.1 require no simulation or lower fidelity simulation in order to train them at this particular ANSP.

Hardest Skills Table 8.2 presents the top six hardest skills to learn as identified by interviewees. There are a larger number of skills and/or knowledge items that interviewees believe are harder for trainees to learn, with seven items being specified in addition to the top six reported in Table 8.2 for a total of 13 different items compared to only four in Table 8.1. This indicates that after the easiest parts of the job are learned, trainees face a significant number and variety of skills to learn during the middle and latter stages of their training program.

According to interviewees, the knowledge and/or skills items in Table 8.2 form the key skills needed to be a successful controller. It is also worth noting that the majority of the items mentioned in Table 8.2 are cognitively based skills which are hard to directly measure, making it more challenging to know if a trainee is really developing proficiency in these areas. These six skills are all highly abstract skills, and, while crucial to being a good controller, it was acknowledged by interviewees that they are the hardest to develop expertise in.

Table 8.2 Hardest skills and/or knowledge for enroute ATC trainees to learn as reported by interviewees.

Knowledge/ Skills	Percentage (N=13)	Sample Quotes
Adapting knowledge to new situations	54%	• "Sometimes students look for a formula, but ATC does not respond to a formula as it is a dynamic, changing environment from day to day." • "Understanding how it all fits together, anticipating where the problems are going to be, seeing into the future are the toughest parts to get." • "Take a rule or regulation that you are familiar with and apply it somewhere else, it's still the same rule it's just a completely different application or area."
Prioritization	46%	• "Using priorities to your advantage, it helps us determine what the next action is going to be and allows us to work more efficiently." • "Choosing the right priorities. Not easy to do when you don't have the knowledge to determine what the proper priorities are." • "Difficult as priorities change all the time."
Multi-Tasking	31%	• "Multi-tasking."
Decision making process	31%	• "Sometimes it's not skills, could be the decision making, the speed at which you have to make these decisions. The dynamics of the decision making process, from start to finish."
Situation Awareness	31%	• "… instances of tunnel vision, teaching to see the big picture."
Visualization	23%	• "Being able to extrapolate from a data board and other information what that means for the situation." • "Visualization of airspace."

Based on interviewee comments, the knowledge and/or skills mentioned in Table 8.2 are tied to training during the specialty phase at the ANSP, which relies upon the workstation simulator which is described in section 4.1.3, and during on-the-job training in the operational environment. Learning these skills requires a dynamic, multi-faceted training environment as the skills themselves, such as multi-tasking or prioritizing, are quite dynamic and complex.

Summary Interviewees at an ANSP were asked to identify which skills are easiest for trainees to learn and which are hardest for trainees to learn. The easiest skills for trainees to learn as identified by interviewees are: phraseology, domain-specific knowledge, isolated skills and secretarial skills. The top six hardest skills for trainees to learn are: prioritization, adaptation, multi-tasking, decision making process, situation awareness, and visualization. Upon further discussions with interviewees, a pattern emerged with regards to the types of simulation used for training each of these categories of skills. For the easiest skills, it was clear that no simulation or lower fidelity simulation were primarily used for teaching or learning these skills or knowledge items, whereas training the top six hardest skills is associated primarily with higher fidelity simulation. However, these findings are potentially ANSP-specific; therefore, a larger-scale investigation of industry perceptions regarding the selection of a simulation at a particular level of fidelity for completing a particular task was conducted.

8.2 Industry-wide Perceptions regarding Simulation Selection

To better understand ingrained beliefs regarding simulation use, two types of questions were used in the survey:

1. Investigated the appropriateness of two fidelity anchor points (classroom-based simulation vs. workstation simulation) for accomplishing training, testing and research tasks (Question 12 from online survey in Appendix B).
2. A categorization system usability exercise where survey participants were asked to choose the minimum level of simulation fidelity required to train a particular skill or evaluate a particular concept (Question 25 from online survey in Appendix B).

Survey participants were provided with versions of these questions depending on their responses to a demographic question asking them to specify what their primary use of simulation was (Question 7 in online survey in Appendix B). Participants were separated into either training-oriented scale rating questions or testing-oriented scale rating questions. Sub-section 8.2.1 presents and discusses the responses to the training-oriented questions and is followed by a sub-section that discusses the responses to the testing-oriented questions.

8.2.1 Simulation Selection for Training

Ab-Initio Training Survey participants who primarily use simulation for training purposes were asked to rate their agreement on a 7-point Likert-type scale with regards to the following two statements:

1) A **classroom-based role playing exercise** or case study is **sufficient** to train ab-initio students on skill *x*.
2) A **simulated workstation** with radar display and realistic communications is **required** to train ab-initio students on skill *x*.

Survey participants were asked to rate the following seven skills with regards to the two statements above: phraseology, domain-specific knowledge (e.g. aircraft characteristics, ATC rules, maps, frequencies, etc.), isolated skills (e.g. performing a handoff, issuing a clearance, etc.), visualization, prioritization, sector-specific characteristics and multi-tasking. These skills were identified from the previous section's investigation at the ANSP into the easiest and hardest skills for trainees to learn. A selection of skills from both categories was included. The statements were presented in a random order to survey participants, and the skills for each statement were also presented in a random order.

In addition to the questions above, survey participants were asked to choose the lowest level of simulation fidelity, using the categorization system presented in Chapter 8, which they believe is capable of training the same seven skills presented identified above.

The data gathered from these questions is presented in summary tables to improve understanding of the results. In order to provide context to the information that will be presented in these tables, Figure 8.1 provides a sample of the raw data from the results of all three of these questions with regards to the skill of "Multi-tasking".

Figures 8.1a) and b) presents the results of the rating questions for statements 1) and 2) at the beginning of this section, with the ratings being condensed from the 7-point Likert type scale into the three categories of 'Agree'/'Neutral'/'Disagree'. For the skill of "Multi-tasking", it is clear from Figures 8.1a) and b) that survey participants believe that a classroom-based simulation is not capable of training this skill, but that a workstation simulation is required. Figure 8.1c) presents the results from the categorization system usability exercise, which indicates that survey participants believe a simulation environment of Level D or E fidelity is required to train "multi-tasking". For the full graphical presentation of the data presented in this chapter, please refer to Appendix F.

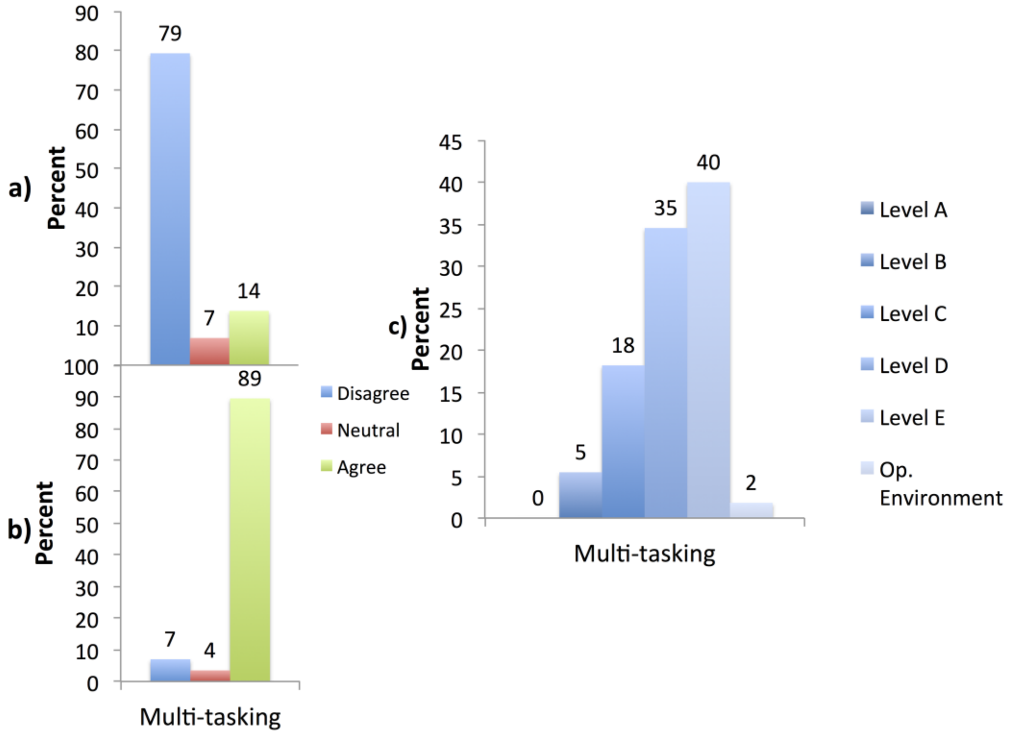

Figure 8.1 Survey participant beliefs regarding simulation selection for ab-initio training.

a) A <u>Classroom-based simulation</u> is capable of training ab-initio students on multi-tasking.

b) A <u>workstation simulation</u> is required to train ab-initio students on skill multi-tasking.

c) Based on the categorization system, what is the <u>minimum level of simulation fidelity</u> required to train ab-initio students in multi-tasking.

In order to present the data in a more concise manner, thresholds were used to summarize the data for each skill. Summarizing the data for Figures 8.1a) and b) consisted of reporting the response, either 'Agree', 'Neutral', or 'Disagree', with the highest response frequency. If another response option was within 10% of the highest reported response, then the data was referred to as 'Unclear'. For Figure 8.1c), the simulation environment with the highest response frequency was reported as well as any other simulation environment within

10% of that top simulation environment. The results of the threshold judgments across all skills are summarized in Table 8.3.

The column labeled "Classroom-based Sim. Sufficient" presents results pertaining to the classroom-based simulation presented to survey participants in statement 1). The second results column, labeled "Workstation Sim. Required", corresponds to results pertaining to the workstation simulation that was presented in statement 2). Finally, the "Minimum Sim. Level" column presents the results pertaining to a survey participant's belief regarding the minimum level of simulation fidelity required to train that particular skill item.

Table 8.3 Summary of survey participant responses to questions about simulation selection with regards to ab initio training of new operators.

Item	Classroom-based Sim. Sufficient (N=58)	Workstation Sim. Required (N=56-58)	Minimum Sim. Level (N=54-55)
Phraseology	AGREE	AGREE	A/C
Domain-specific knowledge	AGREE	UNCLEAR	A
Isolated skills	UNCLEAR	AGREE	C
Visualization	DISAGREE	AGREE	C/D/E
Prioritization	DISAGREE	AGREE	D/E
Sector-specific characteristics	DISAGREE	AGREE	D/E
Multi-tasking	DISAGREE	AGREE	D/E

For the first three items reported in Table 8.3, "Phraseology", "Domain-specific knowledge" and "Isolated skills", there is clearly some ambiguity in the minds of survey participants. There was agreement that the classroom-based simulation environment was sufficient to train these particular items as indicated in the first two results columns in Table 8.3, but participants also believed that the workstation simulation environment was required to train those skills. Perceptions are clearer for some of the more complex cognitive skills, such as "Visualization" and "Prioritization", that survey participants believed a lower fidelity simulation environment was not sufficient to train these skills and that a higher fidelity simulation environment was the only option that could accomplish the training outcomes for those skills and knowledge.

When considering the results from the "Minimum Sim. Level" column, however, lower fidelity simulation environments, such as levels A and B, are clearly favoured for training the first two tasks with a trend towards wanting the highest fidelity simulation environments, levels D and E, for training skills that were perceived to be hardest to train by interviewees in the previous section. There is a strong belief amongst survey participants that Level C, D or E is required to train these harder to learn skills with little support for simulation levels A and B.

The findings from Table 8.3 indicate a strong belief amongst the industry that higher fidelity simulation ought to be relied upon to carry much of the burden in training ab initio students for ATC. It is somewhat less clear what role they believe classroom-based simulation should play in training new controllers given the ambiguity of their response with regards to this particular simulation. This could be due to the lack of literature that exists within the domain that details the benefits and strengths of lower fidelity simulation for training different ATC skills.

This point is illustrated with the skill "Visualization" from Table 8.3. "Visualization" is a skill that requires an individual to cognitively visualize and project the situation based almost entirely on data without relying on visual cues or automated tools, and based on the results from the Table 8.3, it should be trained with a higher fidelity simulation such as the workstation simulation. Based on discussions with SMEs at the ANSP, there are several ways this skill could be trained in lower fidelity simulations as many aspects of it do not rely on the environmental components inherent to a workstation simulation. This could also be said of learning knowledge such as “Sector-specific Characteristics”.

The reliance on higher fidelity simulation was discussed in section *2.4 Impact of Simulation Fidelity on Training*. As shown in that section, there is a significant amount of high-level research that indicates that lower fidelity simulation has much to offer in terms of training benefits and is often overlooked due to their lack of fidelity. One of the challenges facing individuals who design and implement ATC training programs that want to increase the use of lower fidelity simulation will be to counter these ingrained perceptions regarding higher fidelity simulation.

The research presented in this thesis can help to clarify what environmental components can affect fidelity, enabling clearer discussions regarding which components are required to train certain skills. Given what appears to be deeply ingrained perceptions regarding the use of various simulations for ab initio training, this is where further research into the links between simulation fidelity and training outcomes, based on the foundational material presented in this thesis, can provide more clarity regarding these perceptions.

Recurrent Training In addition to ab-initio training, survey participants who primarily use simulation for training purposes were also asked to rate their agreement with regards to the following two statements on recurrent training[7]:

3) A **classroom-based role-playing exercise** or case study is **sufficient** to train qualified controllers on skill x.

---

[7] Recurrent training represents the continual training of qualified controllers whereas ab initio training represents the training of newly hired individuals.

4) A **simulated workstation** with radar display and realistic communications is **required** to train qualified controllers on skill x.

For statements 3) and 4) regarding the training of qualified controllers, survey participants were asked about the following training situations: a new operational procedure (e.g. new traffic flow pattern), a new operational tool (e.g. electronic flight strips), and emergency situations. These training situations were identified during discussions with interviewees at the ANSP, and it was determined that these were three common instances that necessitated recurrent training of qualified controllers. The same presentation format was used as in the previous Ab-Initio Training sub-section, and the same follow on question was provided that asked survey participants to choose the minimum level of simulation fidelity they believe is required to train qualified controllers for these situations (Question 25 in Appendix B).

Table 8.4 presents the responses to statements 3) and 4) as well as the results from the simulation environment categorization system usability exercise. Table 8.4 follows the same presentation rules that were outlined for Table 8.3.

Table 8.4 Summary of survey participant responses to questions about simulation selection with regards to recurrent training of active controllers.

Item	Classroom-based Sim. Sufficient (N=58)	Workstation Sim. Required (N=58)	Minimum Sim. Level (N=54-55)
New procedures	DISAGREE	AGREE	E
New skills	DISAGREE	AGREE	D/E
Emergency situations	DISAGREE	AGREE	E

What is clear from the first two columns in Table 8.4 is that survey participants strongly believe that qualified controllers should be using a higher fidelity simulation in order to complete recurrent training such as practicing emergency situations as well as receiving training on new operational tools and procedures. This perception is strongly reinforced when considering the results from the follow on question presented in the “Minimum Sim. Level” column, where survey participants indicated that they believe simulation fidelity Levels D and E are required to train qualified controllers for these recurrent training situations.

These findings are potentially due to qualified controllers having strongly ingrained work habits in the operational environment, and believing that practicing at a level of fidelity too low may actually force them out of these habits and potentially negate much of the value in the training. Training in lower fidelity simulation could potentially cause a negative or neutral transfer of training when they return to the operational environment thereby negating the value of the training. Practicing in a simulation as close as possible to live operations would allow them to see more clearly how a new procedure or tool will impact their working habits and thus be more prepared when they return to the operational environment. The fact that 60% of those who completed the survey have operational experience, as noted in Section 3.2.3, serves to reinforce the impact of this finding.

Once again these are survey participants’ perceptions, meaning there is room to investigate whether lower fidelity simulation could be designed to target certain environmental components, using the simulation fidelity definition in Chapter 6, which would allow for adequate recurrent training of qualified controllers with a more flexible simulation. There is certainly potential to further investigate the link between fidelity and emergency training in ATC as it was identified by Dahlstrom et al. (2009) that medium fidelity simulation emergency training could be more beneficial as it was better at training the general skills and competencies that are most useful during emergency situations. This study was conducted in the domain of marine transportation, and determining whether or not these findings could transfer to the domain of ATC could prove valuable to ANSPs around the world.

8.2.2 Simulation Selection for Testing and Research

Testing New Operational Concepts Survey participants who primarily use simulation for testing or research purposes were asked to rate their agreement on a 7-point Likert scale with the following two statements:

1) A **role-playing exercise** or case study is **sufficient** to evaluate the acceptability of new operational concepts for the work environment such as *x*.
2) A **simulated workstation** with radar display and realistic communications is **required** to evaluate the acceptability of new operational concepts for the work environment such as *x*.

Survey participants were asked to rate the following six operational concepts or tools: traffic flows, procedures, decision support tools, interface tools, system automation, and information management tools. These concepts were identified based on discussions with SMEs at the ANSP in terms of common updates to the operational environment. The statements were presented in a random order to survey participants, and the skills for each statement were also presented in a random order. The same question presentation format as in previous sub-sections was used, and the same follow on question was provided that asked survey participants to choose the minimum level of simulation fidelity they believe is capable of testing the same six operational tools or concepts for deployment in the operational environment. The results from these questions are summarized in Table 8.5 using the same analysis format as Table 8.3.

Table 8.5 Summary of survey participant responses to questions about simulation selection with regards to testing new operational concepts for the operational environment.

Item	Classroom-based Sim. Sufficient (N=28)	Workstation Sim. Required (N=27-28)	Minimum Sim. Level (N=25)
Traffic flows	DISAGREE	AGREE	D
Procedures	DISAGREE	AGREE	D
Decision support tools	DISAGREE	AGREE	C/D/E
Interface tools	DISAGREE	AGREE	D
System automation	DISAGREE	AGREE	D
Information management tools	DISAGREE	AGREE	D

With a clear pattern in the first two results columns indicating that a classroom-based simulation environment is not capable of testing these operational concepts, combined with a belief that a Level D simulation environment is the minimum level of simulation fidelity required to test most of these concepts, there is little doubt what those in the ATC industry believe higher fidelity simulation is required for testing operational concepts. While there may be room for lower fidelity simulation in earlier prototyping stages for these tools and concepts, this desire for higher fidelity simulation is likely due to the fact that individuals want to be as confident as possible introducing these new concepts into the working environment and the best way to ensure this confidence in their eyes is to test them in a simulation with a high degree of realism.

This view regarding simulation use for testing operational concepts provides the opportunity to investigate whether simulation with a targeted approach for the fidelity of certain key components, similar to the approach discussed for recurrent training in the previous sub-section, could provide more flexibility and options for users of simulation within this area of focus. The biggest challenge with this particular use of simulation, as noted earlier, is the fact that these tools and concepts are being implemented in the operational environment, meaning that safety and functionality must be extensively tested to ensure a smooth transition. Whether or not lower fidelity simulation is capable of accomplishing these objectives requires significantly deeper investigation as it would take considerable evidence to convince ANSPs that this is a viable option.

Researching Human Factors Issues In addition to testing operational concepts, survey participants who primarily use simulation for testing and research purposes were also asked to rate their agreement with regards to the following two statements on research of human factors issues:

3) A **role-playing exercise** or case study is **sufficient** to evaluate the effect of new operational concepts on human factors issues such as *x*.
4) A **simulated workstation** with radar display and realistic communications is **required** to evaluate the effect of new operational concepts on human factors issues such as *x*.

Survey participants were asked about the following five issues: situation awareness, transfer of training, complexity, decision-making, and human-automation interaction. These items were identified as being common human factors related issues in ATC domain based on previous research on enroute ATC. As with the questions from the previous sections, the statements and the human factors issues appeared in a randomized order.

In addition to the questions above, survey participants were asked to choose the lowest level of simulation fidelity, using the categorization system presented in Chapter 7, that they believe is required to research those same five human factors issues. The results from these questions are summarized in Table 8.6 using the same analysis format as Table 8.3.

Table 8.6 Summary of survey participant responses to questions about simulation selection with regards to human factors research.

Item	Classroom-based Sim. Sufficient (N=28)	Workstation Sim. Required (N=28)	Minimum Sim. Level (N=23-24)
Situation awareness	DISAGREE	AGREE	D
Transfer of training	DISAGREE	AGREE	D/E/In the operational environment
Complexity	DISAGREE	AGREE	D/E
Decision-making	UNCLEAR	AGREE	D/E
Human-automation interaction	UNCLEAR	AGREE	C/D

Based on the results from the first two columns, there is once again a strong desire for a higher fidelity simulation environment to be used to conduct research on a variety of human factors issues. However, there are two human factors issues where this may not be the case. Results for decision-making and human-automation interaction were unclear as to the appropriateness of a classroom-based simulation environment being capable of researching these topics. While this makes intuitive sense for decision making as there are several studies which investigate controller decision making strategies with the use of static pictures or scenarios (e.g. Stankovic et al, 2011; Boag et al, 2006; Hyun et al, 2006), this was a more surprising finding for human-automation interaction given the topic requires a certain level of fidelity just to replicate the automation inherent in the operational environment.

When considering the "Minimum Sim. Level" results column, there is once again a strong desire for Level D simulation fidelity. In the context of conducting research, this likely improves the inherent validity of the results if the research is being conducted in a testing environment that is as similar to the target environment. It likely makes the results of this research more attractive to those who seek to make use of its findings, such as ANSPs. A significant amount of survey participants feel "Transfer of training" should be researched in the operational environment, a finding that is likely due to the fact that to analyze how training has transferred one must include the final environment where the learned skills are to applied, in this case the operational environment. "Transfer of training" can be studied looking at any level of simulation fidelity, but the transfer must then be analyzed in context of the operational environment where those skills are used in day-to-day operations.

8.3 Chapter Summary

This chapter presented a variety of perceptions within the ATC industry regarding how simulation environments of varying degrees of fidelity are best used to accomplishing various tasks in training, testing new operational concepts and researching the future ATC environment. These are, however, perceptions and are not necessarily good practice solely based on consensus. In fact, there is contradictory evidence regarding the use of simulation in training, as documented in section 2.4, and the perceptions presented in the Ab Initio Training perception results in Section 8.2.1. This is one of the few areas where there is enough evidence to potentially induce a re-evaluation of what ought to be considered a best practice. The other areas where perceptions were investigated in this chapter are meant to stimulate discussion of how simulation is currently used within the industry as well as offer potential areas for further research to determine if there are other options to the status quo. Investigating more deeply how fidelity is connected to accomplishing objectives in training, testing new operational concepts and future ATC environment research could significantly impact how simulation is used across the industry and is a crucial next step in this area of research.

Chapter 9

Conclusion and Future Work

This chapter summarizes the main findings of this thesis and the potential for future research opportunities. The first section, Research Findings and Conclusions, reviews the objectives laid out in Section 1.3 and summarizes how each one has been achieved. The second section, Contributions, discusses the overall contributions of the research presented in this thesis to the ATC industry and the broader academic community. The final section, Future Work, will discuss future work opportunities that build upon the foundational material presented in this thesis.

9.1 Research Findings and Conclusions

The overall goal of this thesis, as stated in Section 1.2, was to examine how to introduce more consistency to the comparison of simulations within the domain of enroute ATC. This thesis sought to achieve this high level goal by achieving the following four research objectives as stated in Section 1.3.

1. Identify how professionals within the domain of ATC currently make simulation fidelity determinations.
2. Develop an objective enroute ATC simulation fidelity definition.
3. Develop an enroute ATC simulation environment categorization system.
4. Validate the enroute ATC simulation fidelity definition and categorization system within a diverse sample of the industry.

The first objective was achieved by findings presented in Chapters 4 and 5. The key finding from Chapter 4 that helped to achieve this objective was the initial identification of certain parts of the operational environment that individuals were considering when determining the fidelity of a simulation in the context of the limitations and differences, as presented in Section 4.3. It also indicated that further investigation into which specific environmental components individuals were considering when determining the fidelity of an enroute ATC simulation was required. Chapter 5 then identified that simulation fidelity was not well defined for enroute ATC. This was due to the fact that different individuals possessed different interpretations of what components ought to be considered when determining the fidelity of a simulation, as discussed in Section 5.2. These findings helped to identify how professionals within the industry make determinations regarding fidelity and provided clear motivation to for the development of the simulation fidelity definition.

The second objective was achieved by developing a simulation fidelity definition for enroute ATC based on an approach that was adapted from previous research in the general subject area of simulation fidelity, as was discussed in sub-section 3.1.1 and Section 6.1. This definition, presented in Figure 6.1, identified the components of the enroute ATC operational environment that affect the fidelity of a simulation of that environment.

The third objective was achieved by developing an enroute ATC simulation environment categorization system, presented in Table 7.1, based on SME input that was gathered during interviews at the ANSP and the structure of categorization systems from other domains. The categorization system provides five distinct categories of simulation fidelity for differentiating between the various simulation environments used within the ATC industry while using the simulation fidelity definition as the points of comparison between simulation environments.

Finally, the fourth objective was achieved via various validation activities for the two constructs. The fidelity definition was initially validated by the 13 interviewees at the ANSP;

however, due to the small number of interviewees and the similarity of their operational expertise, a larger scale validation exercise was conducted to determine if the definition applied to a wider variety of users. Based on the findings from this larger scale validation exercise, described in section Section 6.3.2, it is clear that survey participants (N=86) strongly believed that the definition captured the environmental components that affect fidelity for the enroute ATC domain. In addition to this, there were no statistical differences with regards to this belief within the four demographic groups, which indicates that the fidelity definition is likely generalizable to a wide variety of users.

The categorization system also used two validation activities: an online survey, results of which are presented in sub-section 7.3.1, and an application exercise, results of which are presented in sub-section 7.3.2. During the survey, several survey participants raised concerns in their comments regarding the existence of gray areas between levels where a particular simulation environment may fall under more than one category. This issue was made clear during the secondary validation activity where enroute ATC simulations used in research papers were to be categorized using the system. It was clear when trying to select the appropriate category for certain simulation environments that different parts of that simulation fell under different levels of fidelity. Potential structural modifications were discussed in Section 7.4.2. Given this, there is a clear belief that a simulation environment categorization system would be useful to professionals within the industry, but the most useful structure of this system likely requires further work.

9.2 Contributions

This thesis has not only made significant contributions to the understanding of simulation fidelity within the domain of ATC, but also to the general process of defining simulation fidelity for any domain.

First, with regards to the domain of enroute ATC, there has been no work on the concept of simulation fidelity within the domain of enroute ATC, as was noted in Section 2.2. The simulation fidelity definition developed for this thesis has helped to bring a clearer understanding of what affects simulation fidelity for enroute ATC, and in doing so provides simulation users within the industry a tool by which they can discuss fidelity in more objective, concrete terms.

The categorization system also introduces more clarity and standardization with regards to simulation fidelity for enroute ATC. The categorization system from Chapter 7 represents a first attempt at developing this type of a system for enroute ATC, and while there are limitations with its current format, it does provide a foundation and important lessons learned in order to develop a final end product for the industry at large.

The third area of contribution is the perceptions regarding simulation use gathered and analyzed in Chapter 8. These perceptions offer important insight into how individuals currently believe simulation of varying degrees of fidelity ought to be used to accomplish different training, testing and research tasks within the industry. These perceptions do not necessarily represent best practices, and explicitly identifying these perceptions allows for discussion regarding their merit. More importantly, it provides the opportunity to investigate and more clearly identify the best practices for choosing the appropriate level of simulation fidelity for a given task within the enroute ATC domain.

The final area of contribution is the process for defining simulation fidelity that was explicitly addressed in this thesis, presenting a potential framework to follow for other domains while also discussing lessons learned throughout the development and best practices. This will hopefully encourage other domains that rely heavily on simulation to explore its effects and identify ways of improving their simulation use by developing a clear and objective definition for their domain using the approach presented in this thesis.

9.3 Future Work

As was stated during Chapter 1, the work completed in this thesis was foundational in nature. The long term goal is to investigate the links between simulation fidelity and simulation use within the ATC industry, potentially identifying more effective and efficient ways of using various levels of simulation fidelity.

First, however, there are opportunities for further work on the constructs developed in this thesis. One of the primary opportunities is the further development of the simulation environment categorization system. As was noted in Chapter 7, the current version of the categorization system has limitations and further work is required to determine the most effective format of the construct. Potential modifications to the current categorization system were discussed in Section 7.4.2, but gathering more SME feedback on the most useful structure of the categorization system as well as its scope would be the most likely next step.

One of those potential next steps with the categorization system involves developing a measurement rubric, similar to those in Table 7.2, for the fidelity of the “Simulation scenario” components from Figure 6.1. As was noted in Section 7.4.1, this category of components was omitted from the categorization system presented in Table 7.1, but there is the potential for introducing these components into the system once a new structure for the categorization system has been developed. Another key extension of this work is determining the equivalency of the level of fidelity across component categories. For instance, what makes “Communication participants” at a Level D equivalent to “Control interfaces” at a Level D. Further SME input is likely required to investigate these topic areas, but it is believed that this work would produce a more robust and useful construct for industry professionals.

There is also an opportunity to investigate the relative importance or weights of each component within the simulation fidelity definition, as was discussed in Section 6.4. It is probable that the value of a particular component is dependent on what the simulation is being used to accomplish, and further investigation into this topic could provide deeper insight into how fidelity needs to be varied depending on the task being trained.

The larger scale next step is to then investigate the links between simulation fidelity and how simulation is used for training, testing new operational concepts and future ATC environment research. One of the best opportunities for future work is to investigate the links between fidelity and transfer of training within the domain of enroute ATC. The work completed in this thesis allows for the structuring of various conditions for transfer of training studies. This type of work would also help to address the perceptions held by professionals within the industry regarding how simulation of varying degrees of fidelity ought to be used, a topic investigated in Chapter 8. Providing more concrete evidence regarding how to structure the use of simulation of various degrees of fidelity would help to either confirm or repudiate the various perceptions that do exist.

Finally, there is also an opportunity to investigate the potential of developing a targeted simulation fidelity methodology for simulation use to achieve a particular outcome. This methodology would identify the components of the simulation fidelity definition that are most relevant to the task being accomplished, and then design and/or use simulation with components at a high level of fidelity in only those areas. This potentially cuts down on needing to rely on overall higher fidelity simulation to increase the validity or effectiveness of training, testing new operational concepts or future ATC environment research.

The opportunities discussed above represent only a handful of future avenues of research on the topic of simulation fidelity within the domain of enroute ATC. It is hoped that the work presented in this thesis will stimulate other opportunities for further investigation into how simulation fidelity affects simulation use within the industry, and potentially within other domains as well.

References

Alexander, A. L., Brunyé, T., Sidman, J., & Weil, S. A. (2005). *From Gaming to Training : A Review of Studies on Fidelity , Immersion , Presence , and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games* (pp. 1–14). Retrieved from http://www.aptima.com/publications/2005_Alexander_Brunye_Sidman_Weil.pdf

Barr, M., Brady, T., Koleszar, G., New, M., & Pounds, J. (2011). *FAA Independent Review Panel on the Selection, Assignment and Training of Air Traffic Control Specialists*. Federal Aviation Administration: Washington, DC.

Beaubien, J. M. and Baker, D. P. (2004). The use of simulation for training teamwork skills in healthcare: how low can you go? *Quality and Safety in Health Care*, *13*, i51–i56.

Blaiwes, A.S., Puig, J.A. and Regan, J.J. (1973). Transfer of Training and the Measurement of Training Effectiveness. *Human Factors.* 15(6), pp. 523-533.

Blanken, M. J. S. (2002). Chapter 13: Theory and Research – Training for Future ATM. In H. Eissfeldt, M. C. Heil, & D. Broach (Eds.), *Staffing the ATM system: The selection of air traffic controllers* (pp. 155–167). Burlington, VT: Ashgate.

Boag, C., Neal, A., Loft, S., & Halford, G. S. (2006). An analysis of relational complexity in an air traffic control conflict detection task. *Ergonomics*, *49*(14), 1508–26. DOI: 10.1080/00140130600779744

Caird, J. K. (1996). Persistent issues in the application of virtual environment systems to training. In *Proceedings Third Annual Symposium on Human Interaction with Complex Systems. HICS'96*, 124–132.

Craighead, J., Murphy, R., Burke, J., & Goldiez, B. (2007). A Robot Simulator Classification System for HRI. *International Symposium on Collaborative Technologies and Systems, 2007.*, 93–98.

Dahlstrom, N., Dekker, S., Van Winsen, R., and Nyce, J. (2009). Fidelity and validity of simulator training. *Theoretical Issues in Ergonomics Science*, *10*(4), 305–314.

DNV GL. (2014a). *Our organisation*. Retrieved from https://www.dnvgl.com/about/in-brief/our-organisation.html

DNV GL. (2014b). *Maritime Simulator Systems*. Retrieved from http://www.dnv.com/industry/maritime/servicessolutions/competence/classificationoftraining/certificationofmaritimesimulatorsystems.asp

Drown, D.F. and Lowry, I.J. (1993). A categorization and evaluation system for computer-based ship operation training. In *Marsim '93. International Conference on Maritime Simulation and Ship Maneuverability, St. Johns, Newfoundland, Canada, September 26 -October 2*. p. 103-113.

Durso, F. T., & Manning, C. a. (2008). Air Traffic Control. In C. M. Carswell (Ed.), *Reviews of Human Factors and Ergonomics* (4th editio, Vol. 4, pp. 195–244). Santa Monica: Human Factors and Ergonomics Society. DOI: 10.1518/155723408X342853.

Eryilmaz, U., Tokmakb, H., Cagiltayc, K., Islerd, V., and Eryilmaz, N. (2014). A novel classification method for driving simulators based on existing flight simulator classification standards. *Transportation Research Part C: Emerging Technologies.* 42, p. 132–146.

Estock, J. L., Alexander, A., Gildea, K. M., Nash, M., and Blueggel, B. (2006). A Model-based Approach to Simulator Fidelity and Training Effectiveness. In *Proceedings of Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. Orlando, FL: National Security Industrial Association.

Eurocontrol. (2012). European ATM Master Plan: The Roadmap for Sustainable Air Traffic Management. Brussels, Belgium: Eurocontrol.

Federal Aviation Administration. (2013). *Title 14 of the Code of Federal Regulations* (CFR). Part 60, Appendix A. Retrieved on March 3, 2014 from http://www.law.cornell.edu/cfr/text/14/part-60/appendix-A

Federal Aviation Administration. (2012). A Plan for the Future: 10-Year Strategy for the Air Traffic Control Workforce. Federal Aviation Administration: Washington, DC.

Hannah, S. D., & Neal, A. (2014). On-the-Fly Scheduling as a Manifestation of Partial-Order Planning and Dynamic Task Values. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 1–20. DOI: 10.1177/0018720814525629

Hays, R., Jacobs, J., Prince, C. and Salas, E. (1992). Flight Simulator Training Effectiveness: A Meta-Analysis. *Military Psychology*. 4(2), p. 63-74. DOI: 10.1207/s15327876mp0402_1.

Hays, R.T. and Singer, M.J. (1989). *Simulation Fidelity in Training System Design: Bridging the Gap Between Reality and Training.* New York: Springer.

Hays, R.T. (1980). *Simulator Fidelity: A Concept Paper*. (ARI Technical Report 490). Alexandria, VA.: U.S. Army Research Institute.

[8]Heintzman, R., Middendorf, M., & Basinger, J. (1999). *Development and validation of a method of evaluating the effectiveness of fighter aircraft simulation force cueing devices.* Retrieved February 25, 2014, from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA375328

Hess, R. and Marchesi, F. (2009). Analytical Assessment of Flight Simulator Fidelity Using Pilot Models. *Journal of Guidance, Control, and Dynamics*. 32(3), pp. 760-770. DOI: 10.2514/1.40645

---

[8] Source included as part of quote from Estock et al. 2006 on p. 20

Feinstein, A.H. and Cannon, H.M. (2001). Fidelity, verifiability, and validity of simulation: Constructs for evaluation. Developments in Business Simulation and Experiential Learning. 28, p. 57-67.

Hilburn, B.G. (2004) Cognitive Complexity in Air Traffic Control: A Literature Review, EEC Note No.04/04: Eurocontrol.

Hilburn, B., & McGarry, K. (2006). *Emerging Human Factors Issues in Future Air Traffic Management* (p. 79). Retrieved from http://chpr.nl/index_htm_files/EMERGE.pdf

Howell, D. (2013). *Statistical Methods for Psychology*. Belmont, CA: Wadsworth.

Hyun, S. H., Kim, K. T., Song, S. M., Yoon, J. S., Lee, J. I., Cho, S. M., & Sohn, Y. W. (2006). Effects of Expertise and Situation Complexity on Visual Attention and Action Planning for Air Traffic Control. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(3), 420–423. DOI: 10.1177/154193120605000344

ICAO. (2012). *Robust Traffic Growth Expected Until 2014*. Montreal, QC. Retrieved from http://www.icao.int/Newsroom/Pages/robust-traffic-growth-expected-until-2014.aspx

ICAO. (2007). *Outlook for Air Transport to the Year 2025*. Montreal, QC: ICAO

Jackson, P., 1993. Applications of virtual reality in traning simulation. In: K.Warwick, J. Gray, and D. Roberts, eds. *Virtual reality in engineering*. London: The Institution of Electrical Engineers.

Jha, P. D., Bisantz, A. M., Parasuraman, R., & Drury, C. G. (2011). Air Traffic Controllers' Performance in Advance Air Traffic Management System: Part I—Performance Results. *The International Journal of Aviation Psychology*, *21*(3), 283–305. DOI: 10.1080/10508414.2011.582456

[9]Lee, A. T. (2005). *Flight simulation: Virtual environments in aviation.* Burlington, VT: Ashgate Publishing.

Lee, P. U., & Prevot, T. (2012). Prediction of Traffic Complexity and Controller Workload in Mixed Equipage NextGen Environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 100–104. DOI: 10.1177/1071181312561041

Liu, D., Macchiarella, N.D., and Vincenzi, D.A. (2008). Simulation Fidelity. In Vincenzi, D.A., Wise, J.A. Mouloua, M., and Hancock, P.A. (Eds.), *Human Factors in Simulation and Training* (pp. 61-73). Boca Raton, FL.: CRC Press.

Loft, S., Finnerty, D., & Remington, R. W. (2011). Using Spatial Context to Support Prospective Memory in Simulated Air Traffic Control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(6), 662–671. DOI: 10.1177/0018720811421783

---

[9] Source included as part of quote from Estock et al. 2006 on p. 20

Loft, S., Hill, A., Neal, A., Humphreys, M., & Yeo, G. (2004). ATC-lab: an air traffic control simulator for the laboratory. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 331–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15354699.

Longridge, T., Bürki-Cohen, J., Go, T., and Kendra, A. (2001). Simulator Fidelity Considerations for Training and Evaluation of Today's Airline Pilots. *Proceedings of the 11th International Symposium on Aviation Psychology, Columbus, OH, 5-8 March 2001.* Dayton, OH; Wright State University.

Maran, N. J. and Glavin, R. J. (2003). Low- to high-fidelity simulation - a continuum of medical education? *Medical Education*, *37*(Suppl. 1), 22–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14641635.

Marsh, A. (2011). *ABCs of Simulators – AOPA*. Retrieved from http://www.aopa.org/News-and-Video/All-News/2011/May/1/ABCs-of-Simulators

Moroney, W. and Lilienthal, M. (2008). Human Factors in Simulation and Training: An Overview. In Hancock, P., Vincenzi, D., Wise, J., and Mouloua, M. (Eds.) *Human Factors in Simulation and Training* (p.4-33). Boca Raton, FL: CRC Press.

Noble, C. (2002). *The Relationship Between Fidelity and Learning in Aviation Training and Assessment.* Journal of Air Transportation. 7(3), p. 33-54.

Packer, M. (2011). *The Science of Qualitative Research*. Cambridge, New York: Cambridge University Press.

Salas, E., Bowers, C. A., and Rhodenizer, L. (1998). It Is Not How Much You Have but How You Use It : Toward a Rational Use of Simulation to Support Aviation Training. *The International Journal of Aviation Psychology*, *8*(3), 197–208.

Sethumadhavan, A. (2009). Effects of Automation Types on Air Traffic Controller Situation Awareness and Performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *53*(1), 1–5. DOI: 10.1177/154193120905300102

Smode, A.F. (1971). *Human Factors Inputs to the Training Device* (Technical Report NAVTRADEVCEN 69-C-0298-1). Orlando , FL: Naval Training Equipment Center.

Stankovic, S., Loft, S., Rantanen, E., & Ponomarenko, N. (2011). Individual Differences in the Effect of Vertical Separation on Conflict Detection in Air Traffic Control. *The International Journal of Aviation Psychology*, *21*(4), 325–342. DOI: 10.1080/10508414.2011.606744

Thomas, M. (2003). Operational Fidelity in Simulation-Based Training : The Use of Data from Threat and Error Management Analysis in Instructional Systems Design . In *Proceedings of SimTecT2003: Simulation Conference* (pp. 91–95). Adelaide, Australia: Simulation Industry Association of Australia.

United States Government Accountability Office (USGAO). (2008). *Efforts to Hire, Staff, and Train Air Traffic Controllers Are Generally on Track, but Challenges Remain*. Washington, DC.

Van de Merwe, K., Oprins, E., Eriksson, F., & van der Plaat, A. (2012). The Influence of Automation Support on Performance, Workload, and Situation Awareness of Air Traffic Controllers. *The International Journal of Aviation Psychology*, *22*(2), 120–143.

Van Merrienboer, J. J. G., & Kirschner, P. A. (2013). *Ten Steps to Complex Learning: A Systematic Approach to Four-Component Instructional Design* (2nd ed.). Burlington, VT: Routledge.

Weber, B., Oberheid, H., & Papenfuss, a. (2013). Visual Aids for Merging Mixed Arrival Traffic: Effects on Controllers' Performance, Situation Awareness, and Mental Workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 630–634. DOI: 10.1177/1541931213571135

Weitz, J., & Abler, S. (1973). The Optimal Use of Simulation. *The Journal of Applied Psychology*, *58*(2), 219–224.

Wickens, C. D. and Hollands, J. G. (2000). *Engineering psychology and human performance: Third edition*. Upper Saddle River, NJ: Prentice Hall.

Appendix A

Interview Questions during Site Visits at ANSP

Topic	Question #	Question
Training Program	1	In your opinion, what skills or parts of the job are most difficult for students to learn? What are the top 3?
	2	What skills or parts of the job are easiest for students to learn? What are the top 3?
	3	What skills or parts of the job are most difficult to design/develop training for? What are the top 3?
	4	What skills or parts of the job are easiest to design/develop training for? What are the top 3?
	5	Where is simulation currently being used most effectively in the training process?
	6	Where do you have the most challenges with the use of simulation in the current training process?
	7	Are there opportunities in the training to take greater advantage of simulation?
	8	Are there opportunities where you think the use of simulation within training should be reduced?
Simulation Capabilities	1	What are the limitations regarding the currently available tools/techniques used for simulation? Top 3. -Do these limitations significantly affect the use of simulation in training?
	2	What does ________ simulation do well with regards to the training process? What does it not do well? (ask for each type of simulation NC uses)
	3	Do you feel that ________ simulation should be used more or less? Why do you feel that way? (ask for each type of simulation)
	4	Are there any cases you can think of where simulation is not able to reproduce critical aspects of the real world environment?

Topic	Question #	Question
Simulation Selection	1	Do you use classroom-based simulations during your training program? Could you offer some examples? - What do you feel this type of simulation is good for in the training process?
	2	In your opinion, what influences the decision to use a specific type of simulation to achieve a certain training objective?
	3	Among the different types of simulation Nav Canada uses in training, which do you feel is the most important and why is that?
Training for New Tools/Procedures	1	What role does the training department play in the development and implementation of training for new operational tools or procedures?
	2	How is simulation used in this type of training?
OJT	1	In your opinion, why is OJT required after high-fidelity simulation?
	2	What are the challenges experienced in OJT that are not experienced in simulation exercises?
	3	On a scale of 1 to 10, with 1 being not similar in any way and 10 being an identical replication, how close is the workstation simulation environment to the actual working environment? Pc-based sim?
Simulation Scenarios	1	When designing a training scenario, what are the information requirements needed to design the scenario and how are those determined?
	2	In your opinion, how does complexity fit into the design of training scenarios? - What challenges are associated with controlling the appropriate level of complexity for a scenario?
	3	What role do you feel complexity plays in the structure and progress of training? - Should its role be increased or decreased?

Topic	Question #	Question
Fidelity	1	In your opinion, what does fidelity mean in terms of an ATC simulation? Provide your definition.
	2	What are the most important components of a simulation that affect the level of fidelity experienced by a user?
	3	Do you feel that the fidelity of a simulation has an effect on the trainee's learning of a skill? -How?
	4	In your opinion, what effect does the simulation scenario have on the fidelity experienced by the trainee?
	5	In your opinion, does the level of fidelity of a simulation affect selection of that simulation when designing the training program? -If yes, how?
	6	How is fidelity related to the training for new operational tools? Are there any things that wouldn't require high-fidelity simulation?
Our Fidelity Work	1	In your opinion, do these components of ATC simulation fidelity **(present components)** accurately represent the components of an ATC simulation that can affect the experienced level of fidelity by the user? -Would you change any of these components? Please explain your choices
	2	Does this graphical representation **(present image)** of the fidelity of an ATC simulation accurately convey the level of fidelity of that particular simulation and how that determination was achieved?
	3	This is an example of different levels of fidelity used by the FAA to categorize different flight simulators **(present flight sim levels)**. Do you feel that a standardized level system for ATC simulations such as this would be useful? -What would your equivalent of these levels be for ATC simulations?

Appendix B

Hard Copy of Online Survey

Consent Form

By digitally signing this consent form, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities.

____

I have read the information presented in the information letter about a study being conducted by Colin Dow and Dr. Jonathon Histon of the Department of Systems Design Engineering at the University of Waterloo. I have had the opportunity to ask any questions related to this study, to receive satisfactory answers to my questions, and any additional details I wanted.I am also aware that excerpts from responses to questions may be included in the thesis and/or publications to come from this research, with the understanding that the quotations will be anonymous.I was informed that I may withdraw my consent at any time without penalty by advising the researcher. This project has been reviewed by, and received ethics clearance through a University of Waterloo Research Ethics Committee. I was informed that if I have any comments or concerns resulting from my participation in this study, I may contact Dr. Maureen Nummelin, the Director, Office of Research Ethics at 519-888-4567 ext. 36005. Answering "No" to any question below will automatically end your participation in the survey.

With full knowledge of all foregoing, I agree, of my own free will, to participate in this study.

○ Yes

○ No

I agree to the use of anonymous quotations any thesis or publication that comes from this research.

○ Yes

○ No

I agree that the data collected will be used in Colin Dow's thesis as well as by other students in the Humans in Complex Systems Lab at the University of Waterloo.

○ Yes

○ No

Introduction to ATC Simulation Fidelity Survey

Welcome to the ATC Simulation Fidelity Survey. This survey consists of three main sections and a potential fourth depending on your area of expertise.

-Background Information

-Simulation Fidelity in the Enroute ATC Domain

-Tools for Comparison of ATC Simulation Environments

-Simulation Selection for Evaluating New Tools and Procedures

We ask that you answer honestly and with as much detail as possible. Please check your answers before proceeding to following pages as you will not be able to go back through the survey for review.

Reminder: This survey should take 30-40 minutes to complete. You may save your responses and continue at a later time by clicking the "Save and continue later" button at the bottom of each page. In order to ensure consistency across all survey participants, the researchers offer the following definition of 'enroute ATC simulation':

Any environment where a user/operator actively practices providing some or all of the air traffic services. Typically this is done with some form and/or combination of tools, objects, or personnel to replicate some part of the real world task environment. Examples of different types of simulation include an instructor moving plastic airplanes around on a tabletop sector, a personal computer part-task program, and a complete reproduction of a radar controller's workstation and work environment.

NOTE: This survey is investigating human in the loop simulations and excludes fast-time simulations. Please remember this when answering the questions.

Part 1 - Background Information

1. Gender

○ Male

○ Female

○ Prefer not to respond

2. Age

○ 20-30

○ 31-40

○ 41-50

○ 51-60

○ 61-70

○ 71-80

3. My background is:

Please check all that apply.

☐ Operational (air traffic controller)

☐ Research-oriented (e.g. university professor, research scientist)

☐ Training-oriented (e.g.instructor, training designer)

☐ Other, please specify... ______________________

4. I have worked in/studied the following domain of ATC:

Please check all that apply.

☐ Tower

☐ Terminal

☐ Low enroute

☐ High enroute

☐ Oceanic

5. In which global region have you spent the most time working in the ATC industry?

○ United States

○ Canada

○ United Kingdom

○ Continental Europe

○ Oceania

○ Other, please specify... ______________________

6. How long have you worked with enroute ATC simulations?

○ 0-5 years

○ 6-10 years

○ 11-15 years

○ 16 years or more

7. For what purposes do you primarily use or work with enroute ATC simulation?

○ Training

○ Testing of new operational tools or concepts

○ Future ATC environment research

8. Have you previously worked with enroute ATC simulation in other ways?

○ Yes

○ No

If Yes, please describe the other ways in which you have worked with enroute ATC simulations.

☐

9. How did you hear about this survey?

○ Public website

○ Workplace

○ Personal contact

○ Other, please specify... ____________________

Introduction to Part 2 - Simulation Fidelity in the enroute ATC Domain

This section of the survey will pose questions related to your understanding of the concept of simulation fidelity, your thoughts on the usefulness of simulations of different levels of fidelity, and on the suitability of our definition of enroute ATC simulation fidelity.

REMINDER

In order to ensure consistency across all survey participants, the researchers offer the following definition of 'enroute ATC simulation':

Any environment where a user/operator actively practices providing some or all of the air traffic services. Typically this is done with some form and/or combination of tools, objects, or personnel to replicate some part of the real world task environment. Examples of different types of simulation include an instructor moving plastic airplanes around on a tabletop sector, a personal computer part-task program, and a complete reproduction of a radar controller's workstation and work environment.

NOTE: This survey is investigating human in the loop simulations and excludes fast-time simulations. Please remember this when answering the questions.

10. What are 3 key words or phrases that come to mind when you think of simulation fidelity?

11. Do you believe that simulation fidelity is a well-defined concept in the ATC domain?

○ Yes

○ No

Please explain your answer.

12. The following tables provide a number of skills and knowledge associated with enroute ATC training. We would like you to provide your level agreement with the following statements for each skill and knowledge point:

A classroom-based role playing exercise or case study is sufficient to train ab-initio students on:

	1-Strongly Disagree	2	3	4-Neutral	5	6	7-Strongly Agree
Phraseology	○	○	○	○	○	○	○
Domain-specific knowledge (e.g. aircraft characteristics, ATC rules, maps, frequencies, etc.)	○	○	○	○	○	○	○
Isolated skills (e.g. performing a handoff, issuing a clearance, etc.)	○	○	○	○	○	○	○
Visualization	○	○	○	○	○	○	○
Prioritization	○	○	○	○	○	○	○
Sector-specific characteristics	○	○	○	○	○	○	○

Multi-tasking	○	○	○	○	○	○	○

A classroom-based role playing exercise or case study is sufficient to train qualified controllers on:

	1-Strongly Disagree	2	3	4-Neutral	5	6	7-Strongly Agree
A new operational procedure (e.g. new traffic flow pattern)	○	○	○	○	○	○	○
A new operational tool (e.g. electronic flight strips)	○	○	○	○	○	○	○
Emergency situations	○	○	○	○	○	○	○

A simulated workstation with radar display and realistic communications is required to train ab-initio students on:

	1-Strongly Disagree	2	3	4-Neutral	5	6	7-Strongly Agree
Phraseology	○	○	○	○	○	○	○
Domain-specific knowledge (e.g. aircraft characteristics, ATC rules, maps, frequencies, etc.)	○	○	○	○	○	○	○
Isolated skills (e.g. performing a handoff, issuing a clearance, etc.)	○	○	○	○	○	○	○
Visualization	○	○	○	○	○	○	○
Prioritization	○	○	○	○	○	○	○
Sector-specific characteristics	○	○	○	○	○	○	○
Multi-tasking	○	○	○	○	○	○	○

A simulated workstation with radar display and realistic communications is required to train qualified controllers on:

	1-Strongly Disagree	2	3	4-Neutral	5	6	7-Strongly Agree
A new operational procedure (e.g. new traffic flow pattern)	○	○	○	○	○	○	○
A new operational tool (e.g. electronic flight strips)	○	○	○	○	○	○	○

Emergency situations	○	○	○	○	○	○	○

12. The following tables provide a number of operational concepts and human factor issues associated with enroute ATC. We would like you to provide your level agreement with the following statements for each operational concept and human factor issue:

A role-playing exercise or case study is sufficient to evaluate the acceptability of new operational concepts for the work environment such as:

	1-Strongly Disagree	2	3	4-Neutral	5	6	7-Strongly Agree
Traffic flows	○	○	○	○	○	○	○
Procedures	○	○	○	○	○	○	○
Decision support tools	○	○	○	○	○	○	○
Interface tools	○	○	○	○	○	○	○
System automation	○	○	○	○	○	○	○
Information management tools	○	○	○	○	○	○	○

A role-playing exercise or case study is sufficient to evaluate the effect of new operational concepts on human factors issues such as:

	1-Strongly Disagree	2	3	4-Neutral	5	6	7-Strongly Agree
Situation awareness	○	○	○	○	○	○	○
Transfer of training	○	○	○	○	○	○	○
Complexity	○	○	○	○	○	○	○
Decision making	○	○	○	○	○	○	○
Human-automation interaction	○	○	○	○	○	○	○

A simulated workstation with radar display and realistic communications is required to evaluate the acceptability of new operational concepts for the work environment such as:

	1-Strongly Disagree	2	3	4-Neutral	5	6	7-Strongly Agree
Traffic flows	○	○	○	○	○	○	○
Procedures	○	○	○	○	○	○	○

Decision support tools	○	○	○	○	○	○	○
Interface tools	○	○	○	○	○	○	○
System automation	○	○	○	○	○	○	○
Information management tools	○	○	○	○	○	○	○

A simulated workstation with radar display and realistic communications is required to evaluate the effect of new operational concepts on human factors issues such as:

	1-Strongly Disagree	2	3	4-Neutral	5	6	7-Strongly Agree
Situation awareness	○	○	○	○	○	○	○
Transfer of training	○	○	○	○	○	○	○
Complexity	○	○	○	○	○	○	○
Decision making	○	○	○	○	○	○	○
Human-automation interaction	○	○	○	○	○	○	○

13. In what situations do you feel a lower-fidelity simulation environment can be just as or more useful than a high-fidelity simulation environment?

Please provide as much detail as possible.

14. Have you ever seen examples where lack of fidelity in a simulation environment had significant consequences?

○ Yes

○ No

If Yes, please provide an example.

15. In your opinion, what are 3 key differences between the highest-fidelity simulation environment you have worked with and the enroute ATC work environment?

Difference 1

Difference 2

Difference 3

16. In your opinion, what parts of the enroute ATC work environment affect the fidelity experienced by someone using an enroute ATC simulation?

Please list as many as you can think of below.

The following is the enroute ATC simulation fidelity definition developed by the researchers.

[image of simulation fidelity definition identical to Figure 6.1 in Section 6.2]

The definition has three main components that are inherent to the simulation environment: physical environment, inter-personal communications, and simulation functionality. Within each component the sub-components provide a list of specific factors that comprise the main component.

The fourth fidelity component, the scenario used, can also affect the perceived fidelity of a simulation; however it is not an inherent component of a simulation environment. The scenario can change drastically depending on if it has been designed to focus on a specific skill for a new recruit, for recurrent training of an experienced controller, testing the viability of a new sector traffic flow, or for a researcher investigating workload and situation awareness.

NOTE: For detailed descriptions of each main component and their sub-components, please click on the following link http://rbhagat.uwaterloo.ca/idea/component_definitions. You may return to this page once you've finished reviewing these descriptions and continue the survey.

17. The four main components and their sub-components in the definition above accurately capture all the relevant components of the enroute ATC work environment that can affect the perceived fidelity of an enroute ATC simulation.

Strongly Disagree	Disagree	Mildly Disagree	Neutral	Mildly Agree	Agree	Strongly Agree
○	○	○	○	○	○	○

18. Would you add to, remove, or change any of the components shown above, or would you keep them as they are?

○ Keep them as they are.

○ Add to, remove, or change components.

If you selected "Add to, remove or change components.", please describe what you would do and why you felt it was necessary.

19. We would like to ask you to think about the highest fidelity simulation environment you have worked with in any context. On a scale of 1-10, with 1 being not similar in any way to the operational environment and 10 being an identical replica, where would this simulation place on that scale?

1-Not similar in any way	2	3	4	5	6	7	8	9	10-Identical replica
○	○	○	○	○	○	○	○	○	○

If you feel comfortable sharing, please tell us specifically what simulation you were thinking of when you answered Question 19.

20. What were two key limitations with this simulation environment?

Limitation 1

Limitation 2

21. To your knowledge, what is this simulation environment used for?

Please check all that apply.

☐ Training

☐ Testing new operational tools or concepts

☐ Future ATC environment research

☐ Other, please specify... ______________________

Introduction to Part 3 - Tools for Comparison of ATC Simulation Environments

This section of the survey will pose questions regarding tools for comparing and discussing different simulation environments and the usefulness of our proposed enroute ATC simulation categorization system.

A simulation categorization system allows for all the available simulations in a particular industry, in this case enroute ATC, to be classified in generic fidelity categories based on their characteristics. Categorization systems are typically generic and not associated with any specific simulators.

REMINDER

In order to ensure consistency across all survey participants, the researchers offer the following definition of 'enroute ATC simulation':

Any environment where a user/operator actively practices providing some or all of the air traffic services. Typically this is done with some form and/or combination of tools, objects, or personnel to replicate some part of the real world task environment. Examples of different types of simulation include an instructor moving plastic airplanes around on a tabletop sector, a personal computer part-task program, and a complete reproduction of a radar controller's workstation and work environment.

NOTE: This survey is investigating human in the loop simulations and excludes fast-time simulations. Please remember this when answering the questions.

22. Do you feel that the terms low, medium, and high for describing simulation fidelity are useful?

○ Yes

○ No

If Yes, why did you feel these terms were useful? ☐

If No, why did you feel these terms were not useful? ☐

Below is a level based simulation categorization system used by the FAA to help differentiate between full flight simulators.

Full Flight Simulators (FFS)

FAA FFS Level A - A motion system is required with at least three degrees of freedom. Airplanes only.

FAA FFS Level B - Requires three axis motion and a higher-fidelity aerodynamic model than does Level A. The lowest level of helicopter flight simulator.

FAA FFS Level C - Requires a motion platform with all six degrees of freedom. Also lower transport delay (latency) over levels A & B. The visual system must have an outside-world horizontal field of view of at least 75 degrees for each pilot.

FAA FFS Level D - The highest level of FFS qualification currently available. Requirements are for Level C with additions. The motion platform must have all six degrees of freedom, and the visual system must have an outside-world horizontal field of view of at least 150 degrees, with aCollimated (distant focus) display. Realistic sounds in the cockpit are required, as well as a number of special motion and visual effects.

*Taken from: US Federal Aviation Administration (FAA): 14 CFR Part 60, Appendices A and C

23. Do you feel that a standardized categorization system similar to the one above but adapted for enroute ATC simulation environments is required?

○ Yes

○ No

Please explain your answer.

☐

The following is the categorization system developed by the researchers to identify the 5 main levels of simulation fidelity used for training, testing new operational concepts, and future ATC environment research in the domain of enroute ATC.

[Table 7.1 from section 7.2 presented here]

24. Please provide examples of simulations that would fit in each category that you've used before or have heard of being used.

Level A ☐

Level B ☐

Level C	☐
Level D	☐
Level E	☐

25. Using the above categorization system, what is the minimum level of simulation fidelity required to train ab-initio students on:

	Level A	Level B	Level C	Level D	Level E	Should be trained in the operational environment
Phraseology	○	○	○	○	○	○
Domain-specific knowledge (e.g. aircraft characteristics, ATC rules, maps, frequencies, etc.)	○	○	○	○	○	○
Skills in isolation (e.g. performing a handoff, issuing a clearance, etc.)	○	○	○	○	○	○
Visualization	○	○	○	○	○	○
Prioritization	○	○	○	○	○	○
Sector-specific characteristics	○	○	○	○	○	○
Multi-tasking	○	○	○	○	○	○

Using the above categorization system, what is the minimum level of simulation fidelity required to train qualified controllers on:

	Level A	Level B	Level C	Level D	Level E	Should be trained in the operational environment
A new operational procedure (e.g. new traffic flow pattern)	○	○	○	○	○	○
A new operational tool (e.g. electronic flight strips)	○	○	○	○	○	○

Emergency situations	○	○	○	○	○	○

25. Using the above categorization system, what is the minimum level of simulation fidelity required to evaluate and approve new operational concepts for regular use such as:

	Level A	Level B	Level C	Level D	Level E	Should be evaluated in the operational environment
Traffic flows	○	○	○	○	○	○
Procedures	○	○	○	○	○	○
Decision support tools	○	○	○	○	○	○
Interface tools	○	○	○	○	○	○
System automation	○	○	○	○	○	○
Information management tools	○	○	○	○	○	○

Using the above categorization system, what is the minimum level of simulation fidelity required to evaluate the effects of new operational concepts on human factors issues such as:

	Level A	Level B	Level C	Level D	Level E	Should be evaluated in the operational environment
Situation awareness	○	○	○	○	○	○
Transfer of training	○	○	○	○	○	○
Complexity	○	○	○	○	○	○
Decision making	○	○	○	○	○	○
Human-automation interaction	○	○	○	○	○	○

26. How useful was the categorization system for determining what type of simulation environment is best suited for each task in Question 25?

Not at all useful	Not very useful	Somewhat useful	Very useful
○	○	○	○

Please explain your answer.

27. Would you keep the categorization system the same or would you make changes?

○ Keep it the same.

○ Make changes.

If you selected "Make changes.", what specific changes would you make? Please be as detailed as possible.

☐

Appendix C

Overall Survey Participant Demographics

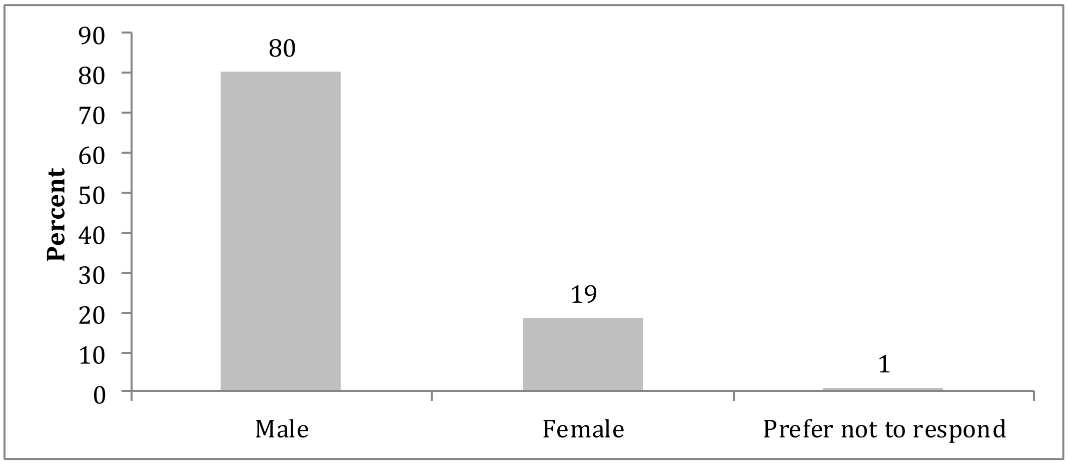

Figure C.1 Gender distribution of survey participants. (N=86)

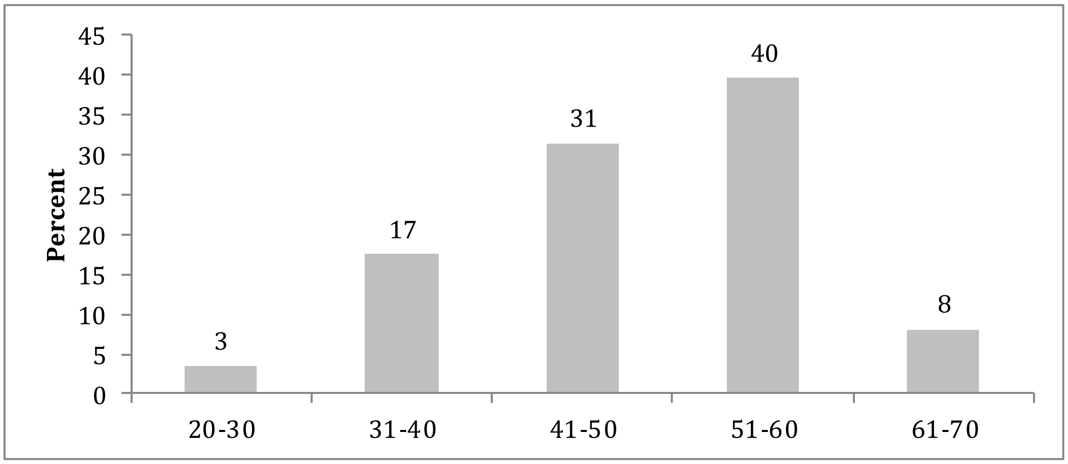

Figure C.2 Age distribution of survey participants. (N=86)

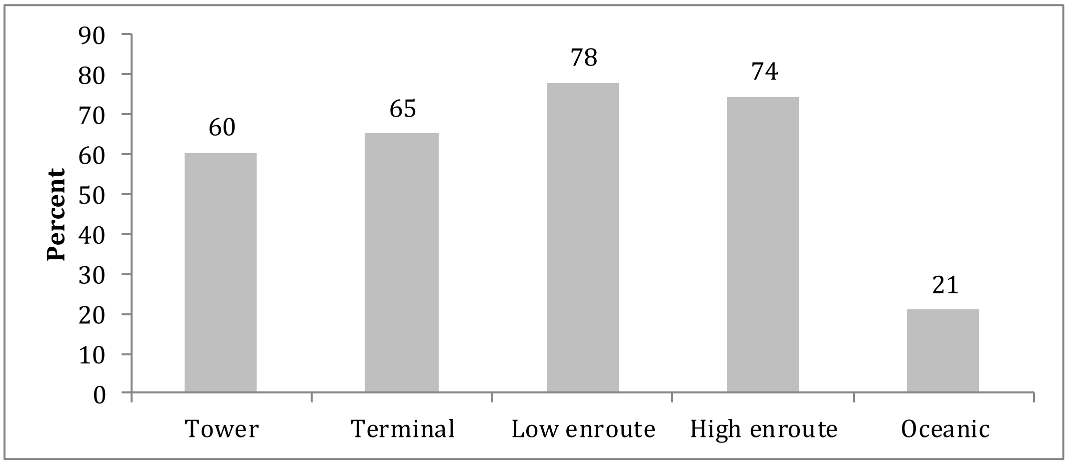

Figure C.3 Area of work/research experience of survey participants. Each column represents a percentage of N. (N=86)

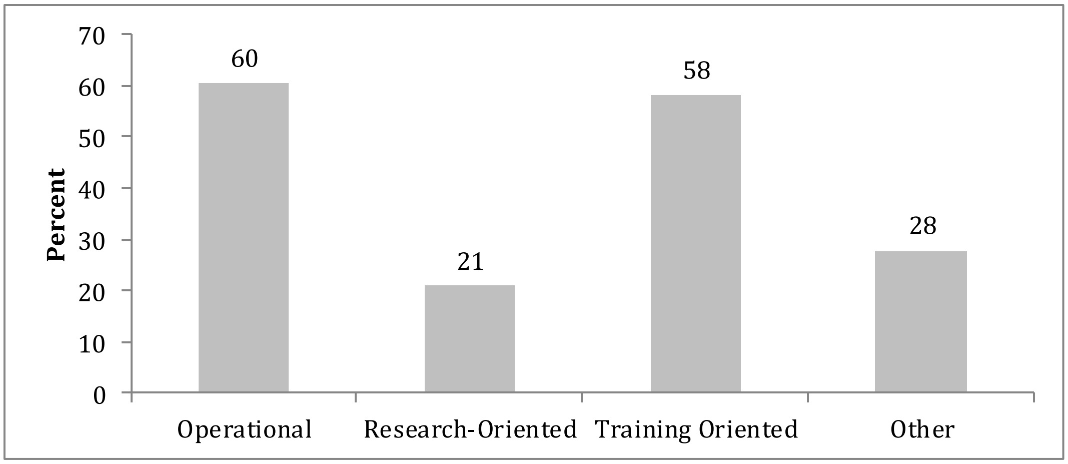

Figure C.4 Background of survey participants. Each column represents a percentage of N. (N=86)

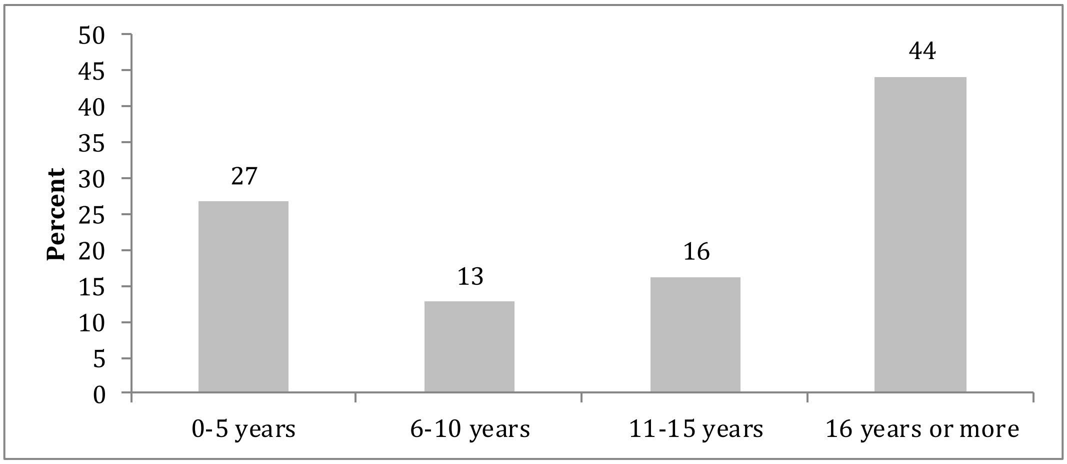

Figure C.5 Distribution of survey participants' years of experience working with simulation. (N=86)

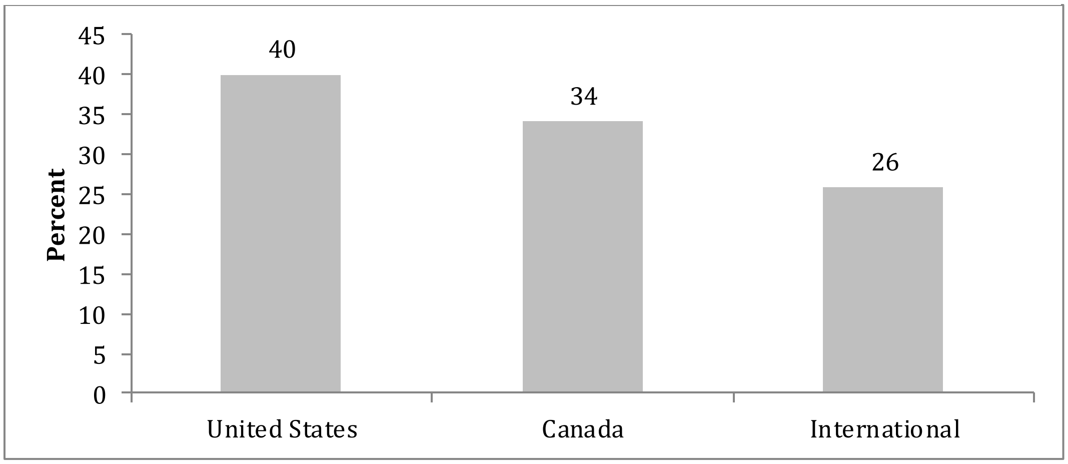

Figure C.6 Distribution of survey participants' global region where they have spent the most time working. (N=85)

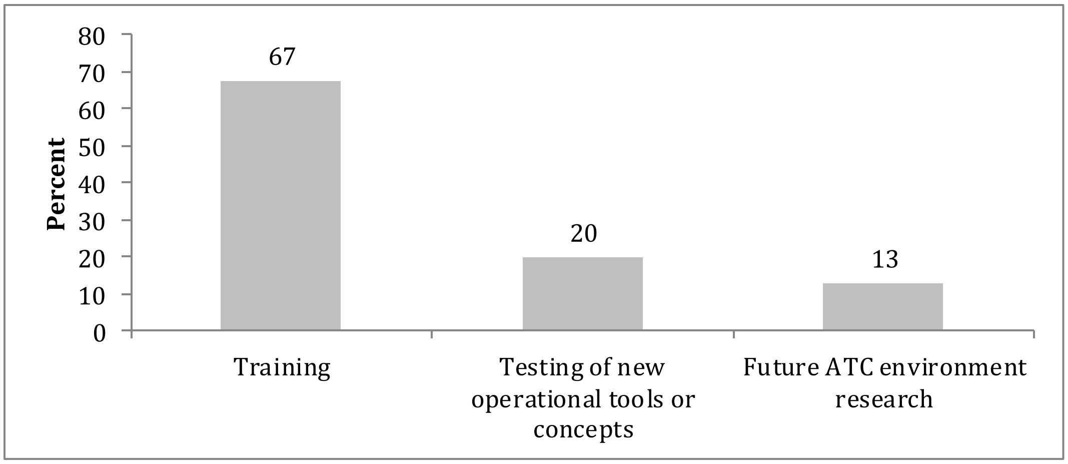

Figure C.7 Distribution of survey participants' primary use of simulation. (N=86)

Appendix D

Demographic Sub-group Comparisons of Survey Participant Fidelity Components

Table D.1 Top ten coded fidelity component response frequencies for all survey participants with gender demographic group comparison.

Fidelity Components	Response frequency (% of N)		
	Overall (N=73)	Male (N=57)	Female (N=15)
Communications	62	53	93
Equipment	42	42	47
Environment	32	28	47
Aircraft performance	30	26	47
System participants	29	33	13
Unpredictability	29	23	53
Traffic	23	21	33
Weather	21	16	40
Automation	19	21	13
Operational stress	11	12	7

Table D.2 Top ten coded fidelity component response frequencies for all survey participants with survey participants' primary use of simulation demographic group comparison.

Fidelity Components	Response frequency (% of N)			
	Overall (N=73)	Training (N=48)	Testing (N=14)	Research (N=11)
Communications	62	58	64	73
Equipment	42	40	50	45
Environment	32	31	43	18
Aircraft performance	30	35	29	9
System participants	29	35	21	9
Unpredictability	29	38	7	18
Traffic	23	19	36	27
Weather	21	29	0	9
Automation	19	21	7	27
Operational stress	11	13	7	9

Table D.3 Top ten coded fidelity component response frequencies for all survey participants with survey participants' years of experience working with simulation demographic group comparison.

Fidelity Components	Response frequency (% of N)				
	Overall (N=73)	0-5 years (N=17)	6-10 years (N=11)	11-15 years (N=12)	16 years+ (N=33)
Communications	62	59	64	67	61
Equipment	42	29	55	50	42
Environment	32	24	27	17	42
Aircraft performance	30	24	36	33	30
System participants	29	12	27	42	33
Unpredictability	29	41	36	25	21
Traffic	23	41	27	8	18
Weather	21	18	27	25	18
Automation	19	12	9	33	21
Operational stress	11	24	9	8	6

Appendix E

Simulation Environment Categorization System

Secondary Validation Exercise List of Papers

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. International Journal of Industrial Ergonomics, 36(7), 623–636. DOI: 10.1016/j.ergon.2006.04.002

Allendoerfer, K. R., Truitt, T. R., & Willems, B. (2010). Evaluations of Data Communications in Tower, TRACON, and En Route Air Traffic Control. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54(1), 116–120. DOI: 10.1177/154193121005400126

Billinghurst, S. S., Morgan, C., Rorie, R. C., Kiken, a., Bacon, L. P., Vu, K.-P. L., … Battiste, V. (2011). Should Students Learn General Air Traffic Management Skills Before NextGen Tools? Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55(1), 128–132. DOI: 10.1177/1071181311551027

Boag, C., Neal, A., Loft, S., & Halford, G. S. (2006). An analysis of relational complexity in an air traffic control conflict detection task. Ergonomics, 49(14), 1508–26. DOI: 10.1080/00140130600779744

Camp, G., Paas, F., Rikers, R., & van Merrienboer, J. (2001). Dynamic problem selection in air traffic control training: a comparison between performance, mental effort and mental efficiency. Computers in Human Behavior, 17(5-6), 575–595. DOI: 10.1016/S0747-5632(01)00028-0

Fothergill, S., Loft, S., & Neal, A. (2009). ATC-lab(Advanced): an air traffic control simulator with realism and control. Behavior Research Methods, 41(1), 118–27. DOI: 10.3758/BRM.41.1.118

Fothergill, S., & Neal, a. (2008). The Effect of Workload on Conflict Decision Making Strategies in Air Traffic Control. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52(1), 39–43. DOI: 10.1177/154193120805200110

Fothergill, S., & Neal, a. (2013). Conflict-Resolution Heuristics for En Route Air Traffic Management. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57(1), 71–75. DOI: 10.1177/1541931213571018

Galster, S. M., Duley, J. A., Masalonis, a. J., & Parasuraman, R. (2001). Air Traffic Controller Performance and Workload Under Mature Free Flight : Conflict Detection and Resolution of Aircraft Self-Separation. The International Journal of Aviation Psychology, 11(1), 71–93. DOI: 10.1207/S15327108IJAP1101

Gronlund, S. D., Dougherty, M. R. P., Durso, F. T., Canning, J. M., & Scott, H. (2009). Planning in Air Traffic Control : Impact of Problem Type. The International Journal of Aviation Psychology, 15(3), 269–293. DOI: 10.1207/s15327108ijap1503

Hannah, S. D., & Neal, A. (2014). On-the-Fly Scheduling as a Manifestation of Partial-Order Planning and Dynamic Task Values. Human Factors: The Journal of the Human Factors and Ergonomics Society, 1–20. DOI: 10.1177/0018720814525629

Hauland, G. (2008). Measuring Individual and Team Situation Awareness During Planning Tasks in Training of En Route Air Traffic Control. The International Journal of Aviation Psychology, 18(3), 290–304. DOI: 10.1080/10508410802168333

Higgins, J. S., Willems, B., Johnson, D. R., & Zingale, C. M. (2012). A Human Factors and Electromyographic Evaluation of Proposed Pointing Devices for Air Traffic Controllers. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56(1), 36–40. DOI: 10.1177/1071181312561028

Homola, J., Prevot, T., Mercer, J., Mainini, M., & Cabrall, C. (2009). Human/automation response strategies in tactical conflict situations. Digital Avionics Systems Conference, 2009. DASC '09. IEEE/AIAA 28th, pp.5.B.6–1,5.B.6–15.

Hyun, S. H., Kim, K. T., Song, S. M., Yoon, J. S., Lee, J. I., Cho, S. M., & Sohn, Y. W. (2006). Effects of Expertise and Situation Complexity on Visual Attention and Action Planning for Air Traffic Control. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(3), 420–423. DOI: 10.1177/154193120605000344

Jha, P. D., Bisantz, A. M., Parasuraman, R., & Drury, C. G. (2011). Air Traffic Controllers' Performance in Advance Air Traffic Management System: Part I—Performance Results. The International Journal of Aviation Psychology, 21(3), 283–305. DOI: 10.1080/10508414.2011.582456

Kaber, D. B., Perry, C. M., Segall, N., & Sheik-Nainar, M. a. (2007). Workload State Classification With Automation During Simulated Air Traffic Control. The International Journal of Aviation Psychology, 17(4), 371–390. DOI: 10.1080/10508410701527860

Kiken, A., Rorie, R. C., Bacon, L. P., Billinghurst, S., Kraut, J. M., Strybel, T. Z., … Battiste, V. (2011). Effect of ATC Training with NextGen Tools and Online Situation Awareness and Workload Probes on Operator Performance. Human Interface and the Management of Information. Interacting with Information, 483–492. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-21669-5_57

Koy, A., & Yeo, G. (2008). BIS Sensitivity, Negative Affect and Performance: Dynamic and Multilevel Relationships. Human Performance, 21(2), 198–225. DOI: 10.1080/08959280801917826

Kraut, J. M., Kiken, A., Billinghurst, S., Morgan, C. A., Strybel, T. Z., Chiappe, D., & Vu, K. L. (2011). Effects of Data Communications Failure on Air Traffic Controller Sector Management Effectiveness, Situation Awareness , and Workload. In M. J. Salvendy, G., Smith (Ed.), Human Interface, Part II, HCII 2011, Lecture Notes in Computer Science, 6772 (pp. 493–499). Springer-Verlag.

Lee, P. U., & Prevot, T. (2012). Prediction of Traffic Complexity and Controller Workload in Mixed Equipage NextGen Environments. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56(1), 100–104. DOI: 10.1177/1071181312561041

Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A theory and model of conflict detection in air traffic control: incorporating environmental constraints. Journal of Experimental Psychology. Applied, 15(2), 106–24. DOI: 10.1037/a0016118

Loft, S., Finnerty, D., & Remington, R. W. (2011). Using Spatial Context to Support Prospective Memory in Simulated Air Traffic Control. Human Factors: The Journal of the Human Factors and Ergonomics Society, 53(6), 662–671. DOI: 10.1177/0018720811421783

Loft, S., Hill, A., Neal, A., Humphreys, M., & Yeo, G. (2004). ATC-lab: an air traffic control simulator for the laboratory. Behavior Research Methods, Instruments, & Computers, 36(2), 331–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15354699

Loft, S., Humphreys, M., & Neal, A. (2004). The influence of memory for prior instances on performance in a conflict detection task. Journal of Experimental Psychology: Applied, 10(3), 173–87. DOI: 10.1037/1076-898X.10.3.173

Loft, S., Neal, A., & Humphreys, M. S. (2007). The development of a general associative learning account of skill acquisition in a relative arrival-time judgment task. Journal of Experimental Psychology. Human Perception and Performance, 33(4), 938–59. DOI: 10.1037/0096-1523.33.4.938

Loft, S., Pearcy, B., & Remington, R. W. (2011). Varying the Complexity of the Prospective Memory Decision Process in an Air Traffic Control Simulation. Zeitschrift Für Psychologie / Journal of Psychology, 219(2), 77–84. DOI: 10.1027/2151-2604/a000051

Loft, S., Smith, R. E., & Bhaskara, a. (2009). Designing Memory Aids to Facilitate Intentions to Deviate from Routine in an Air Traffic Control Simulation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 53(1), 56–60. DOI: 10.1177/154193120905300113

Loft, S., Smith, R. E., & Bhaskara, A. (2011). Prospective memory in an air traffic control simulation: external aids that signal when to act. Journal of Experimental Psychology. Applied, 17(1), 60–70. DOI: 10.1037/a0022845

Loft, S., Smith, R. E., & Remington, R. W. (2013). Minimizing the disruptive effects of prospective memory in simulated air traffic control. Journal of Experimental Psychology. Applied, 19(3), 254–65. DOI: 10.1037/a0034141

Loh, V., Andrews, S., Hesketh, B., & Griffin, B. (2012). The Moderating Effect of Individual Differences in Error-Management Training: Who Learns From Mistakes? Human Factors: The Journal of the Human Factors and Ergonomics Society, 55(2), 435–448. DOI: 10.1177/0018720812451856

Macwilliams, P., & Porter, D. (2007). An Assessment of a Controller Aid for Merging and Sequencing Traffic on Performance-Based Arrival Routes (pp. 1–10). McLean, Virginia. Retrieved from http://www.mitre.org/sites/default/files/pdf/07_0158.pdf

Marchitto, M., Di Stasi, L., & Canas, J. (2012). Ocular Movements Under Taskload Manipulations : Influence of Geometry on Saccades in Air Traffic Control Simulated Tasks. Human Factors and Ergonomics in Manufacturing & Service Industries, 22(5), 407–419. DOI: 10.1002/hfm

Masalonis, a. J., & Parasuraman, R. (2003). Effects of Situation-Specific Reliability on Trust and Usage of Automated Air Traffic Control Decision Aids. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 47(3), 533–537. DOI: 10.1177/154193120304700359

Mercer, J., Callantine, T. J., Lee, P. U., Prevot, T., & Palmer, E. (2005). An evaluation of airborne spacing in the terminal area. Digital Avionics Systems Conference, 2005. DASC 2005. The 24th, 1, pp.2.A.3,21–10.

Metzger, U., & Parasuraman, R. (2001). The Role of the Air Traffic Controller in Future Air Traffic Management: An Empirical Study of Active Control versus Passive Monitoring. Human Factors: The Journal of the Human Factors and Ergonomics Society, 43(4), 519–528. DOI: 10.1518/001872001775870421

Metzger, U., & Parasuraman, R. (2005). Automation in Future Air Traffic Management: Effects of Decision Aid Reliability on Controller Performance and Mental Workload. Human Factors: The Journal of the Human Factors and Ergonomics Society, 47(1), 35–49. DOI: 10.1518/0018720053653802

Metzger, U., & Parasuraman, R. (2006). Effects of Automated Conflict Cuing and Traffic Density on Air Traffic Controller Performance and Visual Attention in a Datalink Environment. The International Journal of Aviation Psychology, 16(4), 343–362. DOI: 10.1207/s15327108ijap1604

Neal, a., & Kwantes, P. J. (2009). An Evidence Accumulation Model for Conflict Detection Performance in a Simulated Air Traffic Control Task. Human Factors: The Journal of the Human Factors and Ergonomics Society, 51(2), 164–180. DOI: 10.1177/0018720809335071

Nunes, A. (2004). An Empirical Investigation of the Effects of Controller Experience on Conflict Detection Ability under Free Flight. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48(1), 161–165. DOI: 10.1177/154193120404800135

Nunes, A., & Kirlik, A. (2005). An Empirical Study of Calibration in Air Traffic Control Expert Judgment. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49(3), 422–426. DOI: 10.1177/154193120504900344

Oprins, E., Zwaaf, D., Eriksson, F., & Roe, R. (2009). Impact of future time-based operations on situation awareness of air traffic controllers. In Eighth USA/Europe Air Traffic Management Research and Development Seminar (pp. 1–10). Napa, CA.

Paubel, P.-V., Averty, P., & Raufaste, E. (2013). Effects of an Automated Conflict Solver on the Visual Activity of Air Traffic Controllers. The International Journal of Aviation Psychology, 23(2), 181–196. DOI: 10.1080/10508414.2013.772841

Pop, V. L., Stearman, E. J., Kazi, S., & Durso, F. T. (2012). Using Engagement to Negate Vigilance Decrements in the NextGen Environment. International Journal of Human-Computer Interaction, 28(2), 99–106. DOI: 10.1080/10447318.2012.634759

Prevot, T. (2002). Exploring the Many Perspectives of Distributed Air Traffic Management : The Multi Aircraft Control System MACS. Proceedings of the HCI-Aero, 2002, 149–154.

Prevot, T., Homola, J. R., Martin, L. H., Mercer, J. S., & Cabrall, C. D. (2012). Toward Automated Air Traffic Control—Investigating a Fundamental Paradigm Shift in Human/Systems Interaction. International Journal of Human-Computer Interaction, 28(2), 77–98. DOI: 10.1080/10447318.2012.634756

Prevot, T., Mercer, J., Martin, L. H., Homola, J., Cabrall, C., & Brasil, C. (2010). Functional Allocation for Ground-Based Automated Separation Assurance in NextGen (Vol. 12, pp. 3–10). Retrieved from http://ntrs.nasa.gov/search.jsp?R=20110008669

Rantanen, E. M. (2009). Measurement of Temporal Awareness in Air Traffic Control. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 53(1), 6–10. DOI: 10.1177/154193120905300103

Rantanen, E. M., McCarley, J. S., & Xu, X. (2002). The Impact of Communication Delays on Air Traffic Controllers' Vectoring Performance. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 46(1), 56–60. DOI: 10.1177/154193120204600112

Rantanen, E. M., Mccarley, J. S., & Xu, X. (2004). Time Delays in Air Traffic Control Communication Loop : Effect on Controller Performance and Workload. The International Journal of Aviation Psychology, 14(4), 369–394. DOI: 10.1207/s15327108ijap1404

Rantanen, E. M., & Nunes, A. (2005). Hierarchical Conflict Detection in Air Traffic Control. The International Journal of Aviation Psychology, 15(4), 339–362. DOI: 10.1207/s15327108ijap1504

Rantanen, E. M., Yang, J., & Yin, S. (2006). Comparison of Pilots' and Controllers' Conflict Resolution Maneuver Preferences. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(1), 16–19. DOI: 10.1177/154193120605000104

Rorie, R. C., Kiken, A., Morgan, C., Billinghurst, S., Morales, G., Monk, K., … Strybel, T. (2011). A Preliminary Investigation of Training Order for Introducing NextGen Tools. Human Interface, Part II, 526–533.

Rovira, E., & Parasuraman, R. (2006). Controller Performance and Attention Allocation in Future Air Traffic Management: Effects of Pilot Intent Information. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(1), 1–5. DOI: 10.1177/154193120605000101

Rovira, E., & Parasuraman, R. (2010). Transitioning to Future Air Traffic Management: Effects of Imperfect Automation on Controller Attention and Performance. Human Factors: The Journal of the Human Factors and Ergonomics Society, 52(3), 411–425. DOI: 10.1177/0018720810375692

Salden, R. J. C. M., Paas, F., Broers, N. J., & van Merrienboer, J. J. G. (2004). Mental Effort and Performance as Determinants for the Dynamic Selection of Learning Tasks in Air Traffic Control Training. Instructional Science, 32, 153–172.

Sethumadhavan, A. (2009). Effects of Automation Types on Air Traffic Controller Situation Awareness and Performance. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 53(1), 1–5. DOI: 10.1177/154193120905300102

Sethumadhavan, A. (2011). Effects of First Automation Failure on Situation Awareness and Performance in an Air Traffic Control Task. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55(1), 350–354. DOI: 10.1177/1071181311551072

Sethumadhavan, A., & Durso, F. T. (2009). Selection in Air Traffic Control: Is Nonradar Training a Predictor of Radar Performance? Human Factors: The Journal of the Human Factors and Ergonomics Society, 51(1), 21–34. DOI: 10.1177/0018720808329845

Shively, R. J., Vu, K.-P. L., & Buker, T. J. (2013). Unmanned Aircraft System Response to Air Traffic Control Clearances: Measured Response. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57(1), 31–35. DOI: 10.1177/1541931213571009

Sollenberger, R. L., & Hale, M. (2011). Human-in-the-Loop Investigation of Variable Separation Standards in the En Route Air Traffic Control Environment. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55(1), 66–70. DOI: 10.1177/1071181311551014

Stankovic, S., Loft, S., Rantanen, E., & Ponomarenko, N. (2011). Individual Differences in the Effect of Vertical Separation on Conflict Detection in Air Traffic Control. The International Journal of Aviation Psychology, 21(4), 325–342. DOI: 10.1080/10508414.2011.606744

Strybel, T. Z., Minakata, K., Nguyen, J., Pierce, R., & Vu, K. L. (2009). Optimizing Online Situation Awareness Probes in Air Traffic Management Tasks. In M. J. Smith & G. Salvendy (Eds.), Human Interface and the Management of Information. Information and Interaction Lecture Notes in Computer Science (pp. 845–854). Springer-Verlag.

Van de Merwe, K., Oprins, E., Eriksson, F., & van der Plaat, A. (2012). The Influence of Automation Support on Performance, Workload, and Situation Awareness of Air Traffic Controllers. The International Journal of Aviation Psychology, 22(2), 120–143.

Vu, K.-P. L., Silva, H., Ziccardi, J., Morgan, C. a., Morales, G., Grigoleit, T., … Battiste, V. (2012). How Does Reliance on Automated Tools During Learning Influence Students' Air Traffic Management Skills When the Tools Fail? Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56(1), 16–20. DOI: 10.1177/1071181312561024

Vuckovic, a., Kwantes, P., & Neal, a. (2011). A dynamic model of decision making in ATC: Adaptation of criterion across angle and time. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55(1), 330–334. DOI: 10.1177/1071181311551068

Vuckovic, A., Sanderson, P., Neal, A., Gaukrodger, S., & Wong, B. L. W. (2013). Relative Position Vectors: An Alternative Approach to Conflict Detection in Air Traffic Control. Human Factors: The Journal of the Human Factors and Ergonomics Society, 55(5), 946–964. DOI: 10.1177/0018720813481803

Weber, B., Oberheid, H., & Papenfuss, a. (2013). Visual Aids for Merging Mixed Arrival Traffic: Effects on Controllers' Performance, Situation Awareness, and Mental Workload. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57(1), 630–634. DOI: 10.1177/1541931213571135

Weiland, M. Z. (2010). Field Evaluation of Advanced Training Technologies in Terminal Air Traffic Control. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54(27), 2257–2261. DOI: 10.1177/154193121005402708

Wiggins, M. W., & Glass, R. (2008). A Comparison between Cue-Based, Exemplar-Based, and Production-Based Training in an Air Traffic Control Scenario. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52(26), 2042–2046. DOI: 10.1177/154193120805202602

Wilson, G. F., & Russell, C. a. (2003). Operator Functional State Classification Using Multiple Psychophysiological Features in an Air Traffic Control Task. Human Factors: The Journal of the Human Factors and Ergonomics Society, 45(3), 381–389. DOI: 10.1518/hfes.45.3.381.27252

Xiao, T., Sanderson, P. M., Mooij, M., & Fothergill, S. (2008). Work Domain Analysis for Assessing Simulated Worlds for ATC Studies. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52(4), 277–281. DOI: 10.1177/154193120805200417

Yang, J., Rantanen, E. M., & Zhang, K. (2009). The Impact of Time Efficacy on Air Traffic Controller Situation Awareness and Mental Workload. The International Journal of Aviation Psychology, 20(1), 74–91. DOI: 10.1080/10508410903416037

Yeo, G. B., Frederiks, E. R., Kiewitz, C., & Neal, A. (2013). A dynamic, self-regulatory model of affect and performance: Interactions between states, traits and task demands. Motivation and Emotion, 38(3), 429–443. DOI: 10.1007/s11031-013-9376-z

Yeo, G., Loft, S., Xiao, T., & Kiewitz, C. (2009). Goal orientations and performance: differential relationships across levels of analysis and as a function of task demands. The Journal of Applied Psychology, 94(3), 710–26. DOI: 10.1037/a0015044

Yeo, G. B., & Neal, A. (2004). A multilevel analysis of effort, practice, and performance: effects; of ability, conscientiousness, and goal orientation. The Journal of Applied Psychology, 89(2), 231–47. DOI: 10.1037/0021-9010.89.2.231

Yeo, G. B., & Neal, A. (2006). An examination of the dynamic relationship between self-efficacy and performance across levels of analysis and levels of specificity. The Journal of Applied Psychology, 91(5), 1088–101. DOI: 10.1037/0021-9010.91.5.1088

Yeo, G., & Neal, A. (2008). Subjective cognitive effort: a model of states, traits, and time. The Journal of Applied Psychology, 93(3), 617–31. DOI: 10.1037/0021-9010.93.3.617

Appendix F

Full Graphs of Survey Participants' Perceptions Regarding Use of Simulation

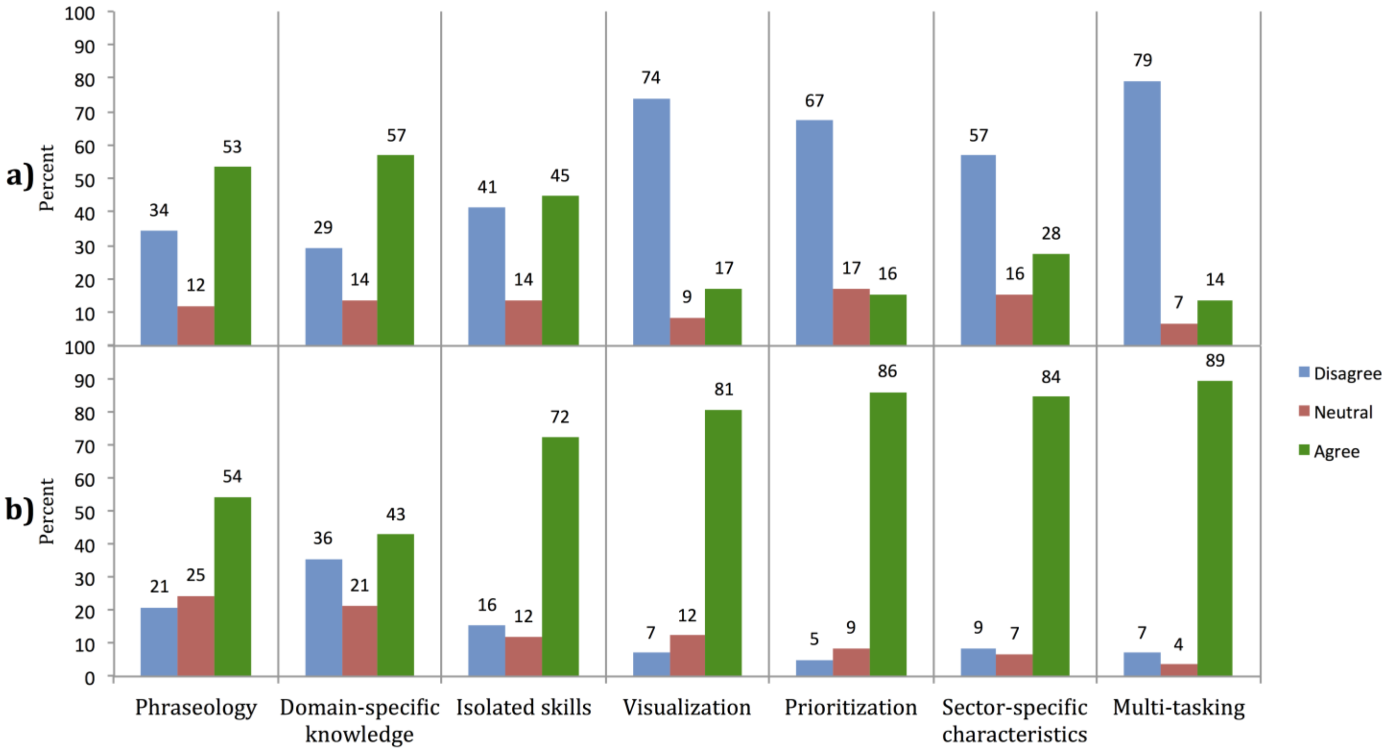

Figure F.1 Survey participant beliefs regarding selection of simulation environment for ab-initio training
a) <u>Classroom-based simulation</u> is capable of training ab-initio students on skill X. (N=58)
b) <u>Workstation simulation</u> is required to train ab-initio students on skill X. (N=56-58)

c)

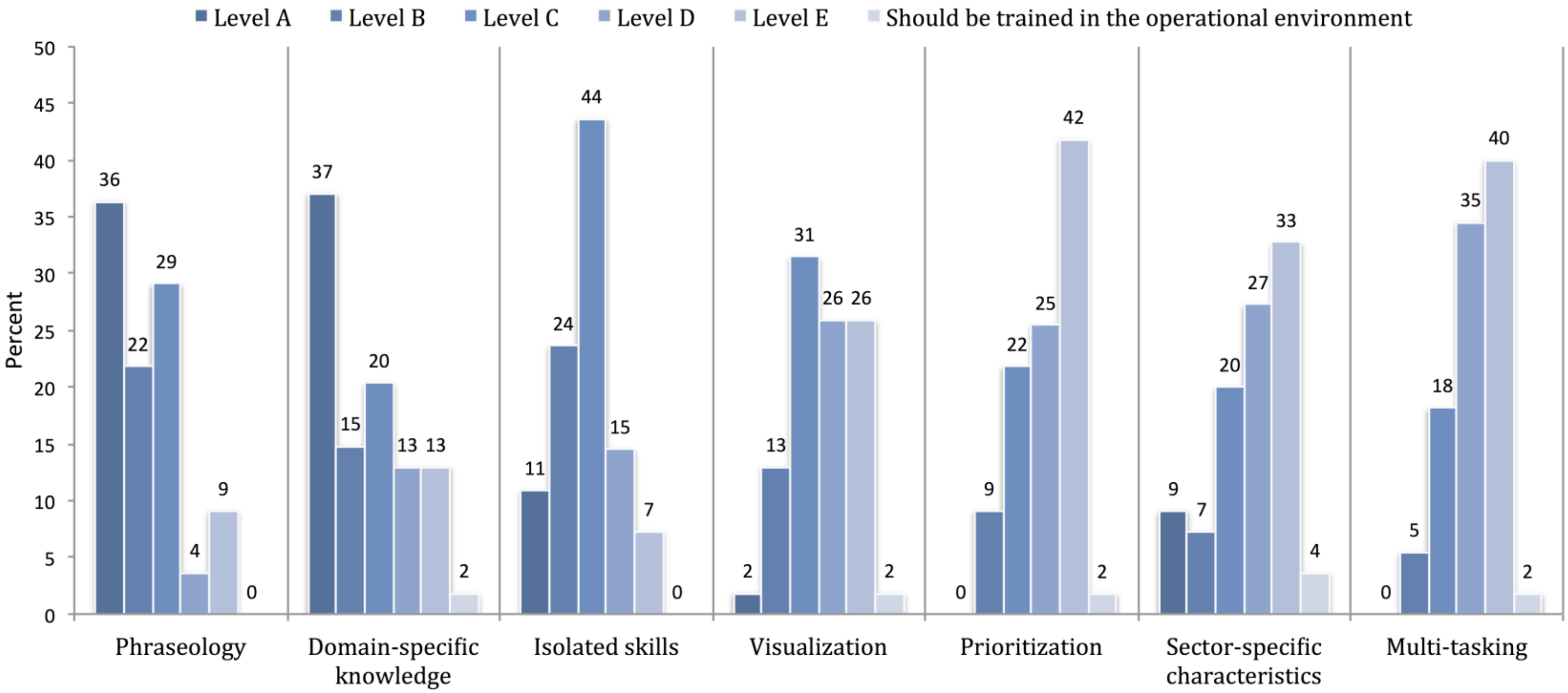

Figure F.1 Survey participant beliefs regarding selection of simulation environment for ab-initio training.

c) Based on the categorization system, what is the minimum level of simulation fidelity required to train ab-initio students in *X*. (N=54-55)

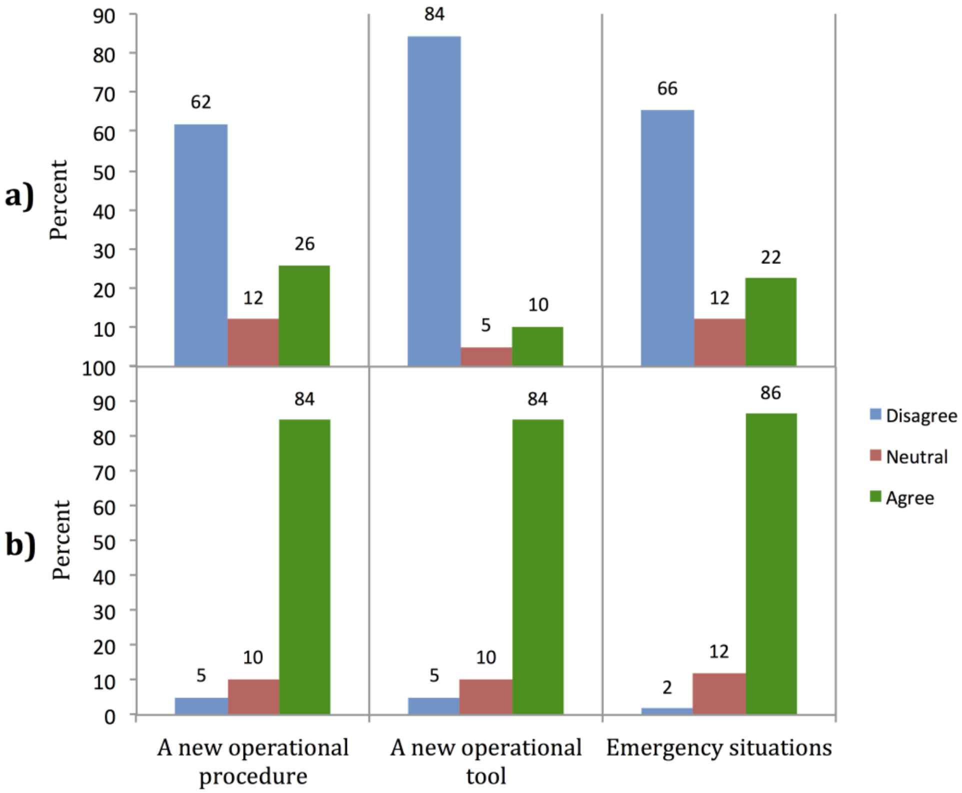

Figure F.2 Survey participant beliefs regarding selection of simulation environment for recurrent training.

a) <u>Classroom-based simulation</u> is capable of training qualified controllers on skill X. (N=58)

b) <u>Workstation simulation</u> is required to train qualified controllers on skill X. (N=58)

c)

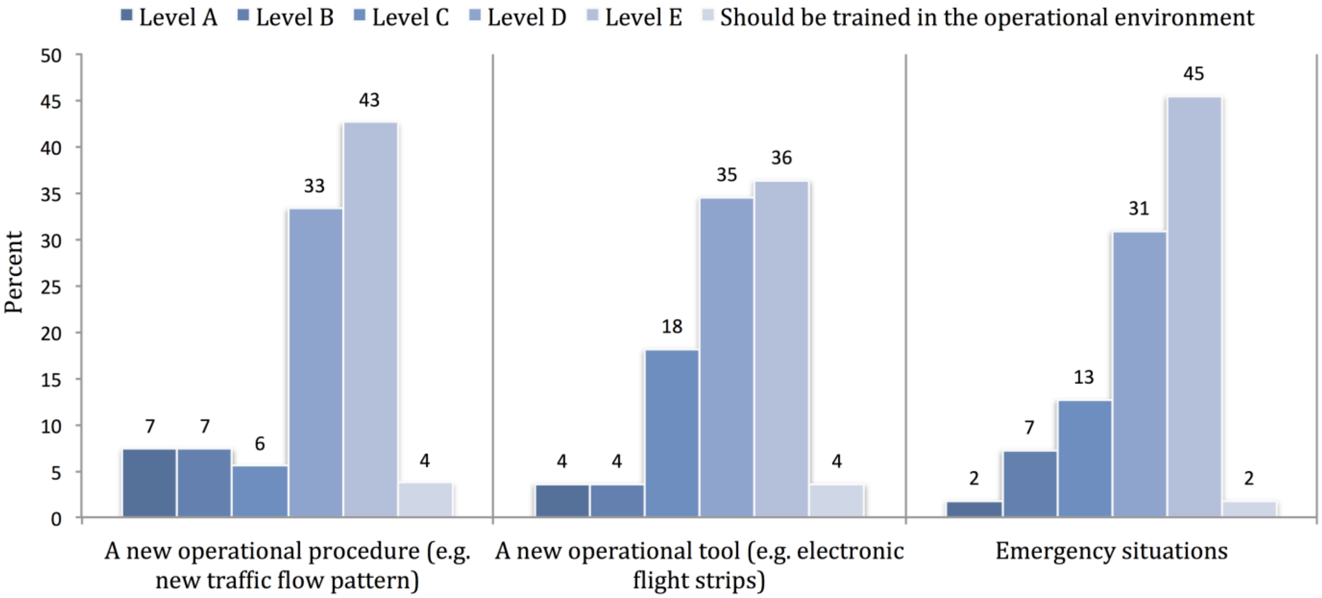

Figure F.2 Survey participant beliefs regarding selection of simulation environment for recurrent training.

c) Based on the categorization system, what is the minimum level of simulation fidelity required to train qualified controllers in *X*. (N=54-55)

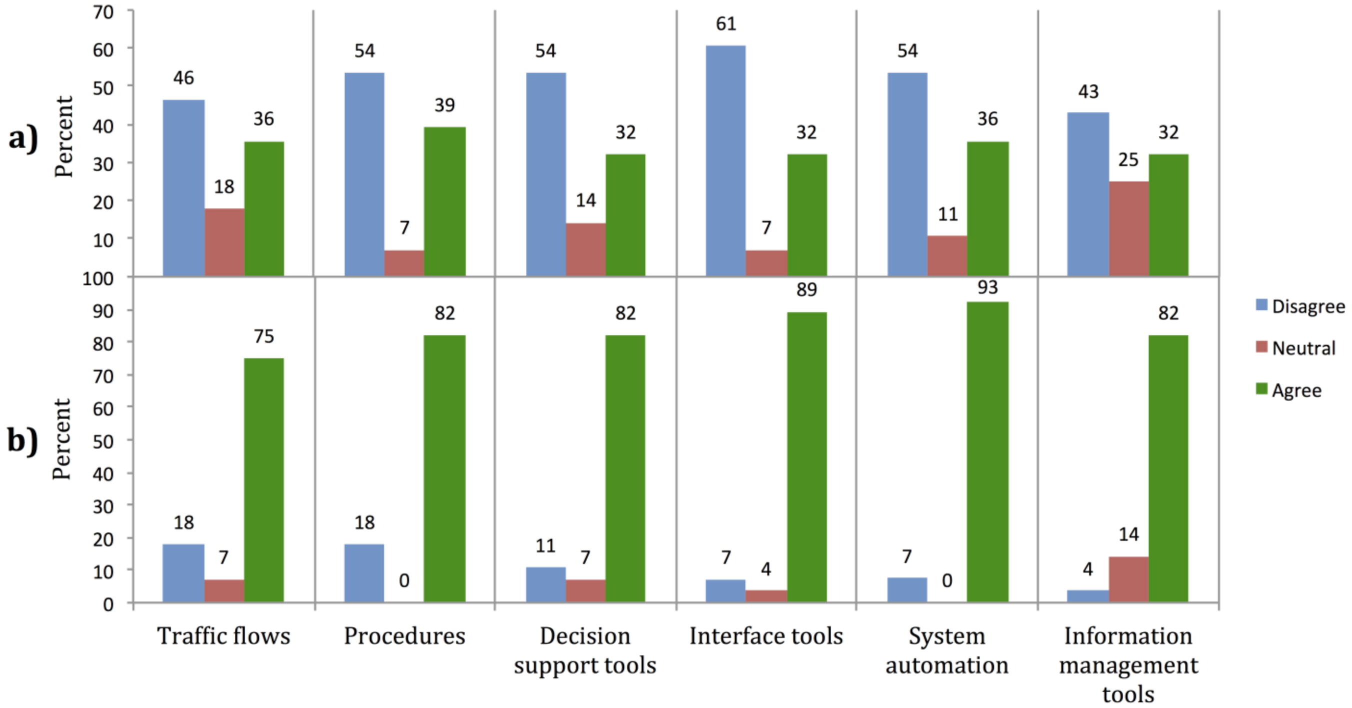

Figure F.3 Survey participant beliefs regarding selection of simulation environment for testing new operational concepts.

a) <u>Classroom-based simulation</u> is capable of evaluating X as a viable new operational concept. (N=28)

b) <u>Workstation simulation</u> is required to evaluate X as a viable new operational concept. (N=27-28)

c)

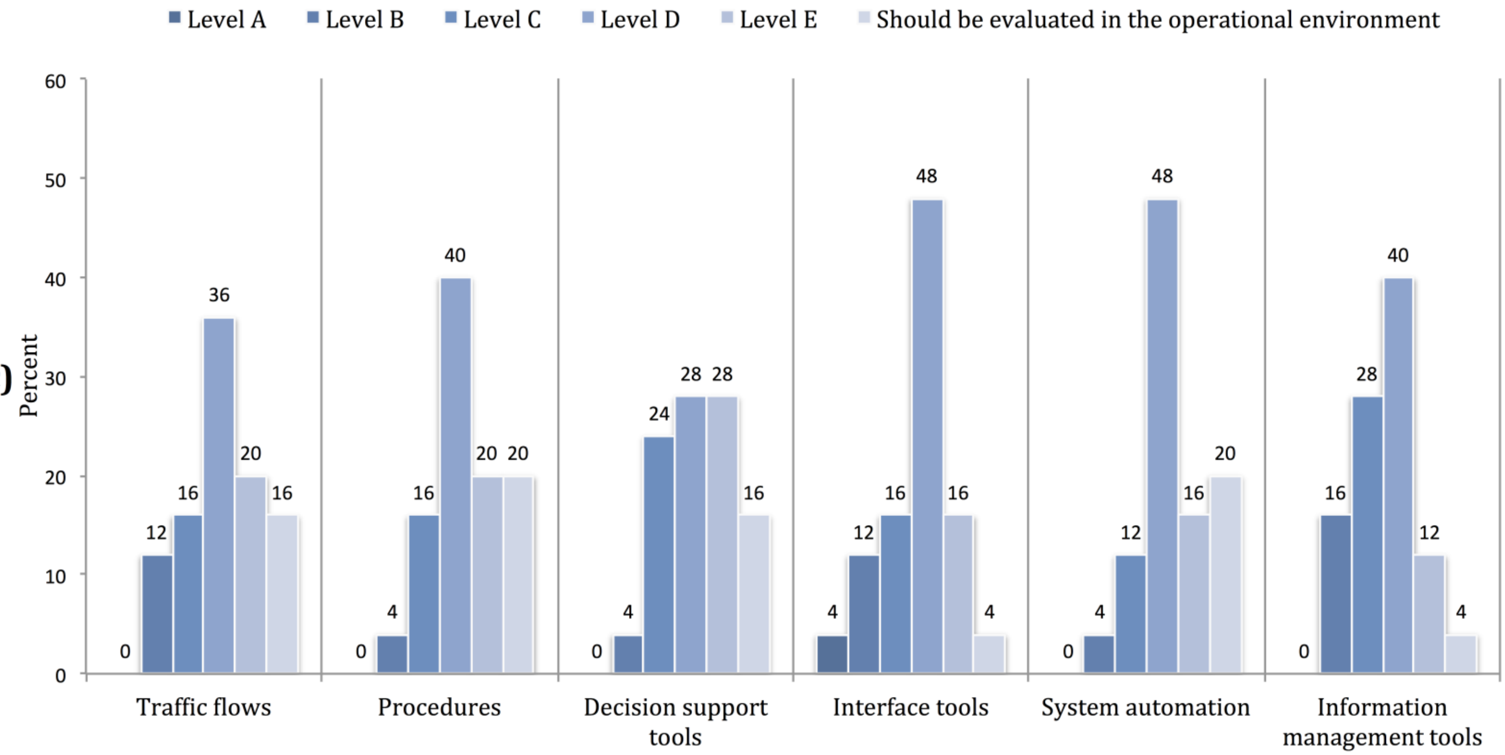

Figure F.3 Survey participant beliefs regarding selection of simulation environment for testing new operational concepts.

c) Based on the categorization system, what is the minimum level of simulation fidelity required to evaluate new operational concept X for the operational environment. (N=25)

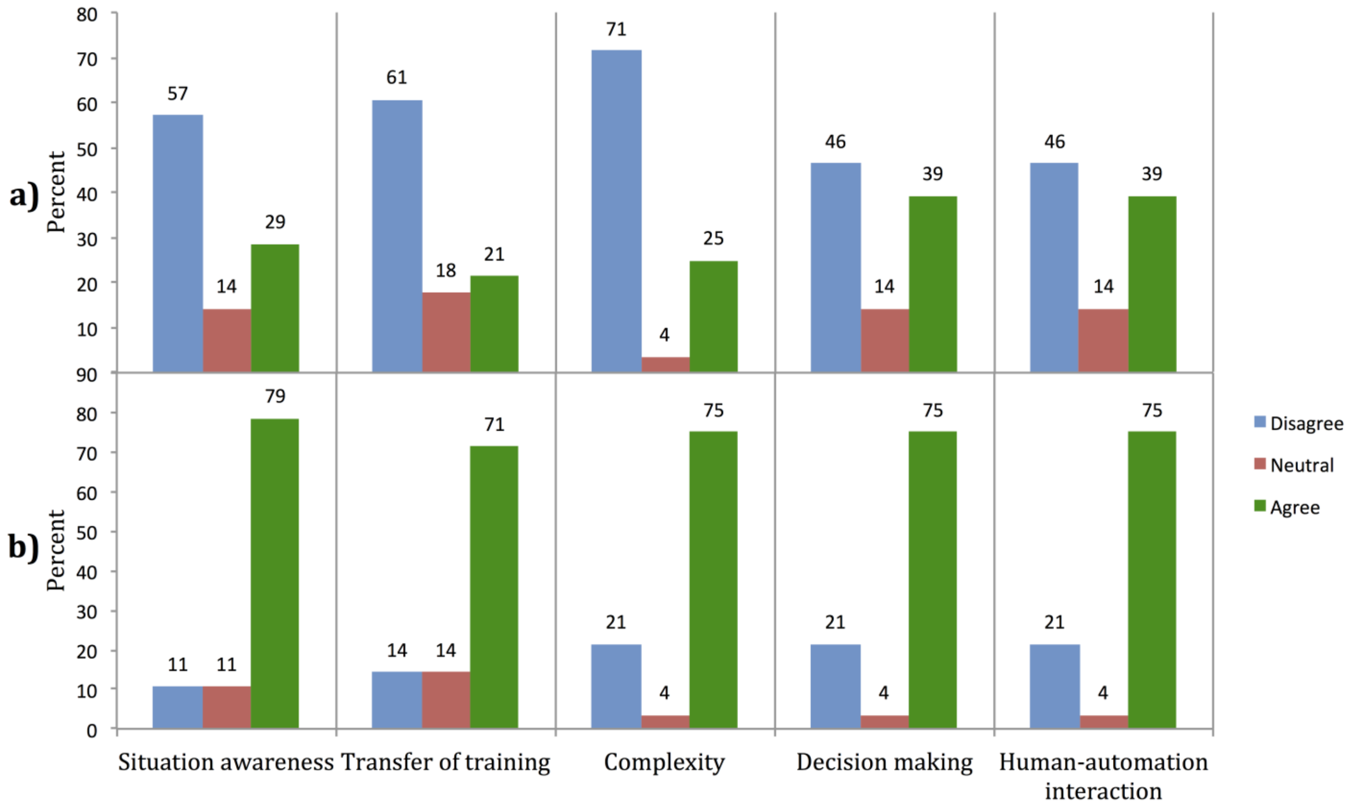

Figure F.4 Survey participant beliefs regarding selection of simulation environment for researching the impact of new operational concepts on human factors issues.

a) Classroom-based simulation is capable of evaluating the effect of new operational concepts on human factors issue X. (N=28)

b) Workstation simulation is required to evaluate the effect of new operational concepts on human factors issue X. (N=28)

c)

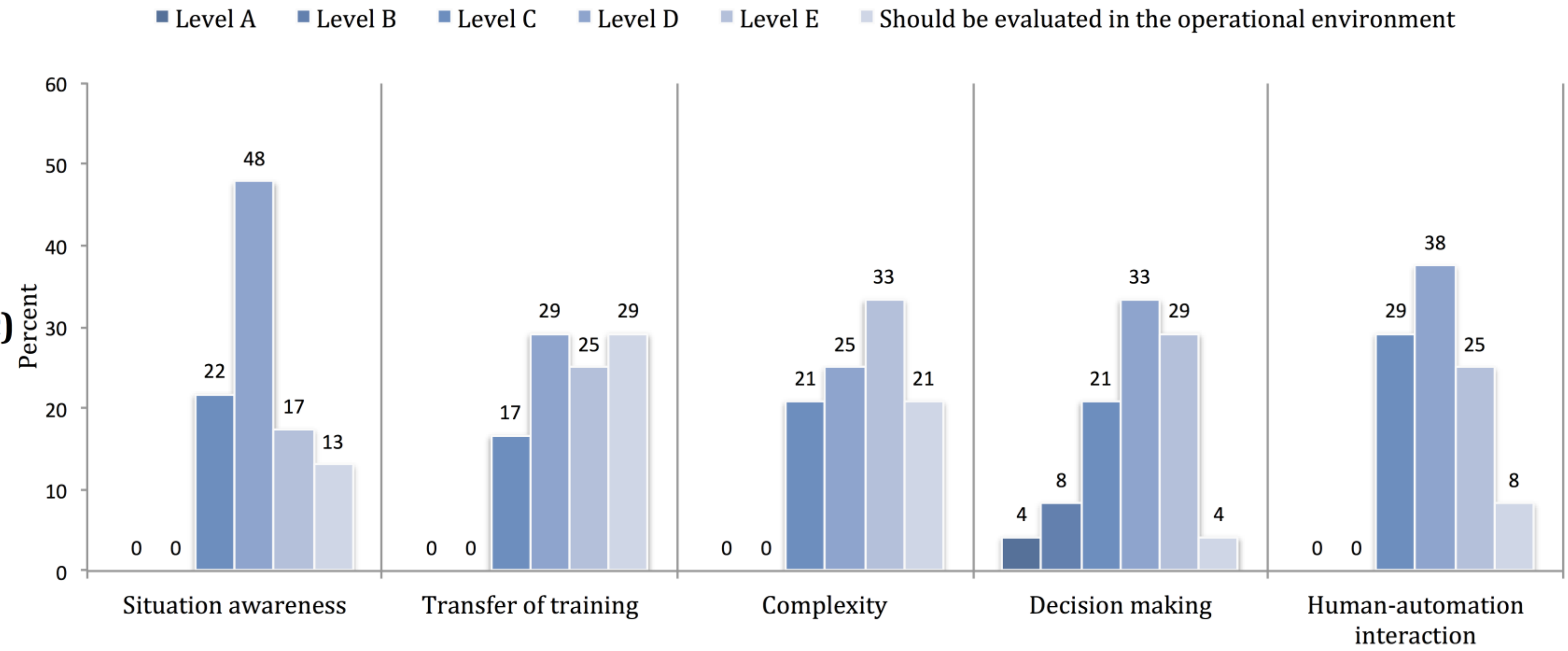

Figure F.4 Survey participant beliefs regarding selection of simulation environment for researching the impact of new operational concepts on human factors issues.

c) Based on the categorization system, what is the minimum level of simulation fidelity required to research the impact of a new operational concept on X. (N=23-24)

Appendix G

Environmental Component Definitions from Chapters 4 and 5.*

Environmental Component	Definition
Aircraft performance	This component captures how closely the aircraft performance characteristics in the simulation mirror those of the real life aircraft (this includes the variability of how different airlines and pilots may fly the aircraft).
Airspace	This component reflects the accuracy of the airspace structure and characteristics created in the simulation scenario to those in real life.
Communications	This component represents communication between the controller and all the operators within the ATC system (e.g. pilots, other controllers, flight information specialists, etc.) as well as the dynamics of these communications (includes things such as delay in responses, garbled transmissions, communicating with multiple actors simultaneously, tasks of the actors being communicated with, etc.).
Environment	Any of the surrounding area outside of what is in the immediate vicinity of a controller's workstation (i.e. the large room, other controllers' workstations, ambient noise, etc.).
Equipment	Any of the control interfaces (keyboard, mouse, etc), primary visual display and secondary information displays or interface tools that are part of a controller's workstation.
Operating system functionality	This component captures how closely the tools and capabilities of those tools of the operational environment operating system are reproduced in the simulation environment.
Scenario	This component represents the traffic situation created by the training, testing or research designers.
System participants	This component refers to the simulation's replication of the other operators within the system aside from the primary operator of the enroute ATC simulation, such as pilots and controllers in other sectors. This component captures how consistent the actions these system participants take within the simulation reflect those of real world operators.
Unpredictability	This component represents whether or not the simulation is capable of capturing off-nominal events that could happen in the real world given the operational environment's inherent unpredictability (e.g. an aircraft going left when told to go right).
Weather	This component captures how accurate weather or turbulence phenomena are reproduced in the simulation environment.

**Note*: Many of these components or alternate versions of these components appear in the simulation fidelity definition in Figure 6.1 along with descriptions in Section 6.2.