University of Central Florida

STARS

Electronic Theses and Dissertations, 2020-

2020

Exploring Delphi Method Generated Synthetic Natural Environment (SNE) Visual Aesthetic Quality (VAQ) Factor Forecasts and Preferences through Conjoint Analysis of End User Assessments

Thomas Kehr
University of Central Florida

Part of the Engineering Commons

Find similar works at: https://stars.library.ucf.edu/etd2020

University of Central Florida Libraries http://library.ucf.edu

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2020- by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Kehr, Thomas, "Exploring Delphi Method Generated Synthetic Natural Environment (SNE) Visual Aesthetic Quality (VAQ) Factor Forecasts and Preferences through Conjoint Analysis of End User Assessments" (2020). *Electronic Theses and Dissertations, 2020-*. 370.
https://stars.library.ucf.edu/etd2020/370

Showcase of Text, Archives, Research & Scholarship

EXPLORING DELPHI METHOD GENERATED SYNTHETIC NATURAL ENVIRONMENT (SNE) VISUAL AESTHETIC QUALITY (VAQ) FACTOR FORECASTS AND PREFERENCES THROUGH CONJOINT ANALYSIS OF END USER ASSESSMENTS

by

THOMAS W KEHR II
B.S. University of Central Florida, 2010
M.S. University of Central Florida, 2013
M.S. University of Central Florida, 2016

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Modeling and Simulation
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2020

Major Professor: Michael Proctor

© 2020 Thomas W Kehr II

ABSTRACT

Traditional techniques used for verification, validation, and accreditation (VV&A) of Synthetic Natural Environments for military applications are time consuming, subjective, and often costly. Due to varying levels of common visual factors, Synthetic Natural Environments (SNE) vary widely in appearance and use case. Early identification of these factors in the SNE life cycle may improve its Visual Aesthetic Quality (VAQ) while reducing VV&A issues downstream and informing future development.

This research explores supplementing existing VV&A techniques with the Delphi Method during the conceptualization phase of an interoperable SNE development in order to identify the level of importance of SNE VAQ factors for distributed, dissimilar simulations earlier in the life cycle. Delphi Method findings on VAQ factors drove the development of four different SNEs for a selected urban city center. The importance of VAQ factors within the SNEs were derived through Conjoint Analysis of data from a survey in which end user participants evaluated each SNE using a design that incorporated fractional factorial screening and Graeco-Latin Squares.

Research findings suggest: (1) using an online Delphi Method enables early identification of a correlated set of expertly accepted primary VAQ factors that affect overall realism and training utility in the virtual domain; (2) Conjoint analysis improves the understanding of the significance and power of identified factors and preferences; (3) VAQ importance rankings differed across the Delphi Method and Conjoint Analysis, nor did the Delphi Method successfully predict the two-factor interactions discovered through Conjoint Analysis of the screening design; and (4) Data mining of historical SNE issue reports did not identify the same level of importance of

VAQ factors as users reviewing SNE representations through a Conjoint Analysis and Delphi panel expert forecasts. Limitations with the proposed technique, as well as recommendations for additional research are provided to further refine the parameters associated with these subjective factors to increase the efficiency and application of the proposed approach.

This dissertation is dedicated to my father, Tom Kehr, for his unwavering commitment to my research and for teaching me the value of hard work, and to my Grandfather, Thomas W Kehr I, for laying the foundation of three generations of commitment to higher learning, education, and excellence within our family. I will carry these principles forward in hopes of inspiring the value of continuous learning among future generations of students, scholars, and peers.

ACKNOWLEDGMENTS

I am extremely fortunate to have so many colleagues and contacts that have been personally invested into the success of my research over the years. First, I would like to thank Dr. Michael Proctor, my adviser and dissertation committee chair, for his patience, wisdom, and vision throughout this research. I would also like to thank the other distinguished members of my committee, Dr. Brian Goldiez, Dr. Mark Johnson, Dr. Liqiang Ni, and Dr. Sumanta Pattanaik for their dedication and valuable feedback throughout this dissertation. I would like to thank my employers and colleagues at the U.S. Army PEO STRI and Cole Engineering Services Inc. for their sponsorship and support of this research. I would also like to thank the U.S. Army Synthetic Environment Core (SE Core) Government and Contractor teams for their support in providing historical information, data, and resources that provided background and context for this research.

This research would also not be possible without the time and support from all the volunteers who participated in the many surveys and panel discussions throughout this research. I would also like to personally thank Dr. Robert Cox, Mr. Ronald Moore, Mr. Farid Mamaghani, COL (Ret.) James Shifflett, and Mr. Sean Sedlak for the sage guidance and expertise they provided in the field of geospatial information systems and synthetic terrain generation.

Finally, this dissertation would not have been possible without the continuous support from my incredible friends and family. Thank you all for helping to keep me focused on this achievement and for your constant encouragement throughout this process.

TABLE OF CONTENTS

LIST OF FIGURES xii
LIST OF TABLES xvii
LIST OF ACRONYMS xxi
CHAPTER ONE: INTRODUCTION 1
Problem Overview 4
Research Question 12
Problem Domain and Background 14
Military M&S Domains 14
M&S Synthetic Natural Environments 17
Terrain Database Generation 24
Immersion, Interactivity, and Realism 29
Qualitative to Quantitative Visual Aesthetic Quality Analysis 31
Verification, Validation, and Accreditation 32
VV&A User Roles and Responsibilities 41
Future Challenges and Issues 45
Overview 46
CHAPTER TWO: STATISTICAL TECHNIQUES FOR FORECASTING AND OBTAINING USER FEEDBACK 47
Hiring the Best and Brightest 47
Iterative Design 48
The Delphi Method 51
The Delphi Process 52
Selection of Delphi Expert Panel 54
Delphi Panel Feedback 55
Delphi: Round One 56
Delphi: Round Two 56
Delphi: Round Three 57
Delphi: Round Four 57
Delphi Data Analysis 57
Challenges and Drawbacks of the Delphi Method 59

Applicable Delphi Studies to the Research 61
Conjoint Analysis 62
Traditional Full-Profile Conjoint Analysis 65
Adaptive Conjoint Analysis 66
Choice Based Conjoint 69
Partial-Profile Choice Based Conjoint 70
Adaptive Choice Based Conjoint 71
Menu-Based Choice 73
Choosing a Conjoint Analysis Method 73
Data Analysis in Conjoint Studies 74
Iterative Design in Favor of Conjoint Analysis 74
Applicable Conjoint Analysis Studies to this Research 75
Fractional factorial experimental design and Conjoint Studies 77
Graeco-Latin Square Designs and Conjoint Analysis 78
Quality Function Deployment 79
Product Planning 81
Assembly/Part Deployment 85
Process Planning 86
Process/Quality Control 87
Limitations of QFD 88
Example QFD Deployments 89
Additional Relevant QFD Literature 90
CHAPTER THREE: PROPOSED METHEDOLOGY 92
Hypotheses 93
SNE Quality Evaluation Use-Case 95
Research Scope 99
Phase One: Conduct a Delphi Study of SNE Quality Factors 99
Delphi Survey Design 101
Delphi Data Analysis 102
Delphi Survey Development 103
Delphi Panel Feedback 107
Selection of Expert Panel 108

Piloting the Delphi 110
Distributing the Delphi Surveys 111
Ethical Considerations 111
Concluding the Delphi 113
Phase Two: Experimental Design and Factor Analysis 113
Phase Three: SNE Generation 114
SNE Generation Software Applications 115
SNE Visual Aesthetic Quality Factor Manipulation 116
Selection of SNE Area of Interest 117
Runtime Simulation Generation - Unreal Engine 4 117
Phase Four: End User-based Conjoint Analysis 119
Conjoint Design 119
Conjoint Participant Selection 123
Piloting the Conjoint Study 125
Conducting the Conjoint Study 125
Phase Five: Data Analysis and Recommendations 126
Conjoint Statistical Analysis 126
Domain-specific Use-Cases and Additional Data Analysis 126
Delphi Study versus Traditional SNE Core VV&A 127
Conjoint Analysis versus Traditional SE Core VV&A 128
Methodology Summary 129
CHAPTER FOUR: EXPERIMENTAL EXECUTION AND DATA ANALYSIS 130
Phase One Results: Delphi Study of SNE Quality Factors 130
Delphi Method Round One: Open-Ended Survey 131
Delphi Method Rounds Two, Three, and Four 134
Delphi Method: Data Analysis 137
Phase Two Results: Selected Experimental Design and Factors 149
Phase Three Results: SNE Generation 152
Factor 1: Elevation 153
Factor 2: Spatial Cues 156
Factor 3: Orientation 157
Factor 4: Weather Effects 159

Factor 5: Vegetation Density ... 160
Factor 6: Building Complexity ... 161
Factor 7: Geospecific Locations ... 163
Factor 8: Time of Day ... 165
Factor 9: Level of Detail (LOD) ... 165
Phase Four Results: End User-based Conjoint Analysis ... 166
Conjoint Analysis Design ... 166
Conjoint Analysis Survey Execution ... 171
Conjoint Analysis Data Analysis ... 172
Comparing Synthetic Environment User Groups ... 177
Comparison of Findings Against Historical Data ... 180
CHAPTER 5: DISCUSSIONS, LIMITATIONS, FUTURE RESEARCH, AND CONCLUSIONS ... 183
Hypotheses 1: For a given synthetic natural environment representation across dissimilar simulators within a multi-domain simulation exercise, there exists a correlated set of expertly accepted and user validated primary VAQ factors that affect overall realism and training utility in the virtual domain. ... 184
Data Analysis Discussion ... 184
Delphi Method Discussion ... 185
Hypotheses 2: Conjoint analysis will improve the understanding of the significance and power of identified factors and preferences ... 186
Conjoint Analysis Discussion ... 187
Comparison of Research Techniques ... 190
Addressing Synthetic Natural Environment (SNE) Visual Aesthetic Quality (VAQ) Factors for Virtual Training ... 193
Discussion of Additional Research Hypotheses ... 195
Hypotheses 3A: A Delphi study using a panel of experts will forecast the same VAQ factor considerations as Conjoint Analysis of end user assessments. ... 195
Hypotheses 3B & 3C: Data mining of historical SNE issue reports will identify the same level of importance of VAQ factors as users reviewing SNE representations through a Conjoint Analysis AND Delphi panel expert forecasts. ... 196
Hypotheses 4: Quality Function Deployment (QFD) can be utilized to abstract the correlated set of expertly accepted and user validated primary SNE VAQ factors into a series of SNE generation process improvements to influence a new SNE VV&A paradigm. ... 196
Limitations ... 200

Lesson Learned 201
Future Research 202
Conclusions 203
APPENDIX A: ELECTRONIC DELPHI SURVEY 205
APPENDIX B: SAMPLE INDIVIDUAL DELPHI FEEDBACK REPORT 215
APPENDIX C: SNE VISUAL AESTHETIC QUALITY CONJOINT ANALYSIS SURVEY 218
APPENDIC D: UCF PARTICIPANT CONSENT FORM 242
APPENDIX E: UCF IRB APPROVAL LETTER 246
REFERENCES 253

LIST OF FIGURES

Figure 1: Global map showing the projected number of megacities as the global population increases through 2025 ... 2

Figure 2. The scope of synthetic natural environment representations ... 18

Figure 3. Generic terrain database (TDB) generation pipeline ... 26

Figure 4. SE Core Standard Terrain Database Generation Capability (STDGC) Diagram ... 27

Figure 5. DoD System Engineering Process Model ... 32

Figure 6. VV&A events highlighted in the SE Core Standard Terrain Database Generation Capability (STDGC) Diagram ... 38

Figure 7. Notional Delphi Process. ... 53

Figure 8. The ACBC Interview Process ... 72

Figure 9. Four-Phased QFD Approach ... 81

Figure 10. Example House of Quality ... 82

Figure 11. Example Concept Selection Matrix ... 85

Figure 12. Example Part Deployment Matrix ... 86

Figure 13. Example Process Planning Matrix ... 87

Figure 14. Example Quality Control Planning Matrix ... 88

Figure 15. High-Level Research Design ... 95

Figure 16. Delphi overview research methodology ... 100

Figure 17. Example Round One Delphi Feedback Summary Chart for SNE Developer Group 108

Figure 18. Location and Size of SNE Core Emerald City dataset. ... 115

Figure 19. Ranking Survey with Radio Buttons ... 121

Figure 20. Ranking Survey with Drop-Down Selection 122

Figure 21. Proposed Conjoint Analysis Survey Design for SNE Visual Quality 123

Figure 22. Word cloud generated through the open-ended survey responses from Round One of the Delphi. Numbers in parentheses indicate the frequency of occurrence of each word. Non-SNE related words were excluded from analysis and illustration. 132

Figure 23. Round Two Pairwise Graph indicates the most important Statement 1 is statistically more important than Statements 5, 7, 12, 15, 17, 21, 22 142

Figure 24. Round Three Pairwise Graph indicates convergence of relevant importance with Statements 5, 7, 15, 21 of the previous set now having six or more statistically significant differences with other Statements 143

Figure 25. Round Four Pairwise Graph indicates the dominance of Statement 1 but also has inconsistency with relevant importance of other Statements identified in Round Three 144

Figure 26. Results of the mirroring versus mimicking questionnaire from the Delphi exit survey. 148

Figure 27. Results of the groupthink questionnaire from the Delphi exit survey. 149

Figure 28. Visualization of low-resolution (30-meter) elevation used for Factor 1 of the SNE design. 155

Figure 29. Visualization of high-resolution (3-meter) elevation used for Factor 1 of the SNE design. Notice the increased terrain relief and shadows near the center of the image. 155

Figure 30. Level 1 of Factor 2 (Spatial Cues) depicted through an ~10-meter tall tree model. Image rendered in UE4. 156

Figure 31. Level 2 of Factor 2 (Spatial Cues) depicted through an ~20-meter tall tree model. Image rendered in UE4. 157

Figure 32. Vector data comparison of Factor 3 (Building Orientation) levels. Blue shapes represent the default building orientation (Level 1) and Red shapes represent the incorrectly oriented buildings (Level 2). Data depicted within QGIS software. 158

Figure 33. Correct Building Orientation (Level 1) depicted in Conform™ 3D visualization software........ 158

Figure 34. Incorrect Building Orientation (Level 2) depicted in Conform™ 3D visualization software........ 159

Figure 35. Comparison of Factor 4 (Weather Effects) and Factor 8 (Time of Day). Frame 'A' displays a daytime scene with no weather effects. Frame 'B' displays a daytime scene with weather effects enabled (note water droplets on the screen). Frame C displays a dusk scene with no weather effects. Frame D displays a dusk scene with weather effects enabled. 160

Figure 36. Comparison of Vegetation Density levels as depicted through geospatial vector features. The Blue dots represent the normal density of trees (Level 1) while the Red dots represent the remaining trees after a 75% reduction in geospatial point features (Level 2)....... 161

Figure 37. Virtual cityscape of Seattle, WA depicting "extruded" models (Level 1). Note the simple repeating textures and simple building shapes. The visual scene is rendered through the Conform™ geospatial visualization tool. 162

Figure 38. Virtual cityscape of Seattle, WA depicting complex "geotypical" models (Level 2). Note the enhanced geometry to include windows, building overhangs, roof ledges, and other

building features. The visual scene is rendered through the Conform™ geospatial visualization tool. 163
Figure 39. Accurate placement and depiction of the Space Needle (Level 1). Visual scene rendered within Unreal Engine 4. 164
Figure 40. Inaccurate placement and depiction of the Space Needle (Level 2). Note fictitious lake underneath the model. Visual scene rendered within Unreal Engine 4. 164
Figure 41. Depiction of visual scene rendered entirely at the highest LOD. Scene rendered within UE4. 166
Figure 42. Depiction of multiple LODs within the visual scene. The lower LOD can be observed in the top section of the figure when compared to the above figure. Scene rendered in UE4. ... 166
Figure 43. Layout of a single "trial" for the Conjoint Analysis survey. Option A represents SNE Design 1, Option B is SNE Design 2, Option C is SNE Design 3, and Option D is SNE Design 4. The video depicts the SNE from the ground perspective. 170
Figure 44. This image depicts the same SNE as seen in Figure 1 but further in the video, presenting an aerial view of the SNE. 171
Figure 45. Word cloud of key words used throughout the comments of the combined SNE conjoint analysis survey responses. The numbers to the right of each word indicate frequency that each word was used. 176
Figure 46. Word cloud analysis of 931 SE Core discrepancy reports across three user assessments. The number in parentheses indicates the frequency of the word used across all discrepancy reports. 182

Figure 47. Depiction of the "floating" building visual anomaly with complex 3D building models. Images depict a high-resolution 3-meter Digital Elevation Model (DEM).................. 193

Figure 48. Example QFD House of Quality implementation for the US Army SNE Core program. The section highlighted in 'A' and 'B' show example incorrect assumptions that may have been revised based on findings highlighted through this research..................................... 198

Figure 49. Reinforcing the central superstructure of the QFD House of Quality through the implemented Research Methodology 199

LIST OF TABLES

Table 1. Live, Virtual, and Constructive Simulation Domains 15

Table 2. Four sub-classes of interoperability problems to facilitate the development a quantitative definition. Adapted from Purdy & Goldiez (1995) 20

Table 3. Summary of Verification, Validation, and Accreditation Errors (Petty, 2010) 35

Table 4. Summary of V&V Methods (Balci, 1998) 37

Table 5. Comparison of V&V User Roles and Tasks 44

Table 6. Strengths and Weaknesses of The Delphi Method 60

Table 7. Experimental design of conjoint analysis for evaluation of a carpet cleaner 63

Table 8. ACA Step 1: Rank attributes in terms of preference 67

Table 9. ACA Step 2: Rate importance of attributes 68

Table 10. ACA Step 3: Pairs using graded rating scale 68

Table 11. Example choice set for CBC analysis 70

Table 12. Summary of when to use each method of Conjoint Analysis 73

Table 13. Summary of the U.S. Army ITE Virtual Training Aids, Devices, Simulations, and Simulators 97

Table 14. Level of Familiarity Likert type scale rating for Delphi panel domain demographic analysis 104

Table 15. SNE Domain Use-Case relation to U.S. Army Virtual Simulation Platforms 105

Table 16. Likert-type scale response: Level of Agreement 107

Table 17. SNE Generation PC Hardware Specifications 116

Table 18. SNE Core Issue Severity Definitions 128

Table 19. Research question and training scenario provided in round one survey.................... 131

Table 20. Synthesized final statements from the Open-Ended Survey. A simplified descriptor for each factor is also provided for efficient analysis purposes later. 132

Table 21. Likert-type scale response: Level of Importance... 135

Table 22. Response rate between each Delphi study round.. 136

Table 23. Rank Order from top to bottom based on mean for the indicated Delphi round. Shaded cells indicate the Statement did not meet the 80% consensus level. 139

Table 24. Results of Wilcoxon Signed-Rank Test visa via ambivalence for each structured Delphi survey round. Shaded cells indicate statements with no statistical difference from a neutral rating, or ambivalence.. 141

Table 25. Final Ranking of Visual Quality Statements based on Kruskal Wallis mean rankings and Consensus from Round Three of the Delphi... 145

Table 26. Comparison of Median Ranking and Response Rate for Statement 16 (single geospatial source provider) across each Delphi survey Round.. 147

Table 27. Synthetic Environment Visual Quality Factors from the Delphi Method to the Conjoint Analysis Screening Design 151

Table 28. Resolution III Fractional Factorial Design used for Synthetic Environment Conjoint Analysis... 151

Table 29. Software Applications and Plugins used to generate the 16 SNEs for the Conjoint Analysis phase of the research.. 153

Table 30. Comparison of digital elevation datasets used for SNE Factor 1 of the screening design 154

Table 31. Conjoint Analysis Survey Design using the combination of a fractional factorial screening design and Graeco-Latin Squares. SNE Designs 1, 2, 3 and 4 are highlighted to show that each SNE combination is only ever compared once. 167

Table 32. Summary of Fit, Analysis of Variance, and Parameter Estimates of the regression model considering only main effects for Combined Expert Responses to the SNE Conjoint Analysis. Bold numbers indicate statistically significant findings at the 0.2 174

Table 33. Analysis of Main Effects and 2-Factor Interactions through the JMP Screening platform using Lenth's Pseudo-Standard Error (PSE). Bold numbers indicate statistically significant findings at the 0.2 level. Orange numbers indicate a p-value<0.01 175

Table 34. Summary of Fit, Analysis of Variance, and Parameter Estimates of the regression model considering main effects AND 2-factor interactions for Combined Expert Responses to the SNE Conjoint Analysis. Bold numbers indicate statistically significant finding 177

Table 35. Comparison of Synthetic Environment visual quality factor importance rankings between military and civilian users queried through the conjoint analysis survey........ 178

Table 36. Comparison of Synthetic Environment visual quality factor importance rankings between users of various simulation-domain backgrounds queried through the conjoint analysis survey. 179

Table 37. Comparison of historical SNE VV&A VAQ Findings from the US Army SE Core program with the implemented research methodology. Shaded cells indicate matching ranks across methodologies. 181

Table 38. Comparison of Synthetic Environment Visual Quality Factor Rankings Across Implemented Methodologies. 190

Table 39. Re-Structured Synthetic Environment Visual Quality Factors from the Delphi Method to the Conjoint Analysis Screening Design to Reflect Low-Fidelity and High-Fidelity Factor levels. Factor Levels in blue have been updated from the original research screening design. . 201

LIST OF ACRONYMS

AOI	Area of Interest
AVCATT	Aviation Combined Arms Tactical trainer
CCTT	Close Combat Tactical trainer
CLI	Command Line Interface
DARPA	Defense Advanced Research Projects Agency
DHS	Department of Homeland Security
DIGEST	Digital Geographic Information Exchange Standard
DIS	Distributed Interactive Simulation
DIUx	Defense Innovation Unit Experimental
DOD	Department of Defense
DoE	Design of Experiments
DR	Discrepancy Report
DSTS	Dismounted Soldier Training System
EDCS	Environment Data Coding Standard System
FACC	Feature and Attribute Coding Catalog
GFT	Games for Training
GIS	Geospatial Information Systems
GUI	Graphical User Interface
HCI	Human-Computer Interaction
IEEE	Institute of Electrical and Electronics Engineers
IG	Image Generator
JCIDS	Joint Capabilities Integration and Development System
JCIDS	Joint Capabilities Integration and Development System
LOD	Level of Detail
LVC-IA	Live, Virtual, Constructive Integrating Architecture
M&S	Modeling and Simulation
M&SCO	Modeling & Simulation Coordination Officer
MDB	Master Database
MEGA	Major Evaluation of Geospatial Areas
MUSE	Multiple Unified Simulation Environment
NGA	National Geospatial-Intelligence Agency
PDF	Portable Document Format
PEO STRI	Program Executive Officer for Simulation, Training, and Instrumentation
PM	Program Manager
QFD	Quality Function Deployment
RVTT	Reconfigurable Vehicle Tactical Trainer

SE Core	Synthetic Environment Core
SEDRIS	Synthetic Environment Data Representation and Interchange Specification
SNE	Synthetic Natural Environment
STDGC	Standard Terrain Database Generation Capability
TAM	Technology Acceptance Model
TDB	Terrain Database
TIN	triangular information networks
TTPs	tactics, techniques and procedures
V&V	Verification and Validation
VAQ	Visual Aesthetic Quality
VBS	Virtual Battle-Space
VV&A	Verification, Validation, and Accreditation

CHAPTER ONE: INTRODUCTION

The battlefield of tomorrow will not take place on the traditional regionally specific environment that has long been the standard for military training doctrine. Increasing global migration and increasing urbanization will lead to the development of densely populated metropolises, or megacities, where armed groups may seek to exploit popular disaffection and weak governance (U.S. Army Training and Doctrine Command, 2010). Figure 1 provides an expert estimation on the rise of megacities as global population increases. Additionally, advances in aerial transport and munitions means that we can now wage combat globally across multiple regions simultaneously.

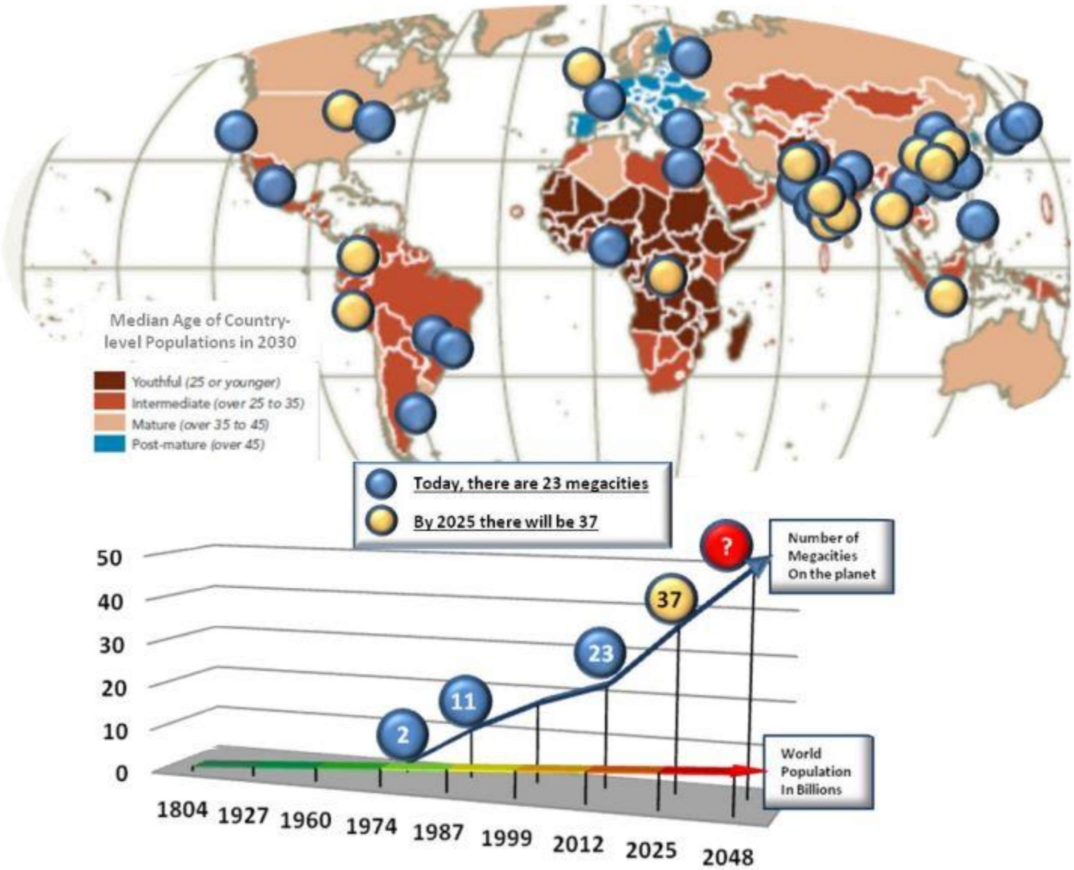


Figure 1: Global map showing the projected number of megacities as the global population increases through 2025

Source: (Chief of Staff of the Army Strategic Studies Group, 2014)

In response to this emerging threat, the military Modeling and Simulation (M&S) domain faces the challenge of rapidly developing larger, more diverse, and more complex synthetic representations of natural environments necessary to facilitate military training. Several organizations anticipate the need for enhanced capability and to harness recent advances in computing power and gaming technology to create early prototypes of these complex globally dispersed synthetic natural environments (SNE's) (Bohemia Interactive Simulations, 2015; Michael Peck, 2015).

The creation of SNE for virtual simulations has historically been tailored to specific simulator hardware performance baselines and training requirements(Lalonde, 2008; Shufelt, 2006). For virtual simulations involving kinetic or asymmetric warfare (Baca & Proctor, 2017), a SNE contains a virtual representation of a real-world geographic battlespace (Smelik, van Wermeskerken, Krijnen, & Kuijper, 2019). Many factors such as geospatial source data resolution and simulator image generator (IG) rendering will influence 'fitness for use' or the Visual Aesthetic Quality (VAQ) of the SNE representation (B. Graniela & Proctor, 2012; Kang, Kim, & Han, 2015; Purdy & Goldiez, 1995; U.S. Army PEO STRI, 2017a). Divergent specialized virtual system requirements within a large set of distributed simulations may lead to increased costs and schedule just to create and maintain a correlated and interoperable synthetic environment (Durall, 2018). As an example, the Synthetic Environment Core (SE Core), the US Army's leading SNE generation program, currently provides SEs for 15 different simulation systems and supports more than 57 unique SNE terrain formats (U.S. Army PEO STRI, 2017b). Easily perceived as costly and redundant (Durall, 2018; "STE, OTW, OTA, and TReX," 2019), the SNE representations, each tailored and optimized for specific warfighting requirements, maximize simulator performance and positive training transfer for the intended warfighters.

The United States Army's synthetic terrain generation approach for simulation and training is currently undergoing a paradigm-shift. Instead of creating a series of specialized terrains for specific warfare training tasks, the One World Terrain (OWT) paradigm diverges from a tailored SNE approach and hopes to provide the US Army with a common 3D whole-Earth CSE suitable for collective land, air, maritime, and space operations and training (U.S. Army STE Cross-Functional Team (CFT), n.d.). The Army's OWT is also being designed to prepare soldiers for future operations in dense urban environments and "megacities"(Alderton, 2019).

In order to meet the demanding SNE requirements of each of these disparate training audiences and their tactics, new and improved verification, validation, and accreditation (VV&A) processes must be developed. Successful VV&A begins with identifying SNE VAQ factors to ensure visual quality across such a large scale. Additionally, the balance between VAQ factors and disparate training audience requirements needs to be identified to enable SNE developers and program managers to balance trade-off analysis so as to adhere to cost and schedule constraints (Stevens, Kincaid, & Sottilare, 2015). V&V are necessary processes before potential accreditation of a given SNE can be deemed suitable for interoperable use cases (Kenyon, 2016). Verification, validation, and accreditation (VV&A) are necessary processes in the system life cycle of a SNE.

Problem Overview

Approved by an authority within an agency designated by an M&S sponsor, SNE accreditation is "the official certification that a model or simulation is acceptable for use for a specific purpose" (DoD Modeling & Simulation Coordination Office, 2013). Accreditation is a necessary standard for SNE quality. Properly accredited SNE promote user capabilities. Non-accredited representations cannot be assumed to be of acceptable quality as they may promote negative user training (Petty, 2010).

M&S system developers devote a portion of available resources to SNE verification and validation in the hope of achieving accreditation. Verification is “the process of determining that a model or simulation implementation accurately represents the developer’s conceptual description and specification. Verification also evaluates the extent to which the model or simulation has been developed using sound and established software engineering techniques” (DoD Modeling & Simulation Coordination Office, 2013). Validation is “the process of determining the degree to which a model or simulation is an accurate representation of the real-world from the perspective of the intended uses of the model or simulation” (DoD Modeling & Simulation Coordination Office, 2013).

As SNE increase in scale, diversity, and complexity to keep up with the similarly increasing demands of the user community, additional fiscal and manpower resources as well as new technologies and techniques may be required to achieve accreditation of a SNE. Given continuation of past budget constraints, current V&V tools and processes will likely incur increasing accreditation risk in light of increasing SNE scale, diversity, and complexity considerations. In order to mitigate accreditation risk, simulation developers must better understand the factors and their associated parameters that experts and users can agree are the most important to SNE quality. Once managers and developers are better able to understand these factors, SNE developers may make better-informed decisions upstream in the SNE development process. Further, V&V agents may provide better inputs to their tools so as to identify and resolve critical issues prior to an accreditation assessment.

The Synthetic Environment (SE) Core program of record within the United States Army’s Program Executive Office for Simulation, Training, and Instrumentation (PEO STRI) may serve as a case study in SNE development. On average, a suite of SE Core interoperable synthetic terrain products incurs direct costs upwards of $2-million. Virtual terrain products like the Close Combat Tactical Trainer (CCTT) and the Aviation Combined Arms Tactical Trainer (AVCATT) account for the majority of these cost due to their complexity and advanced configurations that are required due to strict hardware limitations (U.S. Army, 2016). Approximately 20% of the budget for a given SE Core terrain product is devoted to verification and validation (V&V) efforts (U.S. Army PEO STRI, 2016b). SE Core accomplishes most of this V&V through iterative builds of SNE products. A final validation event occurs at the end of the terrain generation process. Based on the outcome of this final event, accreditation agents will either approve the SNE for training use or require additional development and V&V to fix identified issues (U.S. Army PEO STRI, 2015). Users base accreditation on the number and severity of the issues recorded during the final validation event. If an accreditation failure or partial failure occurs, program management personnel must expend additional cost and schedule to improve the SNE, conduct additional validation processes, and attempt another user accreditation (Kehr, Godwin, & Mcintire, 2014). At this time, no SE Core virtual terrain product (designed for interoperable, human-in-the-loop, real-time processing) has fully passed this final validation event without requiring some level of re-work to achieve accreditation for user fielding. Direct re-work costs can run anywhere from $10K upwards of $50K based on the level of effort and time required to turn-around the product (U.S. Army PEO STRI, 2016b). Additionally, this delay may also result in follow-on terrain products being delayed, thus delaying training capability to the warfighter. Fielding delay may adversely affect soldier training.

Alternatively, a worse case is if the accreditation agent inadvertently approves a flawed terrain product for interoperable training use. Once shipped to the field, undetected flawed terrain potentially leads to negative training imparted to users. Negative training may cost user lives. When the user discovers SNE issues not discovered during VV&A, the developer must fix the SNE product. Besides user costs, such late life-cycle stage fixes carry a much higher financial cost due to having to revert to an older SNE production baseline after moving onto additional projects. Depending on length of faulty SNE fielding, other correlated terrain products may also need revision to maintain a “fair fight” across interoperating Live, Virtual, and Construction (LVC) (defined further below) systems. Recall of fielded, but flawed terrain products, occurred twice in the history of SE Core resulting in expenditures upwards of several million dollars to fix.

Even with automated test tools and a dedicated V&V teams, many SNE issues still get through in final SNE products. Many of the issues identified in final SNE products exist due to the subjective nature of VV&A methods, especially from a Subject Matter Expert (SME) user perspective. Factors that may not seem like an issue to one party may be a significant issue to another (Kehr et al., 2014). This is especially true for location dependent visual aesthetic quality (VAQ) factors, particularly if the developer is not familiar with the terrain location and the SME ascribed aesthetic visual details characteristic of the geographic area.

Defining visual aesthetic quality may be difficult. The Human-Computer Interaction (HCI) community defines aesthetic as “an artistically beautiful or pleasing appearance” or “a pleasing appearance or effect: Beauty” and the term “visual” indicates concentration on the visual sense (Tractinsky, 2013). In terms of “effect”, the aesthetic characteristics of a visual scene in a SNE may provide users important cues that may trigger behavior necessary for training. In contrast, a SNE that is unable to provide a visual scene those cues deemed by the user sufficient to replicate real world use cases is deficient in terms of training value. SNE deemed deficient by a user accreditation authority represents failure. Therefore, this research defines VAQ as: The visually pleasing appearance of a product in accordance with a customer needs and wants.

Failure to accredit a SNE is more likely for a SNE built using automated test tools and traditional V&V practices without first understanding the significance of a SNE VAQ factors. While automated test tools and traditional V&V practices currently catch technical issues, diagnosing and identifying these visual aesthetic quality issues early in development and VV&A is much harder to achieve without improved V&V techniques.

One must clarify the word “quality” before proceeding. Instead of the term “quality” the military community often uses the term “fidelity” when describing a SNE. As far back as DoD 5000.59-P, "Modeling and Simulation Master Plan," October 1995, fidelity is defined as “the accuracy of the representation when compared to the real world” (Under Secretary of Defense for Acquisition and Technology, 1995). The military associates fidelity of a SNE with its accuracy with the real world natural environment. "A Glossary of Modeling and Simulation Terms for Distributed Interactive Simulation (DIS)," August, 1995 defines “accuracy” as “The degree of exactness of a model or simulation, high accuracy implying low error. Accuracy equates to the quality of a result, and is distinguished from precision, which relates to the quality of the operation by which the result is obtained and can be repeated.” Thus, the military associates accuracy with the word quality.

Quality can be a subjective term for which each person, sector, or community may have its own definition (American Society for Quality, n.d.). The American Society for Quality identifies two separate definitions for quality:

1. The characteristics of a product or service that bear on its ability to satisfy stated or implied needs (American Society for Quality, n.d.) ("implied" often inferring satisfaction of a human-based characteristic)
2. A product or service free of deficiencies (the absence of "implied" often inferring free of specification-based deficiencies)

Quality expert, Philip Crosby, defines quality as a "conformance to requirements" (Creech, 1994) leaning heavily toward the specification-based notion of quality. As evidenced by user-based accreditation rejections noted above, defining SNE visual "requirements" often confounds requirements developers and system implementers and thus may be insufficient. Noted quality expert, Joseph Juran, taking a more human focus on quality rather than specification focus, describes quality as a "fitness for use" and iterates that a high-quality product does what its customers want in such a way that they actually use the product (Juran & Godfrey, 1999).

Fitness for use infers user beliefs, attitudes, intentions, and behavior. Connecting belief, attitude, intention and behavior, the Theory of Reasoned Action from Fishbein and Ajzen (1975) and attempts to understand and predict human behavior (Fishbein & Ajzen, 1975). For technology products such as a SNE, Davis (1993) developed the Technology Acceptance Model from the Theory of Reasoned Action to address acceptance of technology by individuals (Davis, 1993). The notion of "acceptance" by users is the essence of military "accreditation". The Technology

Acceptance Model (TAM) includes two elements closely tied to the human-based notion of quality. First TAM defines perceived usefulness as, "the degree to which an individual believes that using a particular system would enhance his or her job performance". Secondly, TAM defines perceived ease of use as, "the degree to which an individual believes that using a particular system would be free of physical and mental effort" (Ajzen & Fishbein, 1977; Davis, 1985 pg. 26).

Attempts to satisfy SNE expectations can also overwhelm a program manager's fiscal and computational resources as well as confound SNE developers, given technology limitations or constraints (Ferwerda, 2003; Mourkoussis et al., 2010). In contrast with trainees, trainers actually put less emphasis on photorealism while emphasizing a SNE's ability to accurately transfer knowledge to a trainee (Mourkoussis et al., 2010). Finding the balance between expectations is a challenge to producing SNE's acceptable among divergent training, management, and developer communities (Hartmann, Sutcliffe, & Angeli, 2008). With SNE's involving "life or death" use cases, from training of airline pilots to military medical applications to military training exercises, extremely divergent and sometimes strongly held positions arise on visual expectations (Department of the Army, 2017; Hackett & Proctor, 2016; Michael D. Proctor & Campbell-Wynn, 2014). To meet ongoing training requirements, acceptable compromise between divergent decision maker expectations is sought (Ferwerda, 2003).

One may describe the degree to which a SNE reflects the real-world natural environment in terms of fidelity, quality, acceptance, or some combination of the three.

Unless otherwise described, this research will focus on quality in the sense of human-based, visual aesthetic of a SNE or Visual Aesthetic Quality (VAQ).

Given the quality linkage, a military-based case study is extensible beyond the military use-case. Further, an important aspect of this research rests in the general process of identifying a user's needs and wants and incorporating them into the design and production of a product. Identifying user needs and wants and incorporating them in design and production processes are common across industrial engineering applications in the form of Quality Function Deployment (QFD). QFD seeks to listen to the "voice of the customer" in order for a developer to produce goods or services that the user actually wants while also adding value to them (Madu, 2006).

As evidenced by the fore mentioned incidents and failures as well as the increasing future SNE complexity, developers and V&V agents need new techniques to better understand SME user wants, needs, and requirements and convert them into measurable operational parameters. Life-cycle-oriented goals of this research include providing terrain developers with the information required to: (1) increase first time accreditation success; and (2) reduce the overall cost and schedule required for V&V activities downstream in the SNE generation process. The general research approach is to implement life cycle process techniques to improve quality upstream input and that through early identification of the major factors that affect an interoperable SNE's visual aesthetic quality downstream. Earlier and improved upstream quality will likely result in SNE products being a much higher quality at the conclusion of their generation process and may alleviate many of the issues identified once they reach the user.

The particular focus of the dissertation is on improving forecasting techniques and identifying levels of preferences for SNE VAQ factors through analysis. Research is case-based with the particular case being SNE V&V within SNE Core. SE Core is a suitable case as on the SE Core program alone, V&V technique improvement has the potential to provide hundreds of thousands of dollars in cost avoidance every year to the U.S. Army. Total cost avoidance may be significantly higher than direct cost avoidance as other organizations and agencies work to apply the research outlined in this dissertation to their own SNE development processes. Further SE Core SNE represent global real world locations (U.S. Army PEO STRI, 2016a). As noted earlier, the research may extend the concepts and techniques considered to industrial engineering notion of quality and QFD. The techniques proposed in this research seek to identify the "voice of the customer" for a SNE product with a focus on VAQ. Users may apply these techniques to both for-profit commercial and government entities. Further these locations and the related techniques developed within this research to improve the VAQ of a SNE may be applicable not only military applications, but also commercial aviation as well as police and fire rescue applications.

Research Question

The question emerges, assuming an "interoperability" (defined further below) perspective taken previously by Goldiez (1995), what is the power of SNE visual aesthetic quality factor forecasting in predicting user preferences for individual use-cases and the simulation interoperability community at large? The research assumes that a better insight into SNE visual aesthetic quality (VAQ) factors will be required to improve future interoperable SNE development and VV&A. As described more fully in Chapter 3, this dissertation considers either conceptually or through case study methodologies the following general research questions:

> *Does the Delphi technique produce VAQ factors and factor preference forecasts during the concept phase of SNE that are consistent with end User generated assessments?*
>
> *Does Conjoint Analysis improve the understanding of the significance and power of identified VAQ factors and preferences?*
>
> *What are the set of primary factors, priorities, and interactions that most affect VAQ and utility of synthetic natural environments for an interoperable training use case?*
>
> *Can the information gathered from the Delphi technique and Conjoint Analysis supplement existing VV&A processes to create a new SNE VV&A paradigm?*

Answering these questions has potentially powerful implications to the future of SNE factor priorities and the life cycle development of a SNE. A detailed understanding of the major factors and their parameters can be used to derive recommendations to improve the realism of SNE (Goldiez, 1996). SNE developers can use analysis of these factors during the outset of terrain generation to ensure conformity to the parameters. Additionally, V&V agents can leverage these parameters as inputs for automated tools and process while accreditation authorities can utilize them for defining acceptable quality metrics. Optimistically, findings may also lead to improvements in Quality Function Deployment (QFD) and possibly add insights important to the construction of a stronger QFD House of Quality matrix.

Within the QFD developmental philosophy, expert panels must forecast critical factors and validate them against user populations through experimental design. If expert and user responses correlate, then researcher can apply statistical methods to extrapolate parameters of these critical factors based on results of the experiment. Expert panels, especially in a physical committee, have a tendency to isolate voices due to physiological barriers associated with conflicting personality's between members (Gordon, 2009; Hsu & Sanford, 2007). This research will address this deficiency through implementation of the Delphi Method.

Problem Domain and Background

This section will give readers and prospective users of this research additional background on core topics associated with this dissertation problem area. Due to the breath of M&S, SNE, and VV&A, this dissertation will provide readers with a high-level picture of these topics and point to influential works where readers may obtain greater knowledge.

Military M&S Domains

Military M&S has been traditionally composed of three simulation domains: *Live*, *Virtual*, and *Constructive* simulations that may interoperate with each other through various communication techniques. Readers may best understand the concept of these three domains by thinking about people, systems, and system operation. *Live* simulation is real people using real systems to participate in a simulated operation. A real person using a simulated system (or simulator) to participate in a simulated operation is engaged in a *virtual* simulation. Finally, simulated people using simulated systems to participate in a simulated operation are said to be in a *constructive* simulation (Tolk, 2012b). Table 1 provides a summary of these concepts.

Table 1. Live, Virtual, and Constructive Simulation Domains

Source: *Engineering Principles of Combat Modeling and Distributed Simulation* (Tolk, 2012b)

People	**Systems**	**Operation**	***Simulation***
Real	Real	Simulated	*Live*
Real	Simulated	Simulated	*Virtual*
Simulated	Simulated	Simulated	*Constructive*

These concept definitions fit with their formal definitions found within the DoD M&S Glossary:

> **Live simulation**: A simulation involving real people operating real systems (DoD Modeling & Simulation Coordination Office, 2013).
>
> **Virtual simulation**: A simulation involving real people operating simulated systems (DoD Modeling & Simulation Coordination Office, 2013).
>
> **Constructive simulation**: Simulations involving simulated people operating simulated systems. Real people can be allowed to stimulate (make inputs) to such simulations (DoD Modeling & Simulation Coordination Office, 2013).

Live simulation allows humans to train in the real operational environment and to experience the physical hardships of traversing terrain (R. D. Smith, 2009). The term “live simulation” is often thought to be a misnomer, but it is possible to think about live simulation as traditional maneuvers, except augmented with simulation devices to engage in mock combat. These mock combat events do not involve real munitions, but instead use global positioning systems, computer, and laser engagement technologies adapted to their weapon platforms to determine the outcome of a duel. Furthermore, the effect of artillery support, close air support, and landmines can also be simulated in this domain to enhance realism (Tolk, 2012b).

Virtual simulation describes a three-dimensional representation of a system that is operating within a three-dimensional environment. The virtual simulation domain is closely aligned with the commonly recognized term, “virtual reality”. In the Military M&S domain, the visual focus of virtual simulations is how objects such as soldiers and military vehicles appear on the battlefield. Virtual simulations are typically constructed to train, test, or measure and individual soldiers or collective team’s ability to respond in a desirable manner by immersing them in a system that generates visual, aural, and tactile stimuli (R. D. Smith, 2009). Flight and driving simulators are a popular example of application of the virtual simulation domain. This research will specifically focus on the *virtual* simulation domain; however, readers can readily apply this research other simulation domains for application. The focus of this dissertation will be on *virtual* simulation.

A constructive simulation represents the aggregation of objects, behaviors, and properties within a system. This simulation domain represents operations across large battlefields while adhering to a format that can run on reasonable computer hardware. Constructive simulation closely mirrors the organization, representation, and information that are used in a formal military organizational hierarchy for purposes of command, communication, and control (R. D. Smith, 2009).

M&S Synthetic Natural Environments

While live simulation is situated in the physical world or natural environment, virtual and constructive simulation requires a synthetically construction of the natural environment. Synthetic Natural Environments (SNE) describe the physical world which all models of a simulation system exist and interact, to include both data and models representing the elements of the environment, their effects on simulation entities, and the entities effect on the environment. A SNE can be present in all three M&S domains: Live, Virtual, and Constructive. Figure 2 illustrates a common representation of SNE components. The M&S domain provides several similar definitions as to what constitutes a synthetic environment representation. The Department of Defense (DoD) M&S Glossary defines an environmental representation as:

> A model, simulation, or database designed to produce an accurate and consistent data set for one or more parameters that characterize the state of the physical environment (DoD Modeling & Simulation Coordination Office, 2013).

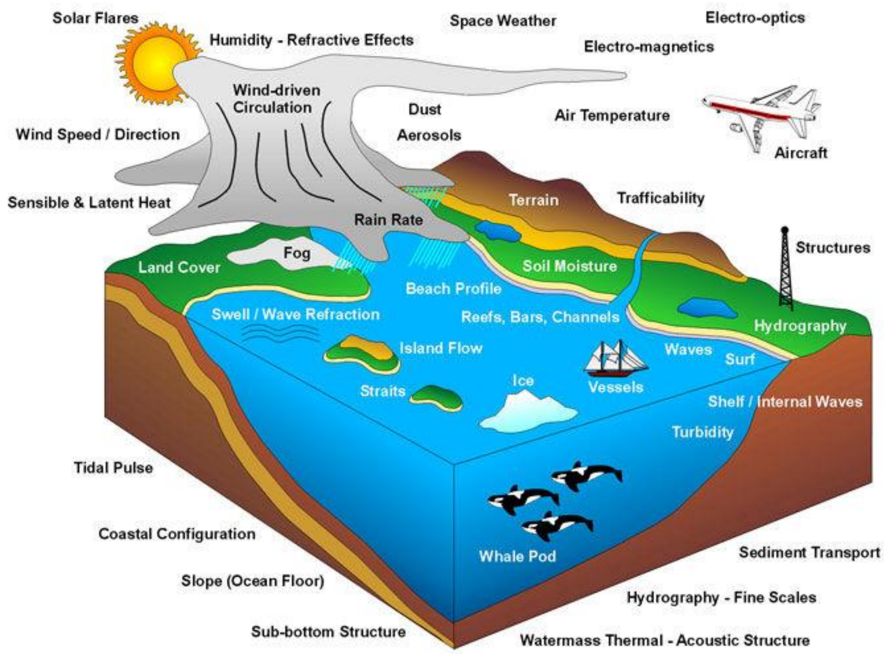


Figure 2. The scope of synthetic natural environment representations

Source: (Mamaghani, 2008).

Similarly, the Synthetic Environment Data Representation and Interchange Specification (SEDRIS) Glossary defines it as:

> An authoritative representation of all or part of the natural environment, including permanent or semi-permanent man-made features (SEDRIS, 2007).

The IEEE Standard for Distributed Interactive Simulation – Application Protocols (IEE Std 1278.1-2012) provides a comprehensive definition for SNE as:

> The integrated set of data elements that define the environment within which a given simulation application operates. The data elements include information

about the initial and subsequent states of the terrain including cultural features, and atmospheric and oceanographic environments throughout an exercise. The data elements include databases of externally observable information about instantiable entities, and are adequately correlated for the type of exercise to be performed. Also known as a virtual world (IEEE Computer Society, 2012).

Correlation and Interoperability

The emergence of distributed simulations in the 1980's brought forth enhanced capability to the military and civilian M&S communities (Wainer & Al-Zoubi, 2010). In order for distributed simulation to effective, M&S developers must pay special attention to the areas of *interoperability* and *correlation*. For distributed simulations, correlation is a measure of the quality of interoperability between virtual environments. Interoperability refers to the ability of a model or simulation to provide services to and accept services from other models and simulations, and to use the services so exchanged to enable them to operate effectively together (SEDRIS, 2007). Correlation is the convergence of the relationships between these interoperable data representations. In terms of interoperable SNEs, correlation can be defined as:

The convergent representation of the same physical environment in two or more separate environments prior to their use in a combined exercise with equal representation of environmental objects at comparable levels of presentation (SEDRIS, 2007).

In essence, a model or feature represented within interoperable SNE 'A' will also be represented in interoperable SNE 'B', based on pre-defined acceptable correlation tolerance. In distributed simulations, accurate correlation of SNE digital elevation models are a critical consideration.

Often, hardware and software requirements of a computer image generator may dictate limitations of the level of correlation provided by those systems and developers must make certain allowances to define an acceptable level of correlation. This process of accepting correlation between dissimilar simulation systems is known as “managed correlation” (SE Core, 2013).

SNE correlation has been a major issues plaguing the M&S industry since the development of distributed interactive simulation (DIS) standards (Standards Committee on Interactive Simulation, 1995). As a result, a large body of published research studies defines and offers solutions to issues of SNE correlation. Purdy & Goldiez provide an early analysis of correlation issues to DIS. In their early work on simulation interoperability, they abstracted the larger issue of interoperability between dissimilar simulators into four smaller, more manageable cases for consideration as summarized in Table 2.

Table 2. Four sub-classes of interoperability problems to facilitate the development a quantitative definition. Adapted from Purdy & Goldiez (1995)

Case Number	**Case Title**	**Case Description**
Case 1	Interaction/virtual incompatibilities	This type of interoperability problem arises when a characteristic or behavior of one type of virtual world entity is not recognized or is incompatible with the same characteristic or behavior of another entity. Firing munitions at an entity that it does not recognize and can not assess probable damage if hit is an example of this problem. A method used by some simulators is to convert the munition into an entity that is known and assess the resulting damage (e.g., treat an AK-47 round as if it were an M-16 round). Such an approach reduces the magnitude of this class of incompatibility but increases the problem of non- uniform simulation fidelity incompatibilities (Case 3).

Case Number	Case Title	Case Description
Case 2	Low interactive fidelity	Virtual world entities should interact with each other and with the environment according to a set of rules such as the laws of physics. Using such laws, vehicles can not drive through solid objects. When entity interactions defy the rules, training efficacy is degraded.
Case 3	Non-uniform levels of simulation fidelity	This interoperability problem occurs when two simulators model objects or behavior at differing levels of fidelity. Such differences skew interaction results and degrade training efficacy. For example, if one tank simulation considers in its mobility model the terrain type (sand versus clay) and adjusts its movement accordingly while another tank simulation does not consider terrain, the differing level of simulation fidelity may result in skewed interactive outcome. Another example of this is two simulators whose image generators differ in rendition capability (e.g., polygon loading, color gamut, & etc.)
Case 4	Differences in the virtual environment	An assumption made when connecting simulators is a single and common virtual environment. When the environment is not common, problems such as intervisibility, floating tanks and subterranean aircraft degrade the realism of the training scenario, skew the interaction outcome, and degrade training efficacy.

In their research, Purdy & Goldiez utilized a combination of data acquisition and manual human factors experiments, including a complex visual scene study, to identify several primary visual factors (Cases 3 & 4) that affect SNE correlation of a virtual environment and its rendition on dissimilar simulators: Luminance, feature size, feature position, and feature texture (Purdy & Goldiez, 1995). Also relevant to case 4, Schiavone *et al*. implemented statistical Bernoulli trials to sample the digital elevation of points between databases and determine a correlation delta between them. Both Goldiez (Goldiez, 1996) and Santiago *et al*. further describe the use of automated correlation tools, based on statistics, a method for efficient calculation of terrain correlation; however, Santiago *et al*. notes:

"When it comes to correlation testing there is no single testing mechanism that can provide a thorough understanding of correlation between databases." (Santiago, Verdesca, Watkins, & de la Cruz, 2012)

Missing from the Purdy & Goldiez case-based definition of interoperability issues are the issues associated with federates using different local time management mechanisms and differences in communications update protocols. Fujimoto & Weatherly describe the local time management issue of interoperability and demonstrate how the time management component of High Level Architecture (HLA) can solve this issue (Fujimoto & Weatherly, 1996). The concept of 'dead reckoning' in a virtual simulation refers to "each object [in the simulation] extrapolates the new positions of remote objects from the states last reported by those object" and was established to address the interoperability issue of distributed simulations utilizing disparate communication update protocols (Calvin et al., 1993; Martin, Jewett, Hollander, & Hicks, 2007). The CPU has traditionally calculated virtual simulation dead reckoning, but modern advances in Graphics Processing Units (GPU) mean GPUs can now dead reckon large quantities of entities much more quickly than the CPU (Martin et al., 2007). Note that these additional issue cases are outside the scope of the research presented in this dissertation.

This research will focus on the physical terrain representation of a *virtual* synthetic natural environment (SNE). The terrain representation, commonly referred to as a terrain database (TDB), is well defined by SEDRIS as:

> The depiction of the terrain environment, which includes data on the location and characteristics of the configuration and composition of the surface of the Earth, including its relief, natural features, permanent or semi-permanent man-made features and related processes. It includes seasonal and diurnal variation, such as grasses and snow, foliage coverage, tree type, and shadow (SEDRIS, 2007).

The TDB also includes the terrain skin which is the geometrical portion of the terrain representation that model's the Earth's surface, including terrain polygons, vertices, and vertex normals (SEDRIS, 2007).

The M&S community utilizes the terms SNE and TDB interchangeably. This research will make every effort to utilize 'SNE' when talking about synthetic terrain, although TDB may appear throughout the research based on the literature.

The major elements of a TDB include the terrain surface, cultural and natural features, textures, environmental data, and 3D models (Mamaghani, 1995). Combined, Graniela and Tolk both provide an comprehensive survey of these major components (Benito Graniela, 2011; Tolk, 2012a)

TDBs can be represented in many forms which contrast sharply based on their use cases. Google Earth™ is a TDB that perhaps most readers are familiar with. While Google Earth is an excellent tool for visualization of global terrain coverage, it does not provide sufficient fidelity for computer simulation entities to inhabit (Benito Graniela, 2011). This is due to the lack of terrain and feature attribution required for entities to interact with and traffic the terrain. Attribution is the additional visible and non-visible terrain and cultural information that specify the state of an environmental object and provide entities and models the ability to sense, plan and navigate in the environment (Benito Graniela, 2011; SEDRIS, 2007). Likewise, TDBs found in the modern video game industry, such as the first-person shooter Call of Duty series, may be sufficiently detailed for linear story-driven simulations, but may lack the large scale terrain representation for unified collective, or team-based, training, as well as the sufficient requirements to ensure military accreditation as a training aid (M Peck, 2012).

Ladner and Shaw provide an excellent overview of TDBs used throughout multiple industries, namely city planning, aerospace, manufacturing, and education (Ladner & Shaw, 2001). In his formative work, Tolk provides a detailed commentary on the major features of a TDB representation and the importance of each of these components to military M&S. He provides a thorough overview of common TDB standards and architectures, to include the SEDRIS U.S. Government terrain standard and its underlying models (Tolk, 2012a).

Terrain Database Generation

TDB generation describes the process of stitching together multiple external geospatial data sources through manual and automated processes to export a TDB runtime format for a targeted computer image generator (CIG). These external data sources are obtained from the real-world through Geospatial Information Systems (GIS). The basic types of data required for TDB generation are remote sensing data (i.e. satellite imagery), triangular information networks (TINs), and thematic layers: raster data (i.e. elevation model) and vector data (Lashlee, Bricio, Holcomb, & Richards, 2012). Features in the form of points, lines, and polygons describe Vector data, while arrays or grids of data values and images represent Raster data (Benito Graniela, 2011). Lashlee *et al.* describe each of these external sources and layers in great detail and discusses their importance to combat M&S (Lashlee et al., 2012).

Feature attribution describes the thematic layers of Vector and Raster data. Attribution describes important meta-data required for simulation entities to reason against the TDB. This can include surface material type, road widths and direction, building heights, hydrology depth, etc. (Benito Graniela, 2011). Feature attribution is also critical when constructing a TDB, since different features may assert an order of importance during the generation process. For example, a deep river may cause TDB generation software to generate a bridge when a road crosses over it, whereas a shallow one will instead generate a culvert. Attributes may also enable dynamic feature rendering so that entities actually have an impact on the terrain feature or surface that they are interacting with like a missile destroying a bridge or heavy construction equipment modifying ground elevation. TDB Developers use data dictionaries to attribute features to common industry standards. This is especially important in the case of interoperable simulations in order to maintain a fair fight and simulation correlation. Two common attribute data dictionaries used in the M&S industry are the SEDRIS Environment Data Coding Standard System (EDCS) and the Digital Geographic Information Exchange Standard (DIGEST) Feature and Attribute Coding Catalog (FACC).

TDB generation is not a recent development. Developers have been implementing TDBs for combat applications for at least 40 years (Schnitzer, 1976). Graniela provides a comprehensive survey of significant TDB generation systems used for military M&S applications (B. Graniela & Proctor, 2012)

There are no standard methodologies for the generation of simulation TDBs. Different approaches will be leveraged depending on available tools, data sources, and target simulation platforms (Mamaghani, 1995). Experts agree; however, that there are a set of common phases shared by all TDB generation processes (Lashlee et al., 2012; Mamaghani, 1995). The primary steps for generating a TDB are Requirements Definition, Data Collection, Value Adding, Transformation and Tailoring, Assembling the Database, and Compiling the Database for Transmission. Mamaghani expands on each of these phases in detail and further discusses tradeoffs that need to be made at each step of the TDB generation process (Mamaghani, 1995). Figure 3 illustrates the most basic form of a TDB generation process.

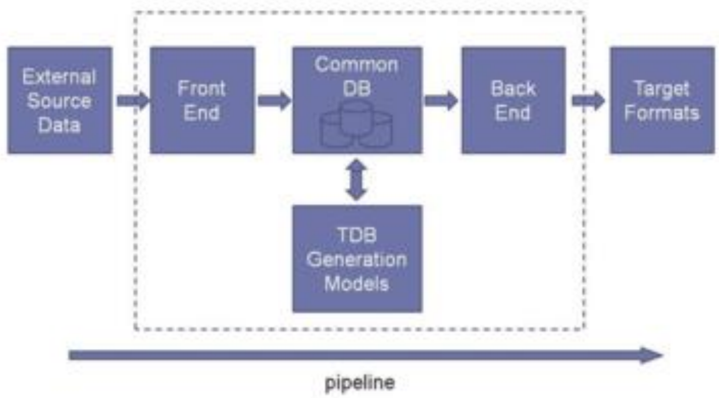


Figure 3. Generic terrain database (TDB) generation pipeline

Source: (B. Graniela & Proctor, 2012)

Beyond the simplistic theoretical generation pipeline put forth by Graniela and Proctor, production houses, such as the Synthetic Environment Core (SE Core), overseen by U.S. Army PEO STRI, produce runtime database products for multiple user needs.

Figure 4 below shows the generation pipeline for PEO STRI.

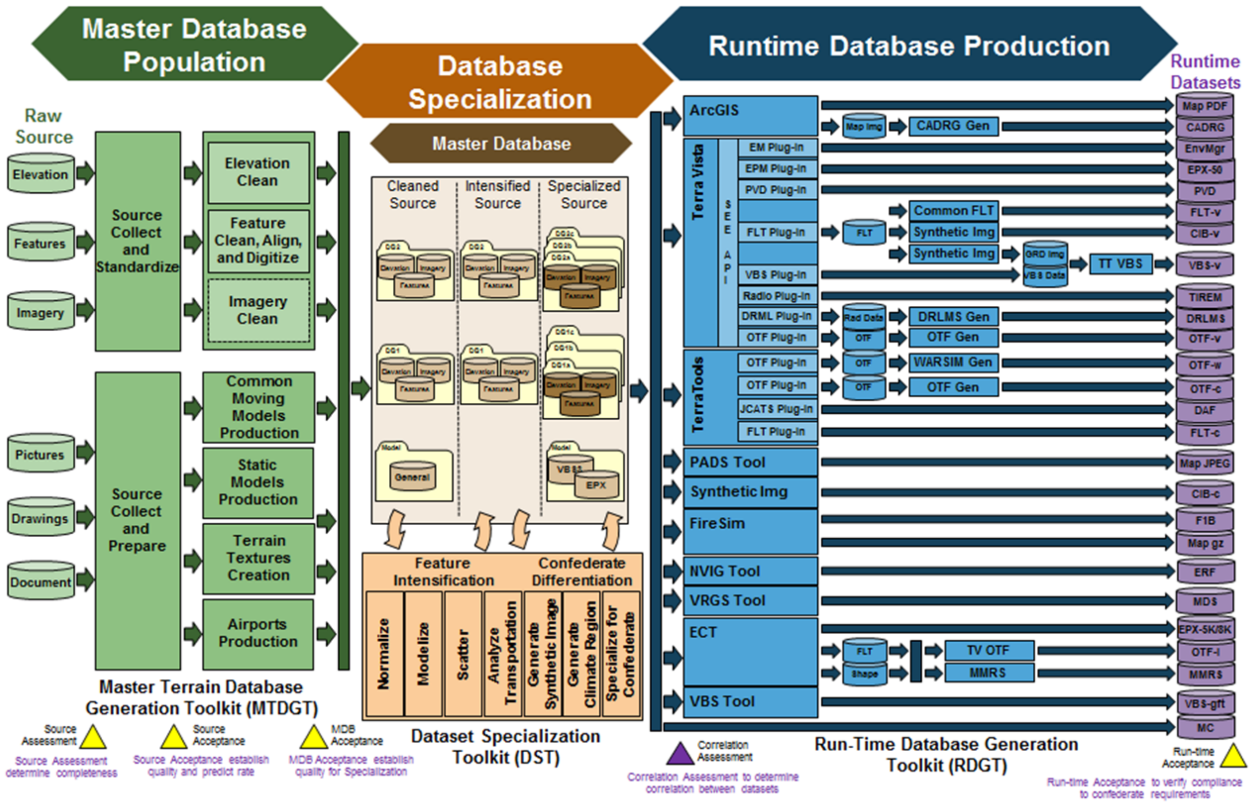


Figure 4. SE Core Standard Terrain Database Generation Capability (STDGC) Diagram

Source: (U.S. Army PEO STRI, 2015)

SE Core's SNE generation process is divided into three major phases and is permeated throughout with varication and validation activity. The first phase of this process is Master Database Population. In this phase, SE Core SNE developers collect raw GIS source data, standardize the data, and prepare it for data specialization. SE Core sources this GIS data from a wide variety of data repositories to include the United States Geological Survey (USGS), National Geospatial-Intelligence Agency (NGA), Department of Homeland Security (DHS), and other commercial and U.S. Government agencies. Because of this heterogeneous collection of source data, SE Core developers standardize the data attribution to a common format –EDCS. Additionally, any aerial imagery collected this phase must be orthorectified and color balanced to composite a complete imagery mosaic across the SNE. This phase also include developing the 3D Model content and other artistic assets needed for SNE rendering. Since SE Core supports several simulation engines of varying fidelity, often developer must generate 3D models with multiple levels of detail (LODs).

The second major phase of the SE Core SNE generation is known as Data Specialization. In this phase, data from the Master Database (MDB) is *Intensified* and *Specialized*. Intensification refers to the "value adding" of the cleaned source data. This includes procedurally generating 3D Model features, adding feature scatter (vegetation, buildings, etc.), and normalizing any inconsistencies across the data. Specialization is the process of reading the source data for a given confederate runtime format. For example, if the target simulation system supports climate regions, developers must specifically add extra attribution to account for this. This is also the phase where developers configure which 3D models and textures are used for various levels of detail rendering based on SME guidance and target simulation system performance parameters. For example, the AVCATT system displays higher LODs at long distances since rotary-wing aircraft can 'see' far into the distance through their sensors, whereas CCTT, a ground based simulator, only displays high LODs at close range since it is more concerned with the up-close fight. SNE developers must carefully balance the rending of LODs with system performance.

The final major phase in the SE Core SNE generation process is Runtime Database Production. This phase is concerned with building the final runtime SNE formats for each of SE Core's confederate programs. Since SE Core supports many simulation programs, each requiring varying levels of fidelity across the Live, Virtual, Constructive, and Gaming domains, developers must use an array of SNE generation software, each with a variety of SE Core specialized software plug-ins. SE Core implements incremental iterative builds when generating SNE products. This allows developers to continually test and verify their SNE products throughout their generation. Depending on the size and complexity of the SNE, this entire SNE generation process can last anywhere between nine to twelve months.

Outside of the three major phases of this process,

Figure 4 also illustrates a series of events along the bottom of the process chart. Each of these is a critical verification and validation milestone event that the SNE must undergo and pass prior to moving into the next stage of production.

Immersion, Interactivity, and Realism

Three factors determine the quality of virtual environments: content, interactivity, immersion, and presence. The content and the sensory stimuli together can cause changes in the user's psychological and physiological state (Whitton & Loftin, 2009). The research presented in this dissertation focuses on determining the quality of the virtual environment *content* as a first step to understanding immersive and interactive qualities. A virtual environment is said to be *interactive* when a user performs an action that generates an almost immediate input to the system (Whitton & Loftin, 2009). *Immersion* is a psychological state characterized be perceiving oneself to be enveloped by, included in an interacting with an environment that provides a continuous stream of stimuli and experiences (Witmer & Singer, 1998). Factors that affect *immersion* include isolation from the physical environment (physical immersion), perception of self-inclusion in the virtual environment (mental immersion), natural modes of interaction and control (physical immersion), and perception of self-movement (mental immersion). A virtual environment that effectively isolates users from their physical environment, thus depriving them of sensations provided by that environment, will increase the degree to which they feel immersed in the virtual environment (Witmer & Singer, 1998). The definition of *presence* is slightly more difficult to obtain as it is sometimes contested throughout the M&S community (Slater, 1999). Witmer and Singer offer two useful descriptions of *presence*:

1. "Presence is defined as the subjective experience of being in one place or environment, even when one is physically situated in another."
2. "Presence refers to experiencing the computer generated environment rather than the actual physical locale." (Witmer & Singer, 1998)

This definition is important because it highlights the fact that *presence*, like mental immersion, is subjective but are unlike *interactivity* and physical *immersion* that are objectively measurable (Cox, Cairns, Shah, & Carroll, 2012; Dede, 2009; Scoresby & Shelton, 2011). Self-reporting, task time completion, eye tracking, or a hybrid combination of these are modern attempts to measure the level of *presence* within virtual environments.

Ultimately, this research will focus on the factors affecting the aforementioned concepts as they apply to the quality of virtual environment *content*. Good immersion requires that the system first have acceptable content, to include models, sufficient object detail, sufficiently subtle behaviors of entities, etc. (Whitton & Loftin, 2009).

Winkler (2001) discusses limitations on the use of the word ‘fidelity’, i.e. the accuracy of the visual reproduction of the original on a display, to prediction of quality “even if sophisticated models of the human visual system are used” (Winkler, 2001). Winkler was able to achieve more reliable quality ratings for images and identified “sharpness” and “colorfulness” as important attributes. Ferwerda (2003) identified three types of “realism” – “physical”, “photo”, and “functional” – that corresponded respectively to equivalence of “visual stimulation”, “visual response”, and “visual information” in the scene (Ferwerda, 2003). Mourkoussis et al. (2010) found that human visual cognition is relatively unaffected by fidelity level as long as the visual scene looks acceptably realistic and that simulations do not degraded visual fidelity in the form of decreased polygon count and rendering performance (Mourkoussis et al., 2010). Moorthy & Bovik (2011) found the assessment of visual aesthetics “is highly subjective and…is a far more difficult problem than that of [visual] quality assessment” and that future studies into visual appeal should allow for ratings of quality, aesthetics, and content (Moorthy & Bovik, 2011).

Qualitative to Quantitative Visual Aesthetic Quality Analysis

Subjective aesthetic appeal plays a critical role in how one perceives quality and utility of everyday products and services. Factors such as age, experience, education, community, venue, and other contextual factors impact aesthetic appeal (Bloch, 1995). Within SNE for industrial and government applications, aesthetics also plays a part as one may reject an artistically created virtual character as “uncanny” while another accept the same character (Hodgins, Jörg, O’Sullivan, Park, & Mahler, 2010). Rejection by a user may result in them disengaging from the entire SNE experience due to “repulsion” (Hodgins et al., 2010). Beyond aesthetics, content of the scene and aspects of realism also feeds into user expectations, where unfulfilled expectations may negatively affect the perception of the suitability of not-so-realistic SEs (Herz & Macedonia, 2002; M Peck, 2012; Warfare History Network, 2014). Unfulfilled expectations undermine the use of SEs for industrial or government training applications, as a negative first impression may yield waning interest or engagement, possibly to the point of complete disengagement (Beeland, 2002; Dobrian et al., 2011).

Verification, Validation, and Accreditation

Verification and Validation (V&V) is an integral part of the DoD Systems Engineering process (Figure 5) and must be integrated throughout all Modeling and Simulation (M&S) activities, including model selection, development, and integration (Tolk, 2012c).

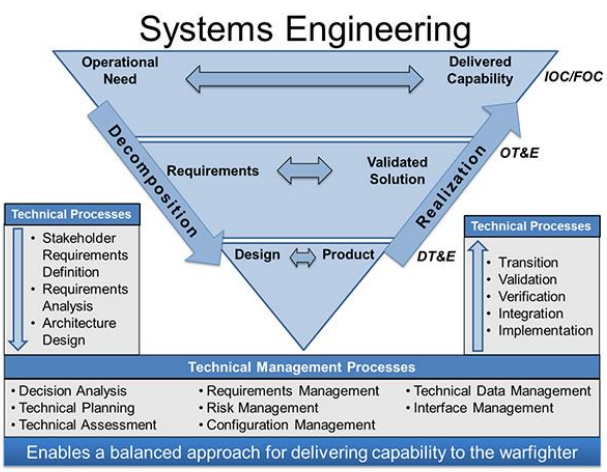


Figure 5. DoD System Engineering Process Model

Source:(Defense Acquisition University, 2014)

As an M&S procedure, *verification* is typically defined as the process of determining if an implemented model is consistent with its specification (Petty, 2010). Verification also explores if a designed model will satisfy the requirements of its intended application or use case. The DoD Modeling and Simulation Coordination Office (M&SCO) M&S Glossary defines *verification* as:

> *The process of determining that a model or simulation implementation and its associated data accurately represent the developer's conceptual description and specifications.* (DoDI 5000.61)

Validation refers to a testing process and determines the degree to which a model is an accurate representation of the simuland. *Validation* examines representational accuracy. This required accuracy should be considered with respect to the models intended use case (Petty, 2010). The M&SCO M&S Glossary defines *validation* as:

> *The process of determining the degree to which a model or simulation and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model.* (DoDI 5000.61)

Practitioners often groups *Accreditation* with V&V, but is an entirely different process. V&V processes focus on technical test cases, where *accreditation* is a non-technical decision process. Accreditation is the official certification by a responsible authority that a model is acceptable for use for a specific purpose (Petty, 2010). The M&SCO M&S Glossary defines *accreditation* as:

The official certification that a model or simulation and its associated data are acceptable for use for a specific purpose. (DoDI 5000.61)

M&S users conduct Verification, Validation, and Accreditation (by extension) to avoid three main error categories concerning the use of M&S. A *Type I Error* is that a valid simulation result is not accepted (Tolk, 2012c). Type I Error is the model developers risk and can result in a waste of model development costs if the model is never used. Any potential benefits that using the model might have produced, such as reduced training costs or improved decision analyses, are lost (Petty, 2010). A Type II Error is that non-valid simulation results are trusted and used. This is a model user's risk and is much more serious than a Type I error (Tolk, 2012c). This may occur when validation is done incorrectly but convincingly, erroneously persuading the accrediting authority to certify the model for use (Petty, 2010). A Type III Error occurs when an irrelevant model is used for an unintended target application or use case. This differs from Type II error in that the model is in fact valid for some purpose, but not suitable for the intended application. A Type III error is the model accreditor's risk. Table 3 summarizes the types of VV&A errors and risks.

Table 3. Summary of Verification, Validation, and Accreditation Errors (Petty, 2010)

	Model valid	Model not valid	Model not relevant
Results accepted, model used	Correct	**Type II error**. Use of invalid model; Incorrect V&V; Model user's risk; **More** serious risk	**Type III error**. Use of Irrelevant model; Accreditation mistake; Accreditor's risk; **More** serious risk
Results not accepted, model not used	Type I error. Nonuse of Valid model; Insufficient V&V; Model builder's risk; **Less** serious error	Correct	Correct

Balci proposed four primary categories for V&V that M&S professionals can use to support V&V activities: *Informal*, *Static*, *Dynamic*, and *Formal* testing methods (Balci, 1998). Table 4 provides a summary of these methods.

Informal methods are those that rely heavily on human intuition and subjective evaluation without rigorous mathematical analysis. Subject matter experts mainly conduct these tests based on their experience with comparable solutions that can be used as a reference (Tolk, 2012c). V&V of synthetic natural environments largely falls under this category. Experts assert that almost all synthetic terrain testing is done via visual "flyovers" or "driving around" to search for visual anomalies (Santiago et al., 2012). This results in V&V agents not thoroughly testing terrain, with only patches of the environment being evaluated closely. Although in the case of SNE Core (and presumably others), priority areas of interest (AOIs) are reviewed first.

Additionally, V&V agents can randomly sample and evaluate representative areas to provide users with a high-degree of confidence in the accuracy of those areas across the SNE.

Static methods are those that conduct an assessment based on the characteristics of code and model design without actual execution. Experts merely evaluate a model "blueprint" (Tolk, 2012c). Static methods are more often performed by model developers, as compared with informal methods, which depend more on subject matter experts (Petty, 2010).

Dynamic methods assess a model or simulation by executing the system and evaluating results. The evaluation may involve comparing the results with data describing the behavior or the simuland or the results of other models. Since the comparisons in dynamic methods are typically of numerical results and data, dynamic methods are generally objective and quantitative, but may not be entirely encompassing based upon the robustness of the simulation and quality of scenario execution (Petty, 2010). The bulk of the research presented in this dissertation will focus on V&V through dynamic methods, specifically through regression analysis and hypothesis testing.

Formal methods perform V&V through rigorous mathematical and statistical proofs and correctness (Tolk, 2012c). Statements about the model are developed using a formal language or notation and manipulated using logical rules; conclusions derived about the model are unassailable from a mathematical perspective. These methods are quite difficult to apply in practice, as the complexity of most useful models is too great for current tools and methods to deal with practically (Petty, 2010)

Table 4. Summary of V&V Methods (Balci, 1998)

Informal Testing Methods	**Static Testing Methods**	**Dynamic Testing Methods**	**Formal Testing Methods**
Audit	Cause-effect graphing	Acceptance testing	Induction
Desk checking	Control analysis	Alpha testing	Inductive assertions
Documentation checking	Data analysis	Assertion testing	Inference
Face validation	Fault/failure analysis	Beta testing	Logical deduction
Inspection	Interface analysis	Bottom-up testing	Lambda calculus
Reviews	Semantic analysis	Comparison testing	Predicate calculus
Turing test	Structural analysis	Statistical techniques	Predicate transformation
Walk-throughs	Symbolic evaluation	Structural testing	Proof of correctness
…	Syntax evaluation	Submodel/correctness module testing	…
	Traceability assessment	Visualization/animation	
	…	…	

Figure 6 again illustrates the SE Core SNE generation process as an example. As alluded to in previous section, SE Core implements a series of V&V milestone events as depicted by yellow and purple triangles across the bottom of this process chart. The first three of these V&V milestone events occur during the Master Database Population phase. The first event, *Source Assessment*, is an informal testing method designed to give SNE developers and managers peace of mind that source collection personnel are obtaining source data commensurate with the SNE requirements and allotted budget. During this assessment, the SE Core management personnel use informal auditing and inspection to review the raw source data for gaps in geographic coverage, especially near areas of importance (AOIs). Once SE Core engineering management is satisfied with the quality and coverage of this raw source data, SE Core developers can begin to clean and prepare the source data for the next V&V milestone.

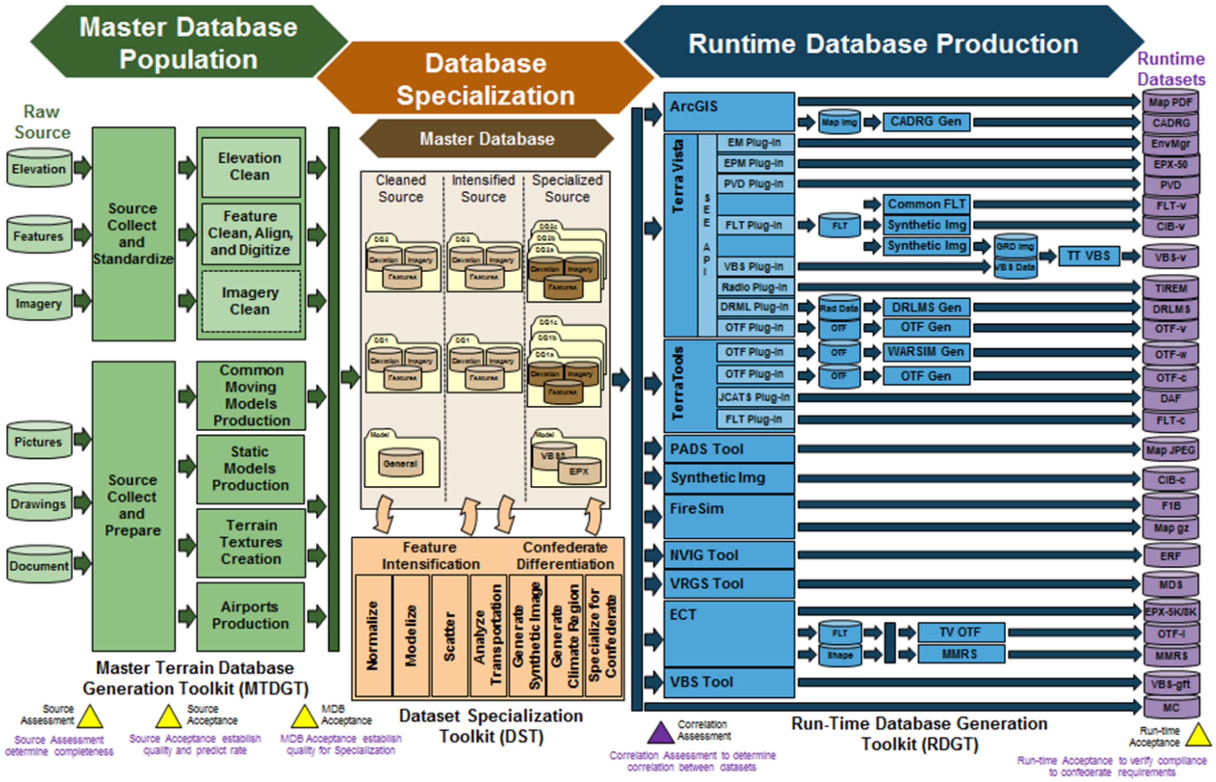


Figure 6. VV&A events highlighted in the SE Core Standard Terrain Database Generation Capability (STDGC) Diagram

Source: (U.S. Army PEO STRI, 2015)

SE Core conducts a series of *Source Data Acceptance Tests* as SE Core developers complete cleaning and preparation of the three major source data components: Imagery, Vectors, and Models. Each of these three SNE component acceptance tests are again broken down into a 50%, 90%, and final validation event. This allows SE Core managers to provide guidance and change early in the source data standardization process. It is critical to ensure that source data is properly validated and standardized prior to committing it to the master database, since future changes to this source will be much more costly downstream during runtime database generation. SE Core

V&V agents implement a combination of Informal and Dynamic Testing methods for these Source Data Acceptance events. For example, during the Vector Acceptance Test, SE Core V&V agents run automatic scripts to ensure that the vector data meets proper attribution for the EDCS standard. SE Core managers then review the reports generated from these scripts and conduct a broad visual inspection of the vector data to verify the correctness of the reports. Similar activities occur during the Imagery and Model Acceptance Tests.

As a final sanity check, SE Core conducts the MDB Acceptance Test prior to committing data to the Master Database. This is largely an Informal Testing Method, designed to give SE Core and its stakeholders a big picture look at the combined entirety of the SNE source data. SE Core V&V agents use visual inspection, combined with the automated validation reports, to ensure geographic data coverage, attribution, and alignment of vector data to imagery (if required).

After SE Core developers specialize, intensify, and pass the MDB data to the runtime software tools, V&V agents run a Correlation Assessment. This automated process is a dynamic testing method that quantifies the delta of correlation between SNE representation formats. This assessment takes into account SNE elevation, 3D models, and feature attribution. In a process known as *Managed Correlation*, SE Core stakeholders and accreditation agents determine acceptable levels of correlation mismatch between SNE formats. SE Core V&V agents conduct this assessment throughout the SNE generation process in order to account for errors introduced through iterative SNE builds.

Runtime Acceptance is the final V&V milestone of the SE Core SNE generation process. SE Core implements a process called the Major Evaluation of Geospatial Areas (MEGA) or MEGA

Review for short. While stringent, this process relies on Informal Testing Methods primarily conducted through SME and Stakeholder visual inspection. SE Core recruits SMEs from across the Army to evaluate each of the SNE runtime formats. This includes SMEs familiar with the target simulation platforms as well as SMEs familiar with the real-world geographic location that SE Core is developing as a SNE. This allows SMEs to evaluate both the visual content of the SNE as well as the SNE runtime performance on the target simulation system, to include Levels of Detail (LODs), electro-optical and infrared (EO/IR) sensor representations, and image generator (IG) overload potential. While the developers of each simulation system pre-define the number of LODs and the recommended transition range for each target runtime system, the LOD transition range can be modified in order to optimize SNE performance.

SE Core breaks the MEGA Review into two major milestones for each SNE format: Initial Site Review and Final Validation. The Initial Site Review provides SMEs with a chance to provide feedback on the SNE while at the 50-60% completion level. This allows them a chance to influence the visual SNE development relatively early in the process, before SNE developers complete major irreversible work on the runtime formats. SE Core then invites the SMEs back for the Final Validation Event once SE Core management considers the database to be in a complete or near complete status. SMEs again provided comments and feedback on potential issues. These issues are recorded as Discrepancy Reports (DRs) and a committee of SE Core stakeholders and SMEs assign a severity to the DRs based on the potential impact to training. The SE Core accreditation agent having reviewed and accepted the list of DRs determines successful completion of the MEGA Review. The content and severity of the resulting MEGA Review DRs can be very subjective, especially since different SMEs have varying interests and loyalties to a given component of the SNE. For example, Aviation SMEs may be more concerned on the accuracy of an airfield model than that of a Ground/Armor SME; whereas, a Ground SME is more concerned about proper transportation network attribution than an Aviation SME would be. Additionally, this subjectivity can be further complicated by levels of experience between SMEs, especially in the area of modeling and simulation.

VV&A User Roles and Responsibilities

There are six primary roles defined to account for the V&V process. In order to ensure V&V is as objective as possible, it is strongly recommended that the set of people conducting V&V are not liaised with the simulation developers. These affiliations can hinder the ability to conduct independent assessment by obscuring objectivity (Tolk, 2012c). In order to minimize costs and risks associated with redundant V&V efforts, M&S developers should not disjoint V&V development processes since it is imperative that errors and issues discovered be reported as early as possible. The six roles defined by the DoD MSCO are provided below (DoD Modeling & Simulation Coordination Office, 2011). It is important to note that these definitions are industry agnostic and do not solely relate to the defense industry.

1. **M&S User**: This is the person or group responsible for the application of the simulation. The user is the entity who defines the requirements, establishes simulation fitness criteria, determines by what means the simulation will be accredited, and ultimately accepts the results of accreditation. This research assumes a user group of U.S. military soldiers for the study of TDB quality factors, although users of the research results could be any organization seeking to improve TDB quality.

2. **M&S Program Manager (M&S PM)**: The M&S PM is the entity responsible for planning managing, and directing resources for simulation development.
3. **M&S Developer**: This entity is responsible for constructing the simulation and providing technical insight and expertise to the other V&V roles. The M&S Developer is the TDB architect. They are responsible for sourcing and stitching together the external source data formats into a functioning runtime terrain format. The research presented in this dissertation can be provided to the M&S Developer to affect TDB quality upstream early in the TDB generation process.
4. **Verification and Validation Agent (V&V Agent)**: The V&V agent is the entity responsible for proving a simulations fitness for the intended use by faithfully carrying out all V&V tasks. The results generated from this dissertation research can be fed as input to the V&V Agent’s tasks to ensure that the TDB conforms to acceptable quality standards, especially in the case where automated V&V tools are being leveraged.
5. **Accreditation Agent**: This entity is responsible for conducting accreditation activities for the simulation. They provide guidance to the V&V agent for providing necessary evidence for simulation fitness in the form of V&V reports.
6. **Subject Matter Expert (SME)**: The SME’s role is to provide specialized insights and knowledge to all the V&V roles into the systems that are being modeled (Tolk, 2012c). A SME for a TDB may be a live-fire range operation officer, a master gunner, a local representative from the geographic area, or anyone else with a knowledge of the terrain and how it will be used to support a simulation.

At each step in the VV&A process, these 6 roles are responsible for one of six different tasks: Perform, Assist, Lead, Monitor, Review, and Approve (DoD Modeling & Simulation Coordination Office, 2011). Table 5 delineates these responsibilities against sample VV&A processes.

Table 5. Comparison of V&V User Roles and Tasks

Activity / Role	User	M&S PM	Developer	V&V Agent	Accreditation Agent	SME
Define Requirements	Lead Approve	Monitor	Assist	Review	Review	Assist
Define Measures	Lead Approve	Monitor	Assist	Assist	Assist	Assist
Develop V&V Plans	Review	Assist Approve	Review	Lead	Assist	
Verify Requirements	Lead Approve	Monitor	Assist	Lead (primary)	Assist	Assist
Verify Design	Approve	Monitor	Assist	Lead		Assist
Implement Design		Monitor Approve	Perform			
Verify and Validate Data	Approve	Monitor	Assist	Lead		Perform
Test Implementation	Approve	Monitor	Lead	Assist		Assist
Validate Results	Assist Approve	Monitor	Assist	Lead		Assist
Prepare V&V Report				Perform		
Configure for Use	Assist | Lead Approve	Lead | Assist	Assist			
Conduct Accreditation Assessment	Monitor				Perform	Assist

Review	Participation normally limited to reviewing results of task and providing recommendations
Perform	Actually executes the task. Normally involves little active participation from others
Monitor	Observes the task to ensure it is done appropriately but does not normally participate
Lead	Leads the task. Normally involves active participation from others
Assist	Actively participates in task (e.g., conducting tests, providing information)
Approve	Determines when an activity is satisfactorily completed and another can begin. Determines what activity should be pursued next (e.g., whether to continue to the next scheduled activity or return to a previous activity).

Source: Adapted from the VV&A Recommended Practice Guide (DoD Modeling & Simulation Coordination Office, 2011)

Future Challenges and Issues

In summary, correlation issues rest with inconsistent representation of underlying elevation, feature (vector), and model data between distributed, but interoperating systems. Correlation issues often manifest themselves in elevation difference in number, selection, location, and numerical value of elevation posts. Correlation issues often manifest themselves in feature difference in number, selection, location, and attribution of feature (vector) data. Inconsistencies arise between levels of detail elevation, feature, and model representations and between different morphing algorithms between the levels of detail. Inconsistencies also arise in visualization due to hardware specific limitations. Currently developers must apply a combination of automated tools, interactive user V&V, and visualization to properly understand the level of correlation between the distributed, interoperating TDB's. There is no single correlation "silver bullet".

The aim of this research is not to study TDB correlation, but instead to explore and possibly better identify aspects of SNE source data and their impact to VAQ of distributed, interoperating TDB's that may result in development of uncorrelated TDB's. The hope is that through better understanding of the factors impacting SNE VAQ, improved interoperating TDB correlation may result.

As Schiavone *et al*. notes, one-hundred percent validity of terrain databases will likely never occur (Schiavone, Nelson, & Goldiez, 1994). This is especially true when almost all SNE evaluation is accomplished by conducting visual "flyovers" or in the context of "driving around"(Santiago et al., 2012). The introduction of automated statistical tools strengthens the validation of correlated SNE representations, but do not provide sufficient quality feedback for validity of a stand-alone SNE. The aim of this research is to identify these SNE factors and quantify limitations on factor values in hopes that developers of VV&A applications can implement them.

Overview

This first chapter of this dissertation provided the problem and motivation for this research. Chapter 1 also provided the reader with a high-level overview of the M&S domain as it applies to VV&A, SNE generation, and SNE interoperability issues. Chapter 2 analyses methods and practices from commercial marketing and quality control. Chapter 3 presents a thorough methodology to identify VAQ factors and quantify limitations affecting SNE quality by utilizing proven applications identified in Chapter 2. Chapter 4 provides an in-depth analysis of the research results and discusses the primary factors and factor parameters identified. Finally, Chapter 5 seeks to make recommendations for SNE quality analysis and generation based on the research results. Chapter 5 also provides discusses the strengths and weaknesses with the proposed methodology for applying this body of research by would-be adopters and provides recommendations for further research efforts.

CHAPTER TWO: STATISTICAL TECHNIQUES FOR FORECASTING AND OBTAINING USER FEEDBACK

In order to identify methods for addressing quality factors and parameters, this research will explore industries and research communities outside of the modeling and simulation domain. This section will identify proven statistical decision-making, market research, and quality practices utilized across a number of industries and domains to detect and measure the “voice of the customer”.

Hiring the Best and Brightest

Robust investments on talent management and recruitment is one way which many companies and organizations have been able to increase the quality of their products. Former Apple, Inc. CEO, Steve Jobs demonstrates one of the best examples of this in commercial industry. Jobs and his top executives never compromised with the talents and qualifications required of their employees. Steve Jobs believed hiring was his most important duty. Jobs focused his hiring on what he called A-List players. He firmly believed that an A-List person could accomplish 50 or 100 to 1 of that of a normal employee, thus he recruited the cream of the crop since a small group of A-list players could run circles around giant teams of B and C players (Elliot, 2012). Steve Jobs found that by having really good people, you don’t have to baby them and by expecting them to do great things, you can get them to do great things (Isaacson, 2015). Many top tier technology companies like Google, Facebook, IBM, and Amazon echo these sentiments on hiring the best of the best employees (Rawson, 2013).

For small to medium size companies, competing for talent can be difficult or even infeasible to replicate the hiring process of these multibillion-dollar technology powerhouses due to financial and prestige limitations. This is especially true in the Government sector, where potential top-tier employees can be turned-off based on relatively low salaries as compared to private industry as well as the metaphorical red tape of Government hiring practices. Patriotism, humanism, love of fellow man, religion, or other esoteric values may motivate some talented individuals in government. Organization can learn from these tech giants by applying some of their overarching hiring lessons (Rawson, 2013):

- Evaluate candidates based on tests that directly correlate to the work they are doing on the job.
- Create an inspiring vision that attracts candidates to your company.
- Treat your employees well.
- Be very selective and put a lot of time and attention into your recruitment process.

Iterative Design

The human element is only one aspect of VV&A. Others may duplicate successful methods and techniques of leading successful firms. Iterative design is another area in which Apple, Inc. leads. Iterative design is a methodology based on a cyclic process of prototyping, testing, analyzing, and refining a product or process over time to ultimately improve the quality and functionality of a design. In the Case of Apple, it is a process they *discover* the product through constantly creating new iterations. Whereas many companies may do six or seven prototypes of a product, Apple will do a hundred. Steve Jobs did not wake up one morning with a vision of iPhone in his head. He and his team discovered it through this exhaustive process of building prototype after prototype (Kahney, 2012). At Apple, these iteration cycles take 4-6 weeks at a time and are run many times over a product's development lifecycle. At each iteration, product designers pass the product to engineering program managers for test and evaluation. They are then passed back to the development team with comments for another iteration. This is an extremely costly approach, but is one of the reasons that Apple has a reputation for high quality products. The more you invest in design, the more likely you are to build incredible market changing products (Lashinsky, 2012).

Recently, the U.S. Government have begun to embrace the benefit of rapid and iterative prototyping as a means to stay innovative and relevant in an increasingly technological age. In his 2015 directive, Frank Kendall, the then Undersecretary of Defense for Acquisition, Technology, and Logistics (USDAT&L), specified the in use of increased prototyping as a key tenant to his Better Buying Power 3.0 policy (Under Secretary of Defense, 2015). Since this proclamation, the Department of Defense have stood up two new organization to better mirror the rapid and iterative prototyping capabilities of private industry. The first of these is the Defense Innovation Unit Experimental (DIUx) whose mission is serve as bridge between the U.S. military and companies operating at the cutting edge of technology. The goal of DIUx is to provide a mechanism to accelerate technology into the hands of soldiers by continuously iterating on how best to identify, contract, and prototype novel innovations through sources traditionally not available to the Department of Defense (Defense Innovation Unit Experimental, 2016). DIUx has setup outposts in the heart of technology hubs at Silicon Valley and Boston, MA.

The U.S. Army has adopted a similar strategy to DIUx, but focuses on materiel development versus rapid contracting mechanisms. The task of the Army Rapid Capabilities Office is to expedite critical technologies to the field through rapid materiel prototyping and delivery efforts to address immediate, near-term, and mid-term Combatant Commanders' needs. The office will incorporate early and prominent warfighter involvement into the requirements gathering and iterative prototyping process to ensure that materiel solutions are not only vetted by Army operators but also delivered to units as a holistic capability with the right support and tactics, techniques and procedures (TTPs) in place (Stalder, 2016).

From a SNE perspective, the Synthetic Environment Core (SE Core) program has embraced iterative prototyping and design as valid means for conducting V&V. SE Core incorporates SMEs and stakeholders early in the SNE generation process to provide them multiple opportunities to provide feedback on incremental builds of SE Core SNE products. This increases the level of trust between SE Core and the end user and ensures that there are no major surprises at the end of the process. This is however not a true iterative prototyping process because SE Core is still very much in control of the product and does not release control of the SNE to the end user for operational testing. Iterative prototyping would certainly provide much greater and earlier insight to SNE issues and deficiencies. On the other hand, iterative prototyping infers: (1) increased number of interactions between developer and operational user; (2) increased time necessary to enable such interactions; and (3) increased manpower/funding to speed timely incorporation of SNE changes before the next interaction. Thus, iterative prototyping would come at a significant cost and schedule impact outside the current SE Core program budget. Without increased funding to and time commitment from developer and operational user to support increased interactions and SNE turn around, adopting such a strategy would significantly reduce the SE Core yearly throughput of SNE products to the user. DARPA is well known in the simulation industry for fielding SIMNET through intense rapid prototyping (Loper & Turnitsa, 2012). SNE VV&A does not currently warrant the resources that DARPA designated projects receive (DARPA, 2016).

The Delphi Method

Developed in the early 1960's by the RAND corporation, the Delphi process is a decision-making technique that relies on the judgment of experts to achieve a convergence of opinion on a specific real-world issue (Hsu & Sanford, 2007). Researchers found that traditional round-table style discussion groups with the object of achieving a group consensus were plagued by negative psychological impediments, such as dominant personalities or those who had the tendency to want to appease all parties (Brown, 1968), a la "the loudest voice rather than the soundest argument may carry the day (Gordon, 2009)".

The Delphi reduces psychological factors, by eliminating the physical meeting of the committee altogether. Direct debate is replaced by a series of carefully planned and sequential questionnaires or surveys intermixed with feedback derived from computed consensus from earlier iterations (Brown, 1968). Often, researchers request members of an expert panel to provide reasons for their response, which is then subject to critique be fellow experts. The Delphi attempts to improve a committee's interaction by subjecting the views of an individual expert to group expert in an anonymous fashion in order to avoid the stigma of face-to-face confrontation. During this controlled debate, more often than not, an expert panel moves towards a consensus;

but in the event that this does not occur, the reasoning for dissimilar opinions is made obvious (Gordon, 2009).

A downfall of the Delphi is that number of respondents in a Delphi study is typically small; therefore, the process does not produce statistically significant results. The panel outcomes are traditionally not able to predict the response of a large population or even that of another expert panel. The value of the Delphi is represented by the expert assessments, the ideas it generates, as well as any differing opinions that are obtained (Gordon, 2009; Joint Research Centre, 2006). Expert selection, questionnaire rating scales, time allotments for conducting questionnaires, potential for low response rate, and unintentionally guiding feedback are factors that should be carefully considered when designing a Delphi study (Hsu & Sanford, 2007).

The Delphi Process

Figure 7 illustrates a high-level notional Delphi process through four iterations.

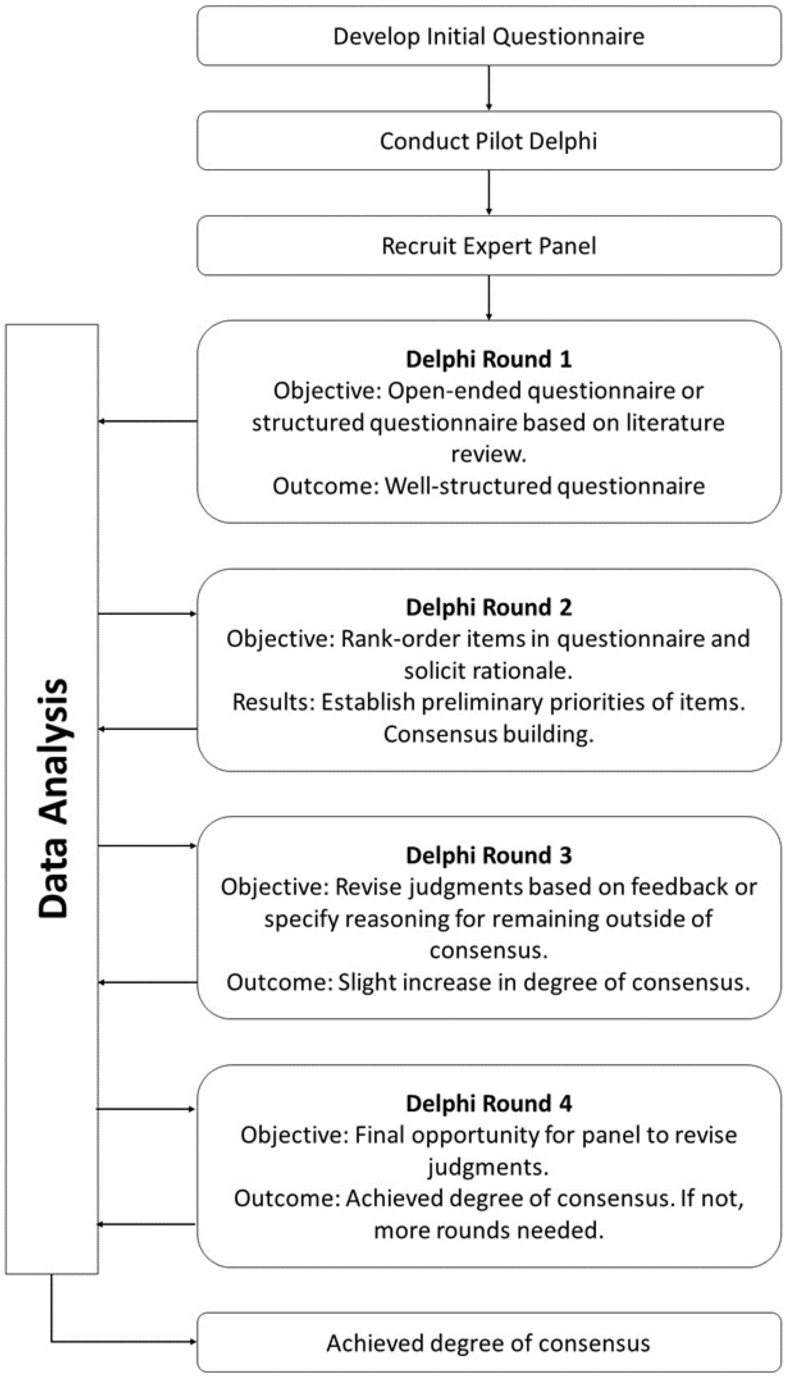


Figure 7. Notional Delphi Process.

Source: Adapted from Gill et al. (Gill et al., 2013)

The first step in the Delphi process is to construct a questionnaire. This questionnaire can be either open-ended, leaving the expert panel to provide specific content, or it can be a closed-form questionnaire based on extensive literature on a subject by the developer. The later should only be undertaken if basic information on the subject issue is widely available (Hsu & Sanford, 2007). Historically, the Delphi questionnaire was constructed and managed through traditional “snail mail”, but recent studies tend to utilize modern web-based survey tools to distribute questionnaires to respondents which has shown to improve expert response rate and reduce respondent dropout (Barrios, Villarroya, Borrego, & Olle, 2011; Gill, Leslie, Grech, & Latour, 2013).

Piloting a Delphi prior to execution is critical in assessing its validity. Researchers should select a trusted group of peers/advisors independent of the chosen Delphi expert panel to complete the pilot Delphi. The purpose of the pilot Delphi is to receive feedback and comments about the statements, process, survey instructions, and ease of completing the survey (Gill et al., 2013; Latour, Hazelzet, Duivenvoorden, & van Goudoever, 2009).

Selection of Delphi Expert Panel

Selection of the expert panel is the single most important step in the Delphi process since it will directly impact the quality of the results generated (Hsu & Sanford, 2007). Delphi subjects should be highly trained and competent within the specialized area of knowledge related to the target issue. Criterion for expert selection is limited and contested within the literature and remains ambiguous. Hsu and Sanford propose that experts be invited to participate if they have related backgrounds and experiences concerning the target issues, are capable of contributing helpful insights, and are willing and open to revise their initial or past judgments for purposes of attaining consensus (Hsu & Sanford, 2007).

The number of expert respondents needed for a Delphi is equally as perplexing; there is no exact size for a panel and can be variable from Delphi to Delphi. If the sample size is too small, critics may not consider the subjects as having provided a representative pooling of insight on the issue. Too large a sample size raises the possibility of low response rates and increased time obligations by respondents and the researcher (Hsu & Sanford, 2007). Most studies use 15 to 35 experts, although studies have been conducted with hundreds and thousands or people (Gordon, 2009). A good rule of thumb is to construct an expert panel by anticipating an acceptance rate between 35 and 75 percent. Gordon notes that researchers should take care to contact perspective experts directly and individually in order to foster a collaborative relationship. Experts should be provided with a description of the project, its objectives, the number of rounds to be included (or the time commitment anticipated), the promise of anonymity, and, if appropriate, a confirmation of the panelist's acceptance (Gordon, 2009).

Delphi Panel Feedback

Between Delphi iterations, it is important to exercise controlled feedback to reduce the effect of noise. The research accomplishes this by providing a summary of the prior iteration to the expert panel, which allows them the opportunity to gain additional insights on the issue and better clarify their answers provided in previous rounds. The application of statistical analysis in the feedback can also reduce the stigma of group conformity. Statistical analysis can ensure that opinions generated by each subject of a Delphi study are well represented in the subsequent iterations (Hsu & Sanford, 2007).

Delphi: Round One

Typically, the first found of a Delphi begins with an open-ended questionnaire for the purposes of soliciting specific information relating to the target issue (Hsu & Sanford, 2007). Researchers then collect the respondent's responses convert them to serve as a data collection instrument for the second round of the Delphi. As discussed earlier, it is acceptable to begin round one of the Delphi utilizing an initial closed-form questionnaire based upon extensive literature review if information on the target issue is widely available. This research will leverage an open-ended question approach due to limited amount of recent literature on this research topic.

Delphi: Round Two

In the second round of the Delphi, the survey asks the expert panel to review a second questionnaire summarized from responses of the first round. The panelists will then rank-order items for the purposes of establishing a priority of the target items. The Delphi panelists should also provide their rationale for their rankings, which researchers will use in follow-on rounds for consensus building. Round two will result in initial agreements and conflicts emerging from the expert panel. The researcher presents the outcomes and expert justifications to the panel in follow-on rounds.

Delphi: Round Three

The third round provides panelists with ratings and responses summarized from experts in the second round. Through the survey, the researcher asks the panel to revise their judgments in light of the new information or to provide additional justification while remaining outside of the consensus. This round gives panelists a chance to make further clarifications about information and judgments of target items. It is expected that “only a slight increase in the degree of consensus can be expected” in the third round (Hsu & Sanford, 2007).

Delphi: Round Four

The fourth round of the Delphi is often the final round (Hsu & Sanford, 2007). In this round, researchers provide the expert panel with a listing of remaining items, ratings, opinions, and items nearing consensus. This serves as a final opportunity for experts to revise their judgments. The degree of consensus required by the researcher dictates further iterations of the Delphi.

Delphi Data Analysis

In the Delphi, researchers use data analysis to discover expert opinions, determine the most important items, and to properly manage opinions. As discussed in the previous section, the number of Delphi rounds will depend on both the time allotted for conducting the Delphi as well as the degree of consensus, which researchers seek to employ. Knowing when to stop the Delphi is crucial since stopping too soon may provide non-meaningful results and stopping not soon enough may cause sample fatigue and tax manpower resources (Hasson, Keeney, & McKenna, 2000).

There is not a universally accepted level of consensus for the Delphi. The level used should be tailored to the expert panel and the research resources. The literature points a variety of consensus levels. McKenna suggests utilizing a 51% consensus among the panel (McKenna, 1994), Sumison points to a 70% consensus (Sumison, 1998), and Green *et al*. recommends 80% consensus (B. Green, Jones, Hughes, & Williams, 1999). Additionally, Ulschak recommends utilizing a seven-point rating scale for the Delphi (Ulschak, 1983), where other experts suggest a four-point Likert-type scale (P. Green, 1982). Some experts even suggest that a percentage measure is inaccurate and instead recommend measurement of the stability of the subjects' responses during each round (Scheibe, Skutsch, & Schofer, 1975).

Data analysis of a Delphi will involve the management of both qualitative and quantitative data (Hasson et al., 2000; Hsu & Sanford, 2007). The first round of the Delphi will primarily consist of qualitative data from the open-ended questions and researchers should manage the data through content analysis techniques. Researchers analyze the data collected at this stage by grouping similar items together to form a universal description. Some studies suggest omitting infrequently occurring items to keep the survey manageable; however, this can be seen as a slight against the primary Delphi principles (Hasson et al., 2000).

The major statistics used in the Delphi to represent information related to collective expert opinions are typically the central tendency statistics of *mean, median*, and *mode*, as well as levels of dispersion such as *standard deviation* and *inter-quartile range* (Hasson et al., 2000). Researchers strongly favor *median* and *mode* and many experts prefer a *median* score based on a Likert-type scale (Hill & Fowles, 1975; Hsu & Sanford, 2007). *Mode* can also be appropriate when clustering of the results around two or more points is apparent vice convergence at a single point (Hsu & Sanford, 2007).

Challenges and Drawbacks of the Delphi Method

Hsu & Sanford summarize several potential shortcoming and weaknesses of the Delphi Method (Hsu & Sanford, 2007) and Table 6 provides a summary of the Delphi Methods strengths and weaknesses. Overcoming low response rates is a constant challenge for Delphi practitioners. Poor response rate is magnified fourfold because a maximum of four surveys are sent to the same expert panel and if a certain portion of the experts discontinue their participation at some point in the process, the quality of the obtained information may be critically scrutinized. Therefore, researchers must seek to motivate respondents to assure their active involvement in the feedback.

Temporal resources also pose challenges to the Delphi. The technique can be time-consuming to implement for both the researcher and the expert panel. Often several days or weeks may pass between iterations, especially when conducting questionnaires through physical mailing. Web-based survey tools can potentially overcome this obstacle.

Studies have shown that the Delphi method can unintentionally be used to 'mold' or 'lead' an expert panel through false feedback (Cyphert & Gant, 1971; Dalkey & Helmer, 1963; Scheibe et al., 1975). Practitioners of the Delphi should exercise ethical investigation techniques and make certain to be cognizant to implement proper safeguards in dealing with this challenge (Hsu & Sanford, 2007).

Uneven distribution of knowledge and experience in the expert panels may also pose challenges. Subjects who may have less knowledge on a certain topic area may be unable to identify important statements identified by experts with a significantly higher amount of knowledge on the topic. Therefore, results of the Delphi could end up only identifying general statements about a topic instead of fully exposing true insights (Altschuld & Thomas, 1991).

Table 6. Strengths and Weaknesses of The Delphi Method

Delphi Strength	Delphi Weakness
Eliminates negative psychological barriers introduced in typical physical committee gatherings (Brown, 1968; Gordon, 2009).	Unintentionally leading an expert panel through false feedback. (Cyphert & Gant, 1971; Dalkey & Helmer, 1963; Scheibe et al., 1975)
Provides an equal voice to all committee participants. (Gordon, 2009)	Potential for uneven distribution of expert panel knowledge and experience. (Altschuld & Thomas, 1991)
Generation of unique ideas and perspectives. (Hsu & Sanford, 2007)	Potential for low response rates. (Hsu & Sanford, 2007)
Encourage true debate through anonymity. (Gordon, 2009)	Can be time consuming to properly implement. (Hsu & Sanford, 2007)
Decades of proven research and wide body of literature available. (Gordon, 2009; Hsu & Sanford, 2007)	Delphi methods do not (and are not intended to) produce statistically significant results. (Gordon, 2009)
Identify gaps in the body of knowledge on a particular topic. (Hsu & Sanford, 2007)	

Applicable Delphi Studies to the Research

There are a number of references to Delphi studies within the M&S industry. The *Simulation Modeling Handbook* specifies the use of the Delphi process as a capable decision-making tool for determining simulation project objectives. The book specifically recommends the use of electronic medium to conduct questionnaires (Chung, 2004).

The medical M&S community have used the Delphi method extensively. Palter *et al.* conducted a Delphi method to determine expert consensus on which virtual reality (VR) tasks are relevant to teaching laparoscopic surgery. They queried a panel of 19 experts in laparoscopic simulation (LapSim) and implemented a five level Likert scale to rate tasks to reach target consensus (80%). Through the study, Palter *et al*. were able to reach an 86.5% consensus for seven basic tasks in two rounds of the Delphi. The *median* statistic for expert scores was used as the benchmark for each task (Palter, Graafland, Schijven, & Grantcharov, 2012). Similarly, Zevin *et al.* implemented a Delphi to define a framework for a simulation-based surgical training curriculum. The expert panel consisted of twenty-four international general surgery experts who were able to achieve 90.1% consensus within a single round of the Delphi. Zevin *et al*. relied upon the web-based SurveyMonkey application to distribute Delphi questionnaires electronically (Zevin, Levy, Satava, & Grantcharov, 2012).

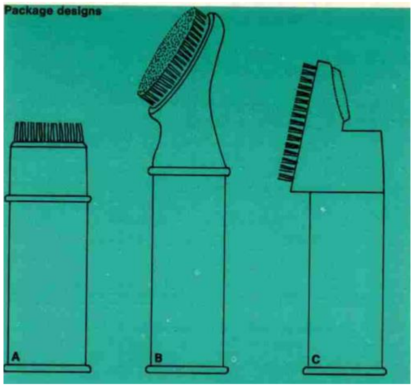
Within the military M&S community, Montijo *et al*. describe a Delphi implementation to reduce flight mishaps in the Air Force by determining the specific root causes of fighter and unmanned aerial system mishaps in order to develop improved behavioral-based simulation training objectives (Montijo et al., 2008). The panel reached a 78% consensus that most Sir Force flight mishaps occur due to channelized attention, task misprioritization, course of action, and cognitive task oversaturation errors. The *Department of Defense Risk, Issue, and Opportunity Management Guide for Defense Acquisition Programs* highlights the Delphi process as a risk identification methodology. The guide recommends the Delphi as systematic methodology to ensure "early and recurring communication between the user and acquisition communities involved in the development of JCIDS documents helps requirements leaders and acquisition leaders identify high risk requirements and inform potential technical risk-based trades" (Department of Defense, 2015). The DoD utilizes the Delphi method throughout the Simulation Based Acquisition (SBA) approach. (M. V. R. S. Johnson, McKeon, & Szante, 1998).

Conjoint Analysis

Conjoint Analysis is a powerful research tools used by marketers to decide which features a product should have and how companies should price the product. Conjoint analyses helps marketers and product research teams decide which of a product's or service's qualities are most important to a user or consumer (P. E. Green & Wind, 1975). Conjoint Analysis, or Conjoint Measurement as it used to be called, was first developed by Green and Rao in 1971 as a process to measure the joint effects of a set of independent variables on the ordering of a dependent variable (P. E. Green & Rao, 1971). Researchers applied the fields of mathematical psychology and psychometrics to develop the technique of conjoint analysis as a tool to sort out the relative importance of a product's attributes. It wasn't until 1975 that the marketing community took notice of conjoint analysis with Green and Wind's publication in *Harvard Business Review* detailing the application of early conjoint analysis to the consumer evaluation of various carpet cleaners (P. E. Green & Wind, 1975).

The goal of conjoint analysis is to determine the combination of product/service attributes that is the most influential on a consumers' purchasing choice or decision-making. In its most basic form, this conjoint analysis accomplishes this by showing respondents a controlled set of hypothetical products or services and analyzing their preferences between the products. The result is the determination of implicit valuation of the individual product elements called 'utilities' or 'part-worth'. This information is significantly useful in modifying current products/services and for designing new products to focused demographics (P. E. Green & Wind, 1975). Certain models based in conjoint analysis findings can also estimate the psychological trade-offs consumers can make when evaluating multiple products attributed together. Table 7 illustrates a simple example of a conjoint analysis survey and corresponding results.

Table 7. Experimental design of conjoint analysis for evaluation of a carpet cleaner



	Package design	Brand name	Price	Good Housekeeping seal?	Money-back guarantee?	Respondant's evaluation (rank number)
1	A	K2R	1.19	No	No	13
2	A	Glory	1.39	No	Yes	11
3	A	Bissell	1.59	Yes	No	17
4	B	K2R	1.39	Yes	Yes	2
5	B	Glory	1.59	No	No	14
6	B	Bissell	1.19	No	No	3
7	C	K2R	1.59	No	Yes	12
8	C	Glory	1.19	Yes	No	7
9	C	Bissell	1.39	No	No	9
10	A	K2R	1.59	Yes	No	18
11	A	Glory	1.19	No	Yes	8
12	A	Bissell	1.39	No	No	15
13	B	K2R	1.19	No	No	4
14	B	Glory	1.39	Yes	No	6
15	B	Bissell	1.59	No	Yes	5
16	C	K2R	1.39	No	No	10
17	C	Glory	1.59	No	No	16
18	C	Bissell	1.19	Yes	Yes	1*

*Highest Ranked

Source: Adapted from Green and Wind, 1975 (P. E. Green & Wind, 1975)

Over the years researchers have developed a number of conjoint analysis variations, each have their own benefits and weaknesses for a given scenario.

Traditional Full-Profile Conjoint Analysis

As its name suggests, Traditional Full-Profile Conjoint Analysis has been a mainstay of the conjoint analysis community for many decades. This variation typically requires respondents to *rank* or *rate* a series of choices (or cards), where each card displays a product concept consisting of multiple attributes (B. K. Orme, 2013). Table 7 is an example of a full-profile conjoint study. Traditional full-profile conjoint gets its name because respondents see the ‘full-profile’ of choices (all attributes at once). While each respondent only views one product per card, in the process of evaluating the full deck of choices, they sometimes compare the cards side-by-side or in sets. Since the respondents are provided with lots of information all at once, research has found that respondents use simplification strategies to key into few important attributes while ignoring others (B. K. Orme, 2010b). This may seem detrimental to the study, but additional research shows that consumers in the real world also simplify when making complex purchase decision, so simplification should not be considered detrimental to full-profile conjoint (Huber, 1997). Traditional full-profile conjoint can also measure interactions between attributes using composite attributes. As an example, a single four-level attribute by combining two attributes each with two levels.

Researchers most often use fractional factorial designs when constructing a full-profile conjoint study. Full-factorial designs are not practical since they require respondents to respond to an extremely large combination set of factor and attribute levels. Fractional-factorial designs show an efficient subset of the possible combinations and provide enough information to estimate utilities and main-effect interactions (P. E. Green & Rao, 1971; B. Orme, 2010).

Experts suggest that full-profile conjoint analysis is adequate for measuring up to six factors, although the number can vary from project to project based on additional factor criteria (B. K. Orme, 2010b). The limitation is that as the number of attributes increases, so too does the number of choices presented to respondents required to obtain statistically significant results (B. K. Orme, 2013). Additionally, traditional full-profile conjoint is limited in its measurement of composite attributes with more than two or three levels.

Several software firms have released software packages intended to facilitate the execution of Traditional Full-Profile Conjoint Analysis, including Sawtooth Software's Conjoint Value Analysis (CVA) tool and statistical packages by SAS and SPSS. Since traditional full-profile conjoint analysis can be thought of as essentially a multiple regression problem, Microsoft Excel can also be a powerful tool to conduct this version of conjoint analysis (B. Orme, 2010).

Adaptive Conjoint Analysis

Adaptive Conjoint Analysis (ACA) was a popular conjoint method throughout the 1990's. The primary advantage of ACA over traditional full profile conjoint is that it could easily measure more attributes than recommended for full-profile; ACA is capable of supporting up to two-dozen attributes without causing respondent fatigue (B. K. Orme, 2013). As its name suggest, ACA is able to handle increased attributes because it adapts sections of the questionnaires to respondent's pervious answers. Additionally, each section of the interview only presents one or several attributes at a time to further reduce respondent fatigue (B. K. Orme, 2013).

ACA combines stated evaluations of attributes and levels with conjoint pairwise comparisons. The first phase of an ACA study utilizes two-step self-explicated process. Respondents rank attribute levels and then assign an importance to each attribute. Research has shown that the assigning of importance to attributes is a challenge for respondents and have suggested that dropping this portion of an ACA survey will result in better prediction and discrimination between attributes with the caveat that researchers use hierarchical Bayes to estimate attribute utilities. ACA utilized the self-explicated question information to construct sets of tradeoff questions to the respondent (King, Hill, & Orme, 2004). Researchers ask respondents to indicate which product is preferred based on a simplified rating scale. The comparisons are tailored to each respondent to make sure that they are relevant, meaningful, and challenging (B. K. Orme, 2010b). Each of products shown is only a partial-profile and consists of only two to three attributes. Table 8 through Table 10 illustrates the three major steps of conducting an ACA survey.

Table 8. ACA Step 1: Rank attributes in terms of preference

Rank these brands from most to least preferred.
Brand 1
Brand 2
Brand 3

Source: Adapted from Orme (B. K. Orme, 2010a)

Table 9. ACA Step 2: Rate importance of attributes

If two brand products were acceptable in all other ways, how important would this difference be?
Brand 1 versus Brand 2
4 = extremely important 3 = very important 2 = somewhat important 1 = not important at all

Source: Adapted from Orme (B. K. Orme, 2010a)

Table 10. ACA Step 3: Pairs using graded rating scale

Which of these brands' products do you prefer?								
Brand 2 Attribute Level 1 Price 1								Brand 3 Attribute Level 2 Price 2
1	2	3	4	5	6	7	8	9
Strongly prefer left				Indifferent				Strongly prefer right

Source: Adapted from Orme (B. K. Orme, 2010a)

As discussed, ACA has several benefits of traditional conjoint analysis. ACA's self-explicated introduction phase, its adaptive survey nature, and its ratings-based tradeoff phase allows ACA to stabilize estimates of respondent's preferences for a greater number of attributes while using a smaller sample size than other conjoint methods (B. K. Orme, 2010b). Studies also find that pairwise comparisons such as ACA reflect real-world purchase behavior of consumers (Huber, 1997). In that regard, ACA is powerful for modeling complex high-price purchases where consumers focus on a number of product attributes before making a careful and thorough decision. It is suggested that smaller purchases involving less consumer interaction that are only described by few attributes be studied by another conjoint method (B. K. Orme, 2010b).

There are several limitation to utilizing ACA. The first is that researchers must use surveys to conduct ACA since the adaptive nature cannot be transferred to traditional paper and pencil questionnaires. Since ACA is a main-effects model, the utilities of each attribute are equal and does not account for interactions between attributes. This can be limiting when studying the effect of price sensitivity for product branding. Experts also find that ACA is further limited in pricing studies since the importance of price may become understated with the inclusion of many other attributes (B. K. Orme, 2013).

Practitioners of Conjoint Analysis still utilize today albeit not as much as previously. Leading conjoint analysis practitioners. Sawtooth Software, report that ACA accounted for roughly 5% of all conjoint studies in 2009(B. K. Orme, 2013).

Choice Based Conjoint

Choice Based Conjoint (CBC) gained popularity and has quickly become the most widely utilized method of conjoint analysis (B. K. Orme, 2013). CBC surveys closely mirrors the true purchasing behaviors of consumers between competitive products. Unlike previous forms of conjoint where researchers asked respondents to rank products concepts, CBC asks respondents to indicate product preference against a set or products shown to them. Table 11 illustrates an example of a CBC question. Much like the real world, respondents have the ability to decline a purchase through CBC by selecting a null (or none) option. Experts recommend that researchers show more rather than fewer product concepts per choice task (B. K. Orme, 2013).

Table 11. Example choice set for CBC analysis

If these were your only choices for a particular product, which would you choose?			
Brand 3 Attribute Level 1 $450 ○	Brand 2 Attribute Level 2 $425 ○	Brand 3 Attribute Level 3 $400 ○	None. If these were the only options, I'd defer my choice ○

Source: Adapted from Orme, 2010 (B. K. Orme, 2010a)

Since CBC shows sets of products against a full-profile, it encourages more respondent simplification than traditional full-profile conjoint. When compared to traditional full-profile conjoint and ACA, CBC shows emphasis on more important attributes and less emphasis on those that are less important (B. K. Orme, 2010b). Additionally, Huber indicates that choice based tasks are more immediate and concrete than abstract rating or ranking tasks (Huber, 1997).

There are also several variations of CBC commonly used by researchers. The below sections outline the three most common CBC variations.

Partial-Profile Choice Based Conjoint

Partial-profile conjoint was developed from the desire to maintain a choice based study while increasing the number of attributes than can be measured effectively (B. K. Orme, 2013). In partial=profile CBC, each choice contains a randomly rotated subset of the total number of attributes being studied. The major limitation of partial-profile conjoint is that because the data is spread very thin (i.e. each task contains may attribute omissions), the responses are less informative and require larger sample sizes to stabilize the results. Despite its limitations, previous proponents of ACA have shifted to partial-profile CBC in order to obtain greater market simulations; however, most experts and researchers still favor a full-profile conjoint method that displays all attributes within a single choice task (B. K. Orme, 2013).

Adaptive Choice Based Conjoint

Adaptive Choice Based Conjoint (ACBA) seeks to combine the best aspects of choice and adaptive conjoint techniques. ABCA offers respondents relevant products for consideration by patterning them after a preferred product that the respondent has first specified using a build-your-own (BYO) phase. Software then builds several dozen-product concepts using this BYO data for the respondent to consider. The considered products are then carried forward to a "choice tournament" to identify the best overall concept (B. K. Orme, 2013). This last phase is similar to a traditional CBC task. Figure 8 simplifies the overall process.

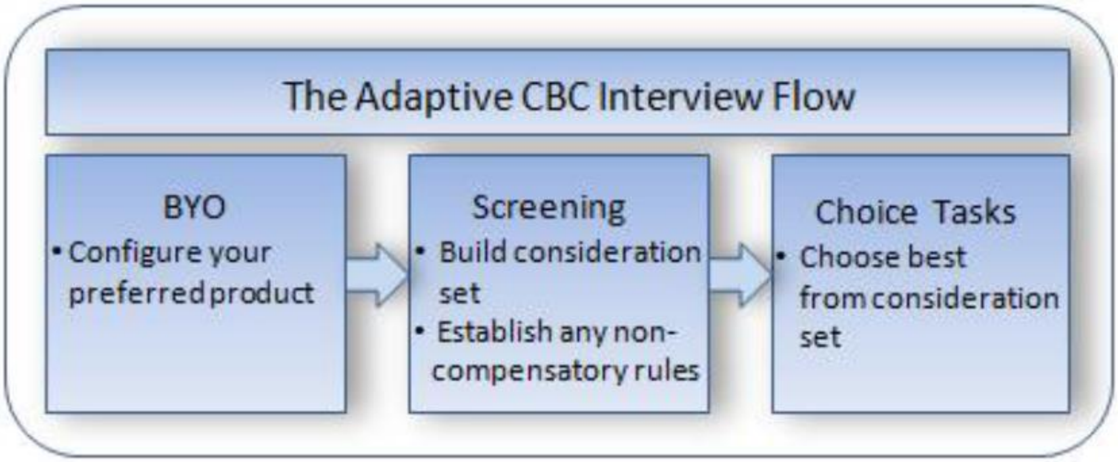


Figure 8. The ACBC Interview Process

Source: (B. K. Orme, 2013)

Even though a typical ACBC survey takes more time to complete over standard CBC surveys, research (and logic) suggests that ACBC surveys are more engaging, realistic, and relevant to respondents compared to traditional CBC tasks since the products presented are centered around their preferred product concept (B. K. Orme, 2010b). Additionally, since each individual captures more information, the sample size required for ACBC can be smaller than standard CBC studies.

ACBC is not a direct replacement for CBC. Experts argue that CBC is still recommended for studies containing four or fewer attributes, while ACBC is better suited for five or more attributes (B. K. Orme, 2010b, 2013).

Menu-Based Choice

Menu-Based Choice (MBC) closely mimic the behavior of a consumer's use of a menu to order products/service *a la carte*. Ultimately, the context of menu choice is different from CBC, which will lead to different utility predictions of consumer behavior. MBC studies are often more complex to design and execute over CBC studies and are recommended for more experienced researchers with significant background in DoE and market modeling (B. K. Orme, 2013).

Choosing a Conjoint Analysis Method

Orme offers several key decision areas that researchers should consider when selecting a conjoint analysis method. Table 12 provides a summary of these expert recommendations.

Table 12. Summary of when to use each method of Conjoint Analysis

Method / **Decision**	**Traditional Full-Profile Conjoint**	**Adaptive Conjoint Analysis**	**Choice Based Conjoint**	**Partial-Profile Choice Based Conjoint**	**Adaptive Choice Based Conjoint**	**Menu-Based Choice**
Number of attributes	≤6 (debated)	≥8	≤3		≥8 (w/ price)	
Mode of interviewing	Paper-and-pencil OR computer	Only computer	Paper-and-pencil OR computer		Only computer	Only computer
Sample size	≤30	small	>100		small	
Interview Time			<5 mins		>8 mins	
Pricing Research			Preferred method		Preferred method	
Menus						Preferred method

Source: Summarized from Orme, 2013 (B. K. Orme, 2013)

Data Analysis in Conjoint Studies

In conjoint analysis, customer indicate their preferences by either *ranking* or *rating* a set of choices. A customer's is the dependent variable and the product attributes are the independent variable. The part-worth utilities are indicated by the coefficients of a regression model and the R-square value gives an indication on how data fits the model –a high value of R-squared close to 1 would indicate the data fit the model well.

The relative importance value of a conjoint study explains how important an attribute is as affecting a consumers' preference for a gives product configuration. One can derive this value from the part-worth utilities for each factor. Two of the most frequently used methods are metric and non-metric analysis. Metric analysis is used if products are *rated*, since ratings are typically scaled at the interval level. Non-metric analysis is used if products are *ranked*, since ranked data is ordinal. The primary difference between metric and nonmetric data is how researchers transform the dependent variable. Metric conjoint analysis implements a linear transformation (commonly Ordinary Least Squares) and the rating data is unchanged. For nonmetric analysis, a monotone transformation (commonly MONANOVA) is conducted and the order of the rating/ranking is preserved but the data have been transformed to make the model fit better (Curry, 2001).

Iterative Design in Favor of Conjoint Analysis

Some experts argue that survey techniques such as conjoint analysis can stifle innovations and that companies who rely on these techniques risk falling victim to the "sameness trap" –the phenomenon of consumers expressing their wants in terms of other popular companies' products

(Ciotti, 2013). Steve Jobs, arguable one of the 21st Century's greatest innovators, echoes this sentiment in his know famous quote:

> *"It's really hard to design products by focus groups. A lot of times, people don't know what they want until you show it to them."*(Shepard, 1998)

While many entrepreneurial experts agree with this methodology, they acknowledge that this technique is best applied to extremely unconventional and circumstantial situations where the products that your company produces are so pivotal as to be creating or redefining their product categories and that you can back up your insights with a hugely expensive creative and iterative design process (Breillatt, 2009). An example of the failure of the Jobs methodology is when Ron Johnson, former VP of retail operations at Apple, became the CEO at J.C. Penny and sought to reform operations by ending product discounts without consulting customer feedback. Shortly after this implementation, J.C. Penny company sales dropped by double-digit percentages and stock plummeted over 40 percent (Ciotti, 2013).

Experts consent that customer can in-fact provide valuable insights for businesses, but surveys must be designed appropriately to capture proper demographic information. Ultimately, companies are at fault if the customer feedback is generic and carries limited utility (Ciotti, 2013).

Applicable Conjoint Analysis Studies to this Research

A thorough literature of Conjoint Analysis studies did not reveal any relevant research in the area of M&S or SNE; however, several useful pieces of literature exist that detail best practices and proper application of conjoint analysis through electronic mediums, which will benefit the research presented in this dissertation.

Sawtooth Software has a considerable market share of conjoint analysis software packages; however, the software and associated plug-ns and extensions come at a significant financial cost. Much like DoE there are several software packages other than Sawtooth Software that provide conjoint analysis packages. Kessels provides an in-depth tutorial for conducting a choice based conjoint analysis study using the *Choice Design* module of the popular JMP statistical software package. In this study, researchers analyzed four attributes of varying levels describing the packaging for laundry detergent, which result in 144 different packaging combinations. The researchers used the choice design module to construct five separate surveys each utilizing 12 tasks to compare three product combinations. This resulted in 487, 344 possible choice sets, but JMP was able to auto-selected the choice sets the provided the most information and resulted in the most precise estimates (Kessels, 2016). Furthermore, the research illustrates the use of JMP's additional statistical packages to interpret the results of a conjoint study. Alternatively, DecisionPro, Inc. provides an excellent tutorial on constructing a conjoint analysis study utilizing the commonly available Microsoft Excel software application (DecisionPro Inc., 2014).

Diener *et al*. discuss the implementation of conjoint analysis studies on mobile devices. Their research has shown that 64% of respondents prefer a smart-phone enabled mobile survey and 79% of these likened the preference due to the "on-the-go" capability (Diener, Narang, Shant, Chander, & Goyal, 2013). The researchers note that in order for mobile conjoint studies to be effective developers must design them with the mobile platforms in mind. They found that respondent experience is extremely poor if researchers proctor a PC-designed survey through a mobile device. Additionally, the research suggests that researchers can treat responses from mobile platforms and personal computers as since researchers can analyze homogenous data and together without risk.

Tang and Grenville tackle the issue of respondent engagement to conjoint analysis in their research. Admittedly, the authors indicate that conjoint surveys can be a monotonous task. Their research investigates how to make conjoint analysis more “fun” and engaging to users (Tang & Grenville, 2013). Their research suggests that three design considerations can significantly increase respondent engagement. First, researchers should augment the conjoint study with adaptive based choices; future choices will be a function of previous respondent choices. Second, the authors suggest improving the look and feel of the survey itself by swapping traditional survey elements for interactive controls (i.e. replacing radio buttons with a literal card-sorting interface to rank choices. In closing, the researchers encourage designers of conjoint studies to design conjoint exercises while considering the respondents’ point of view (Tang & Grenville, 2013).

Fractional factorial experimental design and Conjoint Studies

The work of Jones and Montgomery in fractional factorial experimental design has direct opportunities to constructing full-profile conjoint studies. In their seminal work, they present a wealth of knowledge regarding fractional factorial experimental screening designs utilizing 16 runs. In their research they note that fractional factorial (FF) designs are useful for factor screening because they efficiently identify dominant main effects, they contain a full factorial in fewer factors, and experimenters can easily add runs to resolve difficulties in interpretation (B.

Jones & Montgomery, 2010). In their work they provide the standard 16-run FF resolution III designs for six, seven, and eight factors. The authors note that these designs may not be the most efficient due to confounding of two-factor main effect interactions. This means that the experimenter cannot separate the effects unless they have adequate process knowledge since the effects cannot be separated without conducting additional experiments. The authors propose an alternative economical set of designs based on Hall orthogonal arrays that have no confounding. Their research suggests that these 'non-regular' designs offer a better chance of detecting significant two-factor interactions (B. Jones & Montgomery, 2010). Jones et al. apply this same concept to resolution III FF designs that utilize 9-14 factors for 16-runs. Again, the research indicates that the non-regular designs are capable of unambiguous estimation of main effects provided that there are only several two-factor interactions (Bradley Jones, Shinde, & Montgomery, 2015).

Graeco-Latin Square Designs and Conjoint Analysis

A Graeco-Latin square design is a design of experiment in which the experimental units are grouped in three different ways. It is obtained by superposing two Latin squares of the same size. If every Latin letter coincides exactly once with a Greek letter, the two Latin square designs are orthogonal ("Graeco-Latin Square Design," 2008). Two Latin squares are said to be orthogonal if the two squares when superimposed have the property that each pair of letters appears once. Together they form a Graeco-Latin square design. Latin square designs allow for two blocking factors and are used to simultaneously control (or eliminate) two sources of nuisance variability when running an experiment (The Pennsylvania State University, 2018).

Graeco-Latin Square designs are particularly useful in reducing the number of trials (combinations) for a complete design. This is especially important when conducting Conjoint Analysis among respondents which may not have enough time or interest in a longer-duration survey. The use of a reduced design often raises concerns that only a subset of all-interactions are observed, but experts agree that such an assumption is necessary when using orthogonal arrays (P. E. Green, 1974; Nielsen & Schmidt, 1990).

Quality Function Deployment

Quality Function Deployment (QFD) is a technique leveraged by both commercial and government organizations to design quality into a product versus inspecting for it. In QFD, quality is defined as meeting the customers' needs and providing a superior value (Crow, 2016). QFD is a highly structured approach to defining a customer's needs and requirements then translating them into a specific plan to produce a product to meet those needs. These customer needs and requirements are termed the "voice of the customer" and can be either stated or unstated based on the method of customer data collection. Researchers can capture the voice of the customer through a wide variety of means, to include, direct interviews, surveys, working groups, customer developed specifications, warranty data, and field reports, etc. The customer will often express needs in terms of "how" the need can be satisfied or "what" the need is, but practitioners of QFD should strive to ask "why" until the true root need is identified (Crow, 2016). Researchers can then use a series of planning matrices known as the "House of Quality" to summarize the voice of the customer. These matrices can then be translated to higher or lower levels to describe various aspects of products requirements and technical capabilities such as the "what's" or "how's". The House of Quality is not the final deliverable of a QFD study, but rather a means to an end (Crow, 2016). The House of Quality is a powerful communication tool, but the true power of QFD comes from its value to communication and decision-making at the organization level. QFD performs well in organizations because it is able to involve all functional departments within the organization in a synergistic product development process. Inter-departmental involvement leads to balanced consideration of requirements and can reveal knowledge that only single individuals/departments may know and may not be otherwise communicate throughout the rest of the organization.

QFD allows an organization's development team to focus on the true requirements of a product/service in an effort to minimize misrepresenting customer needs. To this effect, QFD is a powerful communication and quality-planning tool.

QFD involved a four-phased methodology across the process of product development: Product Planning, Assembly/Part Deployment, Process Planning, and Process/Quality Control. Figure 9 illustrates these four major phases. Each phase of QFD utilizes matrices to aid in the planning and communication of critical product and process planning and design of information.

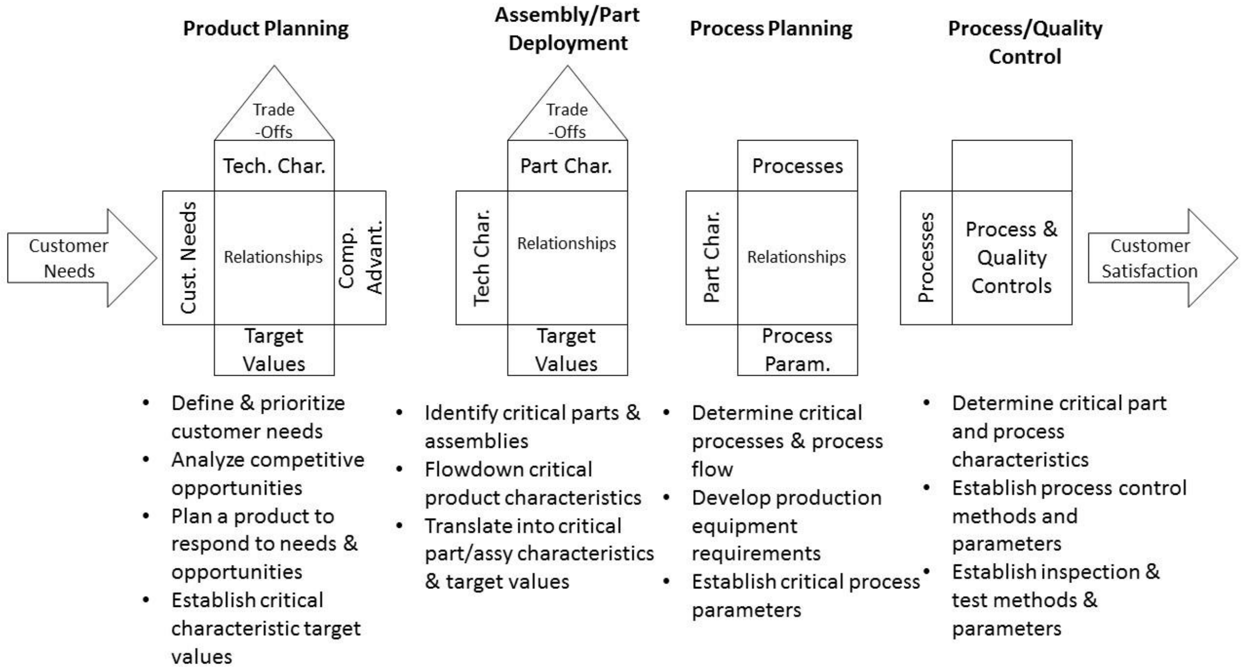


Figure 9. Four-Phased QFD Approach

Source: Adapted from Crow (Crow, 2016)

Product Planning

Once researchers obtain the customer needs and requirements, construction on the "house of quality" can begin. Figure 10 depicts an example House of Quality. This section provides a summary of preparing the House of Quality matrix for product planning.

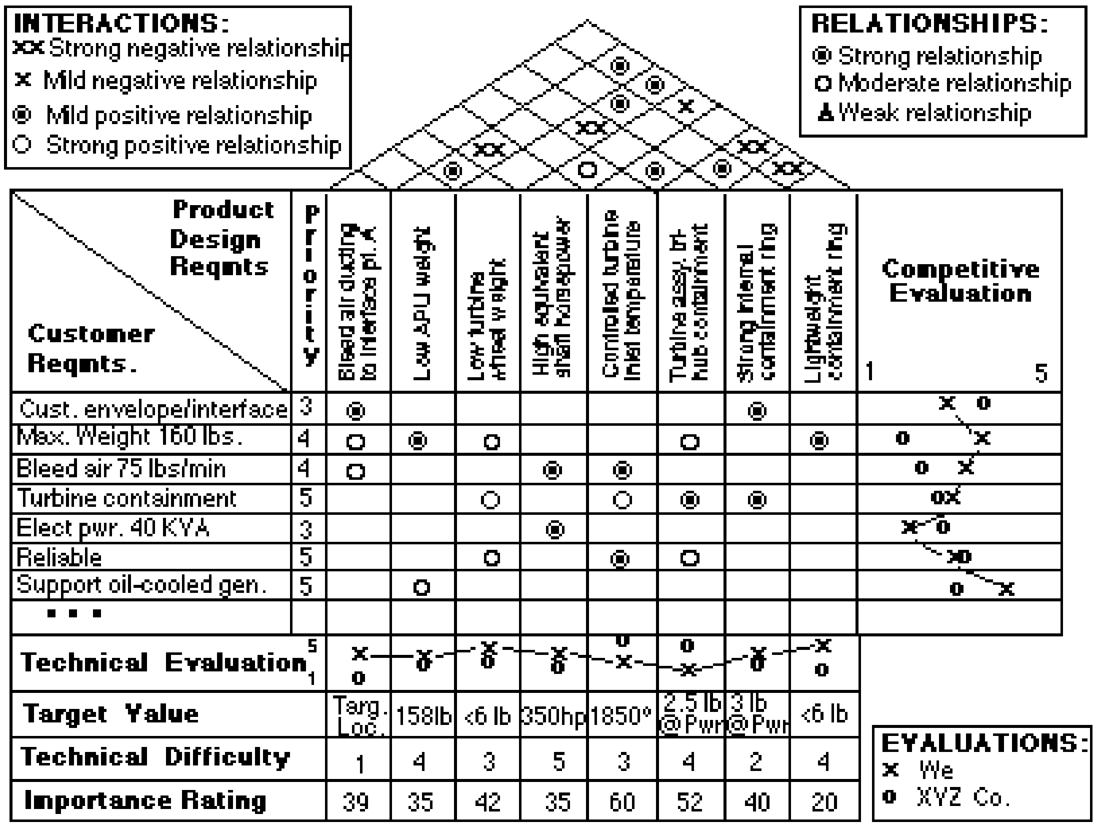


Figure 10. Example House of Quality

Source: (Crow, 2016)

The left side of the matrix (rows) depicts customer needs and requirements. They are organized by category with affinity diagrams and should be reflected by desired market segment. The needs/requirements in the matrix should not exceed between 20-30 items. If this occurs, users should decompose the matrix into smaller matrices to reduce the items. Users should then prioritize each customer need represented in the matrix through a ranking/rating.

Once researchers identify the customer needs, they list the design requirements along the top row of the matrix (columns). These design requirements should be ways of attaining customer needs and are under control of the product manufacturer/designer. These 'engineering characteristics' should be expressed in technical terms and must be measurable (Madu, 2006). It is common and expected for these design requirements to conflict and negatively influence one another. A well-design product or service is likely to involve tradeoffs (American Supplier Institute, 1989). If conflicts do not exist, it is probable that an error has occurred in the design. Conflicts should be resolved productively through the use of QFD else significant engineering changes will be required downstream (Madu, 2006). Additionally, these technical characteristics should be presented in a way as to not constrain designers by implying a specific technical solution (Crow, 2016).

After designers identify the customer needs and design requirements, they must assign relationships between them. Designers should use a standard symbol set to convey Strong, Medium, and Weak relationships. An example of this is seen in the 'Relationships' legend in Figure 10. It is advised that the use of Strong relationships should be used sparingly (Crow, 2016). Developers should also include three important rows at the bottom of the matrix: Target Value, Technical Difficulty, and Importance Rating. Target values are the specifications for a particular design requirement that could be achieved through engineering design (Madu, 2006). They should then assign a Difficulty Rating should to each technical characteristic; typically, a 1 to 5 point scale with five being most difficult/risky. Designers should avoid many difficult items since this will assuredly delay development and exceed budgets. Technical maturity, personnel qualifications, and business risk should be considered when assigning difficulty ratings (Crow, 2016). Lastly, the Importance Rating denotes the importance of each technical characteristic to the customer. A designer then calculates importance ratings by assigning weighting factors to each relationship symbol then multiplying the customer priority rating by the relationship-weighting factor in each cell of the matrix.

The next step in constructing the House of Quality is to add the ‘roof’. This matrix describes the correlation/interaction between different design requirements. Symbols should be used to indicate Strong or Medium, Positive or Negative relationships. This is depicted the ‘Interactions’ legend of Figure 10. Special attention should be paid to negative and strong negative interactions between requirements as this describes a conflict in trying to achieve both requirements jointly (Madu, 2006). In the event this occurs a trade-off should be made and the design requirement with the highest importance rating should be retained (Madu, 2006).

The final step in constructing the House of Quality is the addition of the Completive Evaluation column and Technical Evaluation row. Organizations can use these elements of the matrix for benchmarking the manufacturer’s products to that of competitors. In the Competitive Evaluation, the manufacturer is compared to the competitor on each of the customer requirements/needs identified by the customer (Madu, 2006). Likewise, the Technical Evaluation compares the manufacturer against the competitor for each of the design requirements. Organizations can accomplish this by obtaining competitive products and performing technical benchmarking. Additional data such as warranty and service claims, along with price, can be obtained for the technical evaluation (Crow, 2016).

Assembly/Part Deployment

As discussed, the House of Quality is not the final product of QFD. Once product planning is complete via House of Quality, designers can develop a more complete product specification. A concept selection matrix can be utilized select sources for new product concepts. Figure 11 depicts an example of this. The technical characteristics (criterion), normalized importance ratings, and target values are carried over from the House of Quality to this concept matrix. Designers then evaluate each potential product concept, using a symbolic rating, on how well they satisfy the criteria. If product concepts are weak in certain areas, but strong in others, the matrix can be used to identify areas where concepts can be "synthesized" together to form a new, stronger product (Crow, 2016).

Criteria	Importance Rating	Concept A	Concept B	Concept C
Low APU Weight	4	◉ 20	○ 12	○ 12
Low turbine wheel weight	4	◉ 20	○ 12	◉ 20
Controlled turbine inlet temperature	6	○ 18	◉ 30	△ 6
Acceptable turbine assembly life	5	○ 15	◉ 25	○ 15
Turbine assy tri-hub containment	5	◉ 25	◉ 25	○ 15
High equivalent shaft horsepower	4	△ 4	◉ 20	◉ 20
Strong internal containment ring	4	○ 12	◉ 20	○ 12
Total		114	144	100

Figure 11. Example Concept Selection Matrix

Source: (Crow, 2016)

Based on the concept selection matrix (or other concept evaluation methods), a product concept is selected. A process diagram can then constructed to identify critical subsystems, modules, and parts of the product concept. Designers use this to prepare a part deployment matrix. They can then construct a part deployment matrix very similar to the product-planning matrix (House of Quality). Product requirements now become the rows of the matrix and critical part subsystems, assemblies, and characteristics become the columns. Relationships, importance ratings, and designers again calculate the target values for each critical subsystem. Figure 12 depicts an example part deployment matrix.

Product Design Requirements	Priority	Critical Part Char. / Targ. Value	Turbine wheel					Combustor			
			Balanced	Surface finish	Backface geometry	Grain refinement	Airfoil geo. & thickness	Material	Liner pattern factor	Nozzle throat area	...
Low APU Weight	4	158 lb					◉	○			
Low turbine wheel weight	5	<8 lb			○	◉	○				
Controlled turbine inlet temperature	6	1850° max				○			◉	◉	
Acceptable turbine assembly life	3	3,000 hrs	⊖		⊖	⊖	⊖	⊖	⊖		
Turbine assy tri-hub containment	4	2.5 lb @ PW		◉	◉	◉		○		○	
High equivalent shaft horsepower	4	390 hp		◉		○				◉	
Importance Rating			15	40	35	90	50	39			

◉ Strong relationship
○ Mild relationship

Figure 12. Example Part Deployment Matrix

Source: (Crow, 2016)

Process Planning

Process design continues the same process as assembly and product design, except at the manufacturing process level. In order to evaluate various manufacturing approaches a concept selection matrix is constructed. This informs the development of the process-planning matrix (Figure 13). Again, designers use the higher-level matrix to populate the rows, in this case the critical subsystems and the columns are populated with important process and tooling requirements. It's recommended that engineering and manufacturing teams work closely during this stage to identify trade-offs to achieve mutual goals based on customer needs (Crow, 2016).

Critical Part Characteristics / Critical Process Steps	Turbine wheel					Rating	Part Control Parameters
	Balanced	Surface finish	Backface geometry	Grain refinement	Airfoil geo. & thickness		
Priority	2	4	4	9	5		
Mold preparation	○	⊙			⊙	51	- surface finish - Airfoil geometry
Hot isostatic pressure casting	○	⊙		⊙	⊙	96	- surface finish - inclusions, cracks, porosity - Blade tip fill
Mass center balancing	●					10	- Machine centers
Turbine tip OD & shroud line contour machining		○				12	- Outer diameter - Profile geometry - surface finish
Low stress grind - backface	○	⊙	⊙			46	- Backface geometry - surface finish

Other Data:
- Equipment
- Location
- Tooling

Figure 13. Example Process Planning Matrix

Source: (Crow, 2016)

Process/Quality Control

A final and often overlooked step of QFD is the generation of a quality control matrix (Figure 14). This matrix supports more detailed planning related to process quality control, setup, equipment maintenance, and testing (Crow, 2016). The previous process-planning matrix is used as the basis for planning the specific quality control steps. The result of this planning is that the manufacturing process directly focus on the critical processes and characteristics that will have a significant effect on developing a product to meet the "voice of the customer".

Critical Process Steps	Process Control Parameters	Control Points	Control Method	Sample Size & Freq.	Check Method
Hot isostatic presure casting	Mat'l temp. Mold temp. Remelt %	Mat'l prop. Heat treat FPI	Cert.	100%	N/A
Mass center balancing	Balance mach. calibration Speed	Detailed balance	Cert.	100%	N/A
Turbine tip OD & shroud line contour machining	Set-up Speeds & feeds Tool wear	Dim. insp. Surface finish	X bar & R chart	4 pieces/ lot	Elect. gage Check fixture Visual
Low stress grind - backface	Speeds & feeds Diamond dressed wheel	Dim. insp. Surface finish	N/A	100%	CMM Visual
Florescent penetrant insp. & proof spin	Speed	Cracks Inclusions O.D.	N/A	100%	Visual Spin test

Figure 14. Example Quality Control Planning Matrix

Source: (Crow, 2016)

Limitations of QFD

QFD is not a perfect solution for all applications and has certain limitations that would-be practitioners should be aware of. For example, QFD can be cumbersome if all relational matrixes combine into a single deployment, the size of each of the combined relational matrixes would be very large and make it difficult to draw accurate conclusions (Wolniak, 2018). QFD is a predominately qualitative method. Due to the ambiguity and subjectivity in the voice of the customer, many of the answers that customers give are difficult to categorize as demands. It can also pose challenges in making connections between customer demands and technical priorities (Wolniak, 2018).

Example QFD Deployments

Akao and Mazur provide a summary of worldwide past, present, and future implementations of QFD (Akao, 1997). Many of these implementations are within automotive companies across the globe. The Chrysler Corporation was among the first to embrace QFD in North America. Czupinski and Kerska outline Chrysler's first major QFD project, the Chrysler LH powertrain from 1988. Chrysler formed a cross functional team representing Brand Management, Design Office, Program Management, Engineering, Process Engineering and Finance, to identify the key customer attributes for a mid-size sedan (Czupinski & Kerska, 1992). After extensive marketing research, the team developed an overall priority for each of the powertrains most important requirements. This enabled them to identify four critical subsystems of the LH powertrain as important study areas. The team also conducted a competitive assessment against five competitors' vehicles. The Chrysler team used this information to establish design requirements that more realistically correlated to customer expectations of powertrain performance. Design requirements were then selected from the House of Quality matrix and passed to the Design Deployment matrix and ultimately to the Process Planning matrix to determine the manufacturing operations most critical to creating the desired part characteristics. Chrysler measured the success of their QFD deployment based on positive comments and favorable ratings of these vehicles during executive and media evaluations (Czupinski & Kerska, 1992). Lockamy and Khurna detail additional lessons learned through Chrysler's adoption of QFD in product design (Lockamy & Khurana, 1995).

Government agencies can also successfully implement QFD. The QFD Institute provides of overview of QFD adoption by NATO. After Operation Desert Storm, NATO saw a need to identify infrastructure and support factors that inhibit the mobility of combat aircraft and to formulate approaches to alleviate the effects of infrastructure and support requirements on aircraft mobility (QFD Insitute, n.d.). NATO formed a team comprised of a mix of aviators, logisticians, maintainers, and industry analysis from multiple nations, as well as a retired US Marine Corps colonel acting as the director. While the study was ultimately successful and received high praise from NATO, it suffered from several implantation issues. Overly complex matrices, insufficient resolution in QFD impact ratings, and inadequate front-end analysis of spoken and unspoken Voice of the Customer were causes of these initial issues (QFD Insitute, n.d.). Yamamoto *et al*. detail another application of QFD by a government entity. In this study, they examine the City of Sapporo, Japans application of QFD to obtain customer evaluation and perspective of government public services such as road maintenance. Their study found that government could easily apply QFD to government services that are visible and tangible, but it was difficult to apply to less tangible services, such as law and policy (Yamamoto, Hara, Kishi, & Satoh, 2005). The study validated that QFD is useful for product and service design, but it should not be used to form consensus among customers.

Additional Relevant QFD Literature

QFD and Conjoint Analysis are two very different methods for understanding the voice of the customer. Product designers will typically choose one of these methods to conduct their market research. Pullman *et al*. provide a detailed strength and weakness comparison of the two methods by applying each of them to the design of a widget. Through their research, they found that Conjoint Analysis was easier to compare the most preferred product features with profit maximizing features to maximize profit. They found that QFD was better able to highlight engineering characteristics and design features that had both positive and negative characteristics thus allowing a better analysis of tradeoffs. Ultimately, the study team recommended that QFD and Conjoint Analysis are not competing techniques but are complimentary in the information they reveal. The study proposes a hybrid QFD/Conjoint Analysis approach in which a conjoint analysis study would reveal the "voice of the customer" as input for a subsequent QFD study. Alternatively, QFD could be used to first screen the problem down to a smaller number of features and a conjoint study could be utilized to refine levels and improve predictions (Pullman, Moore, & Wardell, 2002).

In the realm of Modeling and Simulation, Riddle and Olejniczak successfully implemented a QFD study in their pursuit of user requirements for the development of the Advanced Technology Crew Station (ATCS), a modern tactical aircraft crew station simulator. They were able to harness QFD to provide detailed traceability between design requirements and mission objectives, threat assessments, etc. (Riddle & Olejniczak, 2000). The study team used multiple QFD iterations to abstract various levels of system design for evaluation.

CHAPTER THREE: PROPOSED METHEDOLOGY

Using Conjoint Analysis of end user assessments of SNE Visual Aesthetic Quality (VAQ), this research explores the Delphi Method as a technique for generating SNE VAQ factor forecasts and preferences for supplementing existing VV&A techniques in the conceptualization phase of a SNE systems engineering life cycle. First, given an interoperable simulation case-study setting (U.S. Army ITE) consisting of distributed, dissimilar simulations, the research uses an extensive Delphi analysis to poll multiple populations of SNE experts, including developers, users, and managers. The Delphi analysis intends to separate primary factors that affect correlation of virtual SNE visual aesthetic quality across the dissimilar simulations. The Delphi scope does not include user-case fidelity difference that may impact training. Once the Delphi panel reaches consensus on the primary SNE VAQ, SNE's will be generated and subsequently evaluated by end users. The importance and significance of the forecasted factors visa via end user assessments will be explored though a Traditional Full-Profile Conjoint Analysis developed through a Graeco-Latin Square screening design, while also identifying any factor interactions implementation.

Based on the resulting analyses, the research will also recommend identified statistically significant visual aesthetic quality factors be possibly implemented as priorities upstream in the SNE production & VV&A process. Additionally, the research hopes to identity foundations that enables follow-on research to implement changes in SNE VV&A techniques and inform Quality Function Deployment (QFD) processes for SNE generation activities.

To these ends, the research proposes to explore the following research questions posed in Chapter 1:

1. Does the Delphi technique produce VAQ factors and factor preference forecasts during the concept phase of SNE that are consistent with end User generated assessments?
2. Does Conjoint Analysis improve the understanding of the significance and power of identified VAQ factors?
3. What are the set of primary factors, priorities, and interactions that most affect VAQ and utility of synthetic natural environments for an interoperable training use case?
4. Can the information gathered from the Delphi technique and Conjoint Analysis supplement existing VV&A processes to create a new SNE VV&A paradigm?

Hypotheses

The set of hypotheses corresponding to the research questions posed by this research are as follows:

1. For a given synthetic natural environment representation across dissimilar simulators within a multi-domain simulation exercise, there exists a correlated set of expertly accepted and user validated primary VAQ factors that affect overall realism and training utility in the virtual domain.

2. Conjoint analysis will improve the understanding of the significance and power of identified factors and preferences

3. a. A Delphi study using a panel of experts will forecast the same VAQ factor considerations as Conjoint Analysis of end user assessments.

   b. Data mining of historical SNE issue reports will identify the same level of importance of VAQ factors as users reviewing SNE representations through a Conjoint Analysis.
   c. Data mining of historical SNE issue reports will identify the same level of importance of VAQ factors Delphi panel expert forecasts.
4. Quality Function Deployment (QFD) can be utilized to abstract the correlated set of expertly accepted and user validated primary SNE VAQ factors into a series of SNE generation process improvements to influence a new SNE VV&A paradigm.

Follow on research may extend hypotheses 2a, 2b, 3, and 4 to other applications. Variability of SNE applications may limit extensibility of outcomes from hypothesis 1. Due to resource limitations, this research will not attempt to address hypothesis 4 by deploying a QFD study for the development of a new VV&A paradigm but hopes to provide the groundwork for future research on this topic.

To test the remaining three hypotheses, the research will follow a five-phased methodology. Figure 15 provides a high-level depiction of the methodology of this research. The first phase of the research seeks to identify through a Delphi study of industry experts and users the primary VAQ factors associated with the correlation of a SNE and its rendition across dissimilar simulators. The second primary phase of this research will focus on an optimized fractional factorial screening design proposed by Jones and Montgomery (2010). The third primary phase will utilize the expert-validated factors from phase one and the optimized fractional factorial design to develop a set virtual SNE representations within a modern visual game engine software, each slightly different based upon variations in the identified VAQ factors and factor levels. The fourth major phase will then utilize the set of SNE representations as the choice tasks for a traditional full-profile conjoint analysis study. The fifth and final phase will perform data analysis of the structured Delphi and Conjoint Analysis results and conclude the research by making recommendations for implementing the results of the research into SNE generation processes and providing recommended parameters for each VAQ factor identified. This dissertation will also present a path forward for passing the results of the research into a Quality Function Deployment analysis as the "Voice of the Customers" (e.g. end user) to better understand critical processes and design consideration required to achieving a high-quality SNE representation – thus influencing a new SNE VV&A paradigm.

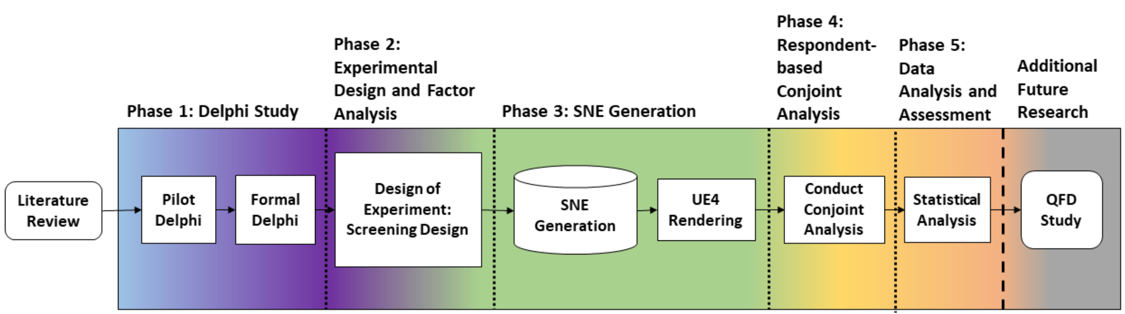


Figure 15. High-Level Research Design

SNE Quality Evaluation Use-Case

This research seeks to extend the earlier complex visual scene study experiments performed by Purdy & Goldiez (1995) by focusing on the Case 3 and 4 definitions of SNE "interoperability"—'Non-uniform levels of simulation fidelity' and 'Differences in the virtual environment', respectively. Therefore, this research will instruct participants in Phase 1 to identify the purely visual aesthetic quality (VAQ) factors for consideration when seeking correlation and rendition of a SNE across dissimilar simulators. During phase four, this research will ask participants to rank order a set of SNE, each representing a unique rendition across a simulator, based on their visual aesthetic qualities -- the visually pleasing appearance of the SNE for utility across a distributed simulation consisting of dissimilar simulators.

Since this research focuses on Synthetic Environment Core (SE Core) terrain as an interoperable SNE use-case and SE Core is the SNE provider of choice for the U.S. Army, this research will reference the suite of confederated virtual simulation systems of the U.S. Army Integrated Training Environment (ITE) as example, but not mandatory, interoperable simulation platforms for VAQ consideration. Table 13 identifies these platforms and provides a summary for each.

As with the study by Purdy & Goldiez (1995), this research does not directly address issues that are specific to any given training task. The research assumes that to implement a specific training task, which involves more than one simulator, a certain level of fidelity, must be present in all simulators to the 'lowest common denominator'. Purdy & Goldiez also specify that 100% interoperability is extremely unlikely in the presence of heterogeneous hardware and software systems, but practical interoperability can be achieved when all the relevant parameters are identified and acceptable quantitative difference limits are defined and adhered to (Purdy & Goldiez, 1995)

Table 13. Summary of the U.S. Army ITE Virtual Training Aids, Devices, Simulations, and Simulators

Virtual Simulation Platform	Virtual Simulation Use-Case	Description	Link for Additional Information
Aviation Combined Arms Tactical trainer (AVCATT)	Aviation (Rotary)	AVCATT is a mobile, transportable, multi-station virtual simulation device that supports unit collective and combined arms training for helicopter aircrews.	http://www.peostri.army.mil/PRODUCTS/AVCATT/
Close Combat Tactical trainer (CCTT)	Ground (Tracked)	The CCTT system consists of computer-driven, manned module simulators replicating the vehicles found in close combat units such as the M1 Abrams Tank, the M2 Bradley Fighting Vehicle (BFV).	http://www.peostri.army.mil/PRODUCTS/CCTT/
Dismounted Soldier Training System (DSTS)	Dismounted (Soldier, Squad)	DSTS is a virtual trainer focused on the individual Soldier and squad-level training that combine gaming technology in a virtual, 360-degree training environment using untethered weapons.	http://www.peostri.army.mil/PRODUCTS/CCTT/
Games for Training (GFT) / Virtual Battle-Space (VBS)	Dismounted (Soldier, Squad, Platoon)	VBS is a 3-D, first-person, games-for-training platform that provides realistic, semi-immersive environments, dynamic terrain areas, hundreds of simulated military and civilian entities.	http://www.peostri.army.mil/PRODUCTS/USAGFTP/
Live, Virtual, Constructive Integrating Architecture (LVC-IA)	Interoperable (LVCG)	The Live, Virtual, Constructive-Integrating Architecture (LVC-IA) is a system of systems providing a net-centric linkage that collects, retrieves and exchanges data among existing Training Aids, Devices, Simulations, and Simulators (TADSS) and both joint and Army Mission Command Systems.	http://www.peostri.army.mil/PRODUCTS/LVCIA/
Multiple Unified Simulation Environment (MUSE)	Aviation (Fixed Wing - UAS)	MUSE provides simulated video feeds for various intelligence-gathering platforms. The MUSE program is the primary UAS training and simulation system used in the Department of Defense for command- and staff-level joint services	http://www.metavr.com/casestudies/uas_sim.html

Virtual Simulation Platform	Virtual Simulation Use-Case	Description	Link for Additional Information
Reconfigurable Vehicle Tactical Trainer (RVTT)	Ground (Wheeled)	RVTT is a system within CCTT that includes the Reconfigurable Vehicle Simulator (RVS), which was originally designed to train the Armored Reconnaissance Platoons and Combat Service Support units supporting the Heavy Brigade Combat Team (HBCT).	http://www.peostri.army.mil/PRODUCTS/CCTT/
Reconfigurable Virtual Collective Trainer (RVCT)	Multi-Domain	The Reconfigurable Virtual Collective Trainer (RVCT) includes aviation platforms (RVCT-A), ground platforms (RVCT-G), dismounted infantry collective maneuver training, collective gunnery training, and mission rehearsal capability. The RVCT is a mobile, transportable, modular, and scalable training capability with the minimum hardware necessary to represent form, fit, and function for the user to execute collective tasks.	https://www.peostri.army.mil/reconfigurable-virtual-collective-trainer-rvct
Non-rated Crew Member Manned Module (NCM3)	Aviation (Rotary)	The NCM3 supports the training of non-rated crew members in crew coordination, flight, aerial gunnery, hoist and slingload-related tasks.	http://www.peostri.army.mil/PRODUCTS/AVCATT/
Synthetic Environment Core (SE Core)	Interoperable (LVCG)	The ultimate objective of SNE Core is to facilitate a common virtual training environment to enhance the training and mission rehearsal capabilities for our Soldiers.	http://www.peostri.army.mil/PM-ITE/SECore.jsp
Common Driver Trainer (CDT)	Ground (Wheeled)	The CDT consists of a simulated vehicle cab, instructor/operator station, After Action Review (AAR) station, visual system, six-degrees-of-freedom motion system and a computational system.	http://www.peostri.army.mil/PRODUCTS/CDT/

Research Scope

This research does not seek to identify all possible interoperability issues nor does this research address the impact of interoperability to specific training tasks and virtual simulation use-cases. Interoperability issues addressed in this research are purely visual and primarily associated with the image generator (IG) component of a simulator. Regarding the Purdy & Goldiez interoperability issue cases, this research does not address Case 1 or Case 2. Case 3 issues are limited to the SNE representation differences across simulators. Case 4 issues relate to the simulator IG and visual scene and therefore account for the bulk of the issues examined in this research. Outside of providing examples for participant consideration, this research makes no distinction between simulator types or weapon platform.

Phase One: Conduct a Delphi Study of SNE Quality Factors

The purpose of this Delphi study will be to query three populations of SNE experts on factors affecting SNE VAQ: Developers, Users/Operators, and Managers. This research will utilize the Internet as a medium construct and distribute the Delphi surveys. A web-based Delphi technique is a cost-effective way to carry out research. This enables participants to easily respond from diverse geographical locations, is time efficient, allows direct import into data analysis software, enables a quick turnaround time between rounds, and improves data quality (Gill et al., 2013). Figure 16 provides a visual overview of the Delphi research phase.

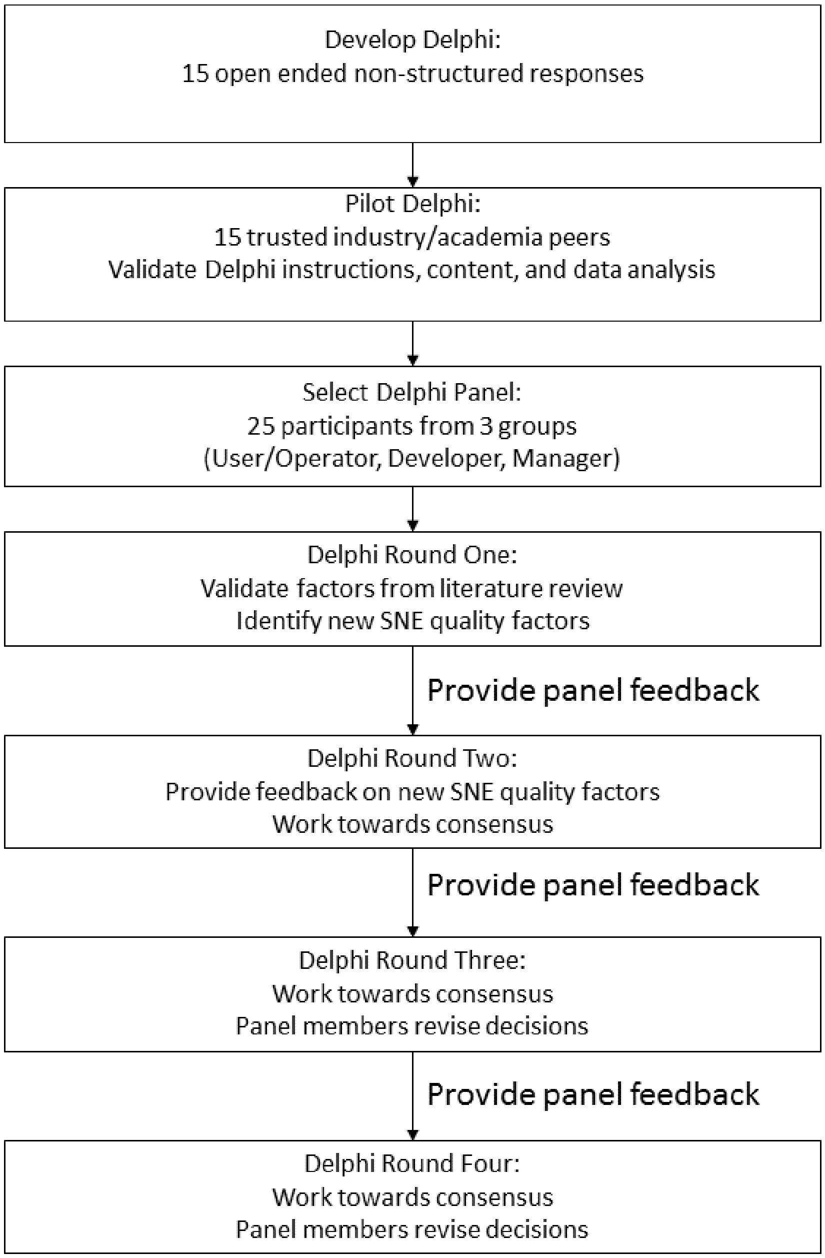


Figure 16. Delphi overview research methodology

Delphi Survey Design

This research will implement the web-based survey tool Survey Monkey™ to design and administer the Delphi surveys associated with this research. Survey Monkey's user-friendly interface, extensive web-browser and operating system compatibility, and its ability to export standard data formats for integration into third-party statistical analysis software tools (JMP, XLSTAT, etc.) were key factors in its selection as a research tool. Gill *et al*. detail multiple studies which have leveraged Survey Monkey as a proven tool for administering successful Delphi studies (Gill et al., 2013). While Survey Monkey offers a free-to-use version of their software, this research will require an upgrade to Survey Monkey's 'Gold' plan in order to take advantage of data export and additional survey customization features.

Unless a comprehensive and extensive literature review has been conducted, it is traditional for the Delphi to begin with an open-ended questionnaire (Hsu & Sanford, 2007). Since SNE development can be a very subjective process consisting of many factors that could potentially impact quality and a thorough literature review only reveals the visual quality factors identified by Purdy & Goldiez (Purdy & Goldiez, 1995), an open-ended questionnaire will be implemented for the first round of the Delphi. The round one survey will contain fifteen blank fields for experts to provide their own open-ended input and justifications to SNE visual quality factors for an interoperable use-case. The literature projects that this Delphi study will require four rounds to achieve the desired level of consensus between experts; however, termination can occur if the study reaches consensus earlier.

Delphi Data Analysis

Expert practitioners recommend data analysis decision rules be determined prior to Delphi survey being developed (Hsu & Sanford, 2007). This will reveal critical details about organizing and assembling insights from the Delphi panel.

The first decision is to establish the level of consensus for the Delphi study. When a certain percentage of expert panel votes falls within a predefined range, consensus will occur. A review of available research suggests that having 80 percent of an expert panel votes fall within two categories on a seven-point Likert scale is sufficient for achieving consensus on a Delphi (Ulschak, 1983). A seven-point scale allows for greater data granularity to support analysis of the responses using continuous data analysis (Gill et al., 2013).

As discussed in the previous section, a defining characteristic of the Delphi method is the process of providing feedback of results to the expert panel to allow them to refine their rankings in order arrive at consensus. The primary statistics used to provide feedback are the measures of central tendency (mean, median, and mode) along with the levels of dispersion (standard deviation and inter-quartile range). The literature favors the *median* and *mode* to reflect the convergence of expert opinion (Hsu & Sanford, 2007), therefore *median* will be the primary measures utilized due to resistance against outliers. Experts warn that *mean* is not an appropriate measure based on the potential for extreme expert outlier opinions as well as the case where scales used in the Delphi are not delineated at equal intervals (Hsu & Sanford, 2007). The researcher will provide the *median* statistic along with graphical representations of respondent distributions to the expert panels as feedback.

Delphi Survey Development

As discussed, this research will use Survey Monkey to construct the Delphi surveys. Survey Monkey offers a comprehensive graphical user interface to design questionnaires and surveys along with additional user-defined enhancements such as image upload, HTML editing, and word processing. The first page of this Delphi study will introduce the research and provide ethics information and considerations. Page two of the Delphi survey will provide specific instructions to the panel members, including information on the case study approach, the specific interoperable problem cases identified by Purdy & Goldiez (1995), and all applicable concept definitions.

The third page of the Delphi survey will contain a demographics questionnaire for two primary purposes. The first is to capture contact information in the form of an email address necessary to provide the required feedback responses back to the expert, the core of the Delphi method. The other purpose of the demographics survey is to help further discriminate the expert populations into virtual simulation domain-specific backgrounds (i.e. aviation, ground, dismounted, etc.). During the Delphi panel selection phase described below, the researcher will strive to obtain an equal distribution of experts from each primary virtual simulation domain. The experts will be required to provide their primary virtual simulation domain experience as well as provide their level of familiarity with other virtual simulator domains in order to capture their specific SNE use-case demographic and help the researcher identify research demographic gaps. The research will implement a 5-level Likert scale specified in Table 14 to investigate expert familiarity across virtual simulation platforms/use-cases.

Table 14. Level of Familiarity Likert type scale rating for Delphi panel domain demographic analysis

Level of Familiarity	
1	Not at all familiar
2	Slightly familiar
3	Somewhat familiar
4	Moderately familiar
5	Extremely familiar

Since this research specifies SNE Core terrain as the interoperable SNE use case, the confederated virtual simulation systems of the U.S. Army Integrated Training Environment (ITE), for which SNE Core is the primary terrain provider, will be used to help experts relate a domain-specific use-case to a real-world simulation platform. Table 15 provides this cross-relational breakdown of SNE use-case to physical U.S. Army ITE simulation platform. This demographic information will provide valuable insight into the forecast of SNE VAQ factors between each primary population group, individual use-cases, and the interoperable SNE as a whole.

Table 15. SNE Domain Use-Case relation to U.S. Army Virtual Simulation Platforms

SNE Training Use-Case	**Example U.S. Army ITE System**
Aviation-Rotary	Aviation Combined Arms Tactical Trainer (AVCATT), Non-rated Crew Member Manned Module (NCM3)
Aviation-Fixed Wing	Multiple Unified Simulation Environment (MUSE)
Ground-Tracked (Armor)	Close Combat Tactical Trainer (CCTT)
Ground-Wheeled	Reconfigurable Vehicle Tactical Trainer (RVTT), Common Driver Trainer (CDT)
Dismounted Soldier/Squad/Platoon	Games for Training (GFT) / Virtual Battlespace (VBS) / Dismounted Soldier Training System (DSTS)
Interoperable/Distributed Interactive Simulation (DIS)	Live, Virtual, Constructive Integrating Architecture (LVC-IA) / Veritas Stealth Viewer
Other (Please Specify)	

Following the introduction pages and the demographics questionnaire, the first round Delphi survey will present the experts with fifteen open-ended response fields to solicit input and feedback statements on the visual aesthetic quality factors affecting correlation of the virtual SNE and its rendition across dissimilar simulators. As discussed earlier, a common issue of the Delphi method is the leading of expert consensus towards a pre-determined direction. The research will strive to maintain a careful balance between requesting purely abstract expert responses on SNE interoperability and leading experts to a pre-determined consensus. To do this, the instructions will ask experts to consider the virtual simulation domains and example platforms identified in Table 13 when providing their comments on visual aesthetic quality factors affecting interoperability of a virtual SNE across dissimilar simulators. Following each open-ended response on the VAQ factor for consideration, there will be an additional textbox where experts can provide their justifications, recommended parameters, and additional comments for each factor. Rules will be setup in the Survey Monkey tool to require experts to provide at least ten factor considerations and justifications prior to completing the survey; this will guarantee that there will be at least eight new factors for the second round of the Delphi – the target ending VAQ factor count.

There is a risk to this research that by providing the option of open-ended responses to the three populations in round one, the Delphi may diverge towards three or more disparate consensuses instead of the goal of a single unified expert consensus. In the unlikely event of this occurring, the research will mitigate this risk by only carrying forward the results of the User expert population, since this is the target 'customer' population for the phase four conjoint analysis.

There will be no open-ended responses for rounds two, three, and four of the Delphi surveys. The research will carefully analyze and synthesize statements from the open-ended qualitative responses from the first-round survey; the research will consolidate similar responses where appropriate. In these subsequent rounds, experts will rate their agreement to each new statement about SNE quality factors using a seven-point Likert-type scale rating. Table 16 provides the Likert scale definitions for this research. Again, a justification text box will be included after each statement for panel members to provide additional information regarding their choices. Due to the subjective nature of factors influencing SNE development, it the research anticipates that experts will propose many unique factors; therefore, the researcher will remove the least popular factors after the second round of the Delphi. Rounds four and five of the Delphi will preserve the twenty top factors. The survey will include a progress bar along with a reduction of questions to reduce survey fatigue by the expert panel.

Table 16. Likert-type scale response: Level of Agreement

Level of Agreement	
1	Strongly disagree
2	Disagree
3	Somewhat disagree
4	Neither agree or disagree
5	Somewhat agree
6	Agree
7	Strongly agree

Delphi Panel Feedback

After the conclusion of each of the four Delphi rounds, only respondents who completed the surveys will be provided (via email) with a feedback package. The package will contain a statistical summary of responses and distributions for each question statement along with a copy of unique ID coded individual responses. Since respondents will only know their own unique ID, they will be able to compare their responses against their assigned group's panel population and the panel as a whole. This will allow the respondent to revise their judgments in subsequent rounds of the Delphi. The research will implement a "Green-Amber-Red" coloring schema to highlight areas where individual experts vary with the group median. Green signifies the expert is within one point of the standard deviation of the panel median. Amber signifies the expert is within one to two points of the standard deviations of the panel median. Red signifies the expert is outside the standard deviation of the panel median and is preventing consensus agreement on this factor. Figure 17 is an example expert panel summary chart that the research will provide with each feedback report.

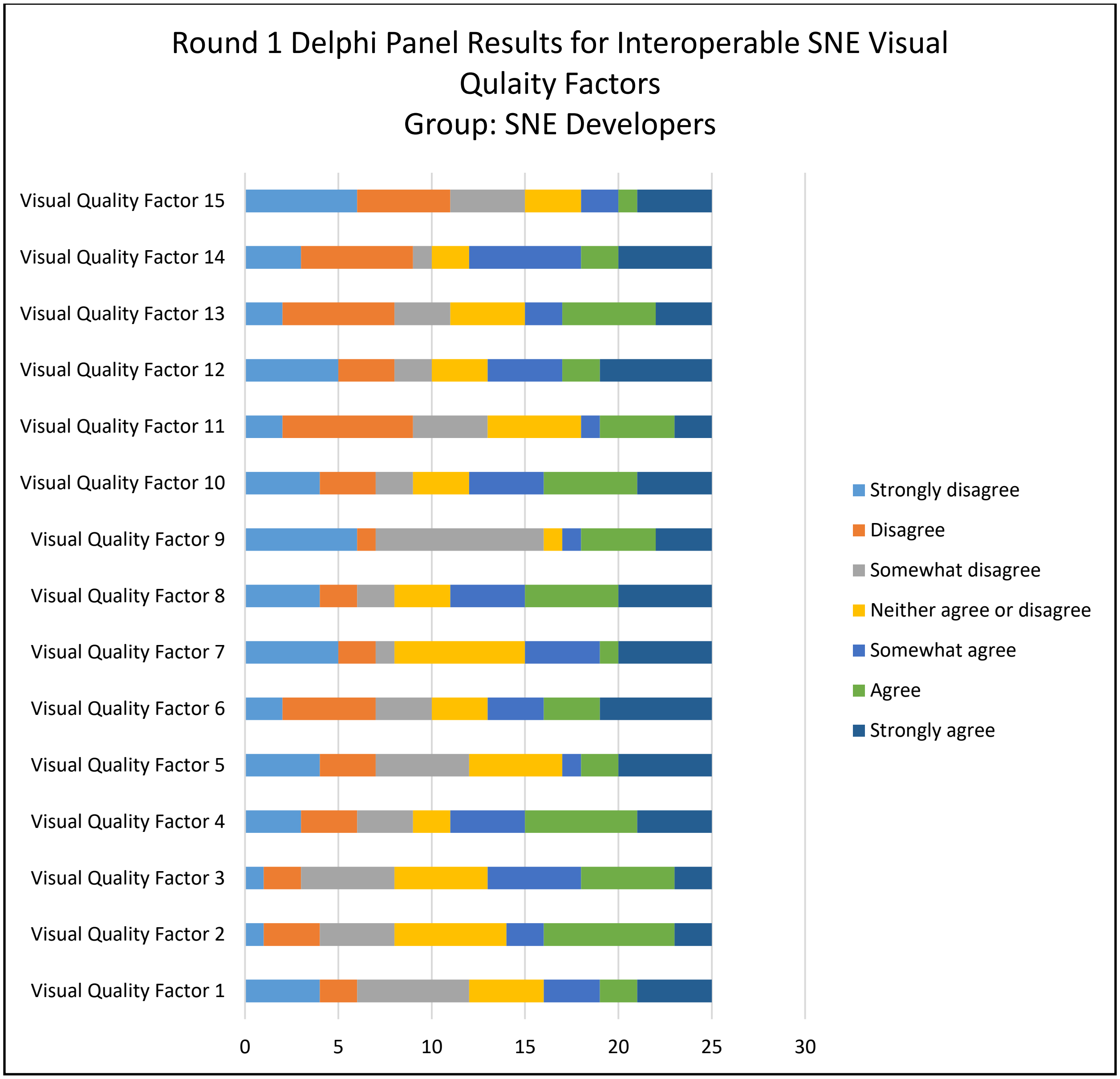


Figure 17. Example Round One Delphi Feedback Summary Chart for SNE Developer Group

Selection of Expert Panel

Synthetic Natural Environment experts often fall into one of three primary populations: Developers, Users/Operators, and Managers. Developers are those experts responsible for constructing the SNE through terrain generation software, GIS processes, and testing. This research defines Users/Operators as SNE end users who either utilize the SNE to train other or are being trained on a particular task. Finally, Managers are the experts who oversee the development and utilization of SNE products, typically program managers, project directors, and chief engineers. This research will therefore integrate these three population groups into a single Delphi study and each group will only receive feedback results within their group. The literature suggests that ten to fifteen experts is typically sufficient for a panel with a homogenous background on the topic of interest; however, if a heterogeneous mixture of populations are involved, the panel size should be closer to 50 (Hsu & Sanford, 2007). After each Delphi round, a Wilcoxon Signed-Rank Test can be computed to test whether an expert panel's rating for each statement was statistically different from a neutral (ambivalence) rating. If the statement is not significantly different from neutral after the fourth round, then the statement can be removed from consideration. Additionally, a Kruskal Wallis (KW) test, the nonparametric form of the single sample ANOVA, will be used to determine if there was a statistical difference in rank between statements. Each statement is assumed to be independent from one another since experts will not provide feedback indicating dependencies between statements between rounds, thus satisfying a critical assumption of the Kruskal Wallis test. The mean ranks calculated by the KW will be used as the importance rankings for each statement. Because of this, the recommended target sample size for each population groups is 25 experts (Cohen, 1992). By targeting 25 respondents, alternative nonparametric significance tests can be conducted in the event of participant dropout (Cohen, 1992). Finally, a pairwise Mann-Whitney U test, the nonparametric form of a two-sample t test, can also be computed across each statement if the Kruskal Wallis test found differences in any round, identified statements that are statistically different from each other in importance. The primary Delphi panel selection emphasis focuses on first obtaining the previously mentioned three user groups, but this research is also interested in the individual SNE use-case background domains of experts for insight into individual use-case forecasting (aviation, ground, etc.) and to better understand domain distribution. The expert selection process will seek to ensure an equal distribution of domain backgrounds between each of the three primary populations. The research will study the domain-specific backgrounds of each expert and their familiarity with virtual domain-specific use cases through the previously discussed demographic questionnaire.

Experts for each of the three primary SNE populations will be selected trough three primary means: professional contact network, peer-nomination, and literature review. Since this research is based on a case study of the U.S. Army ITE and SE Core, the researcher will leverage his professional contact network throughout the U.S. Army civilian and contractor Modeling and Simulation community as the basis for expert panel selection. The researcher will then query this initial set of experts to nominate additional experts whom they believed are qualified for this survey. Finally, the SNE literature review performed for Chapter 1 will serve as a supplement for expert panel solicitation.

Piloting the Delphi

Piloting the Delphi is a crucial step to validate the research design (Gill et al., 2013; Latour et al., 2009). A Delphi pilot study will assess the validity, reliability, and feasibility of the Delphi content, instructions, processes, and data analysis techniques. Additionally, the pilot will gauge ease-of-use and assess time requirements for the surveys. The research will select fifteen panel members for the pilot study, consisting of five members from each population group. The researcher will select these members from a trusted group of government, industry, and academia peers. In addition to providing feedback on the survey utility, the pilot panel will also be responsible for nominating additional experts for the official Delphi study.

Distributing the Delphi Surveys

The experiment will maintain a listing of all expert panel names, email addresses, and their population group. Each panel member will be assigned a unique ID in order to track responses between Delphi rounds (Gill et al., 2013). Using the Survey Monkey email function, the researcher will distribute a personalized email message to each expert panel member along with a link to the survey. The researcher will strip participant names and email addresses out of the data when importing survey results into a statistical software package.

The researcher will use an email scheduling application to distribute link to the panels. Each round of the survey will be open for two-weeks (Fan & Yan, 2010) and three follow-up reminder emails will be sent to non-responders per round. The list of respondents from each round will then be copied into a new recipient list for each subsequent round, thus accounting for dropout.

Ethical Considerations

Since this research will involve the distribution of electronic surveys to human subjects, the appropriate approvals will be obtained by the University of Central Florida Institutional Review Board (IRB) prior to conducting the Delphi or pilot Delphi. The IRB approval letter for this research is provided in APPENDIX E: UCF IRB APPROVAL LETTER.

Since this research will involve the storage of potential Personally Identifiable Information (PII) of panel members, it will take additional security steps. Research conducted over the web pose unique threats to privacy and confidentiality, but the fact that this research is not of particular sensitive nature potentially reduces this threat. Regardless, respect and regard for human participants privacy and confidentiality should be a priority in upholding principles of research ethics (Gill et al., 2013).

Survey Monkey maintains a high level of security from a variety of means. From a physical perspective, they store servers in locked cages, monitor digital surveillance systems, and utilize facility intrusion detections systems. Electronically they use Secure Socket Layer (SSL) password encryption, weekly network security audits, firewall restrictions, and daily hacker safe scans. Additionally, all data is backed up hourly, with daily backup and offsite storage (Gill et al., 2013; SurveyMonkey, 2016b).

The disclaimer will notify participants that access to the survey will be password protected and that Survey Monkey secure servers will host their data. The research will also inform participants them that the researcher will treat their data as confidential, but due to the need to provide feedback, responses will not be anonymous; however, only the primary researcher will be able to link respondents to their responses since PII will be stripped from the feedback and it will only contain the unique ID. As an additional step, once the survey and data analysis is completed, it is possible to contact Survey Monkey to wipe the Delphi data from its servers, although the data can potentially still be retrievable for up to 12 months due to applicable law (SurveyMonkey, 2016b).

Concluding the Delphi

The conclusion of the Delphi will occur after three rounds. In the event that none of the factors reach 80-percent consensus after three rounds, the highest factors will be selected for analysis. In the event of a tie of factor rankings, the highest factor among the User/Operator group will be selected since this is this group represents the end user of a SNE product and therefore acts represents a greater weight of the “voice of the customer”.

The researcher will synthesize the top consensus-reaching statements about SNE VAQ resulting from the Delphi into concise VAQ factors and recommended high and low levels for each factor. The research will then pass these factors and associated levels on to phase two of this research to be used in experimental design for the basis of SNE generation in Phase 3 and conjoint analysis in Phase four.

Phase Two: Experimental Design and Factor Analysis

Using the primary VAQ factors and factor levels generated from the Delphi panel of experts, this research will seek to implement a fractional factorial screening design as the basis for non-confounding design profile to be used for SNE generation. This fractional factorial design will also be used as a basis for the conjoint analysis that the research will conduct in phase four. A screening design is an efficient means to screen a large number of factors while operating under the assumption that only a few factors are of primary importance to users. The primary purpose of this design will be to identify the primary VAQ factors, or main effects. The research will also seek to implement a Resolution III design if the number of factors and factor-levels support such an approach. This will allow a reduction in the number of SNE representations to be generated, since this is a time-consuming effort. The Resolution III design will allow the researcher to efficiently explore the effects of many factors. A Resolution IV design will also be considered if the identified VAQ factors are sufficiently easy to vary within the SNE generation process.

Phase Three: SNE Generation

Once the researcher has designed the experiment, a SNE dataset must be selected that encompasses the desired SNE factors for study. This research will leverage the “Emerald City” dataset from the Synthetic Environment Core (SE Core) program. This dataset is a Publicly Distributed dataset that was used as the foundation for the Operation Blended Warrior (OBW) demonstration at the 2019 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC). OBW focuses on developing a virtual environment for live virtual constructive simulations for training, but from the perspective of what it takes to put it together. OBW is a showcase of industry modeling and simulation capabilities, in conjunction with Department of Defense M&S capabilities, but more importantly, an opportunity to collect data on challenges that arise during the exercise. The Emerald City dataset includes a 10km by 10km area of downtown Seattle, Washington as seen in Figure 18. Location and Size of SNE Core Emerald City dataset. The dataset includes multiple sources of satellite imagery, elevation data, 3D geospecific and geotypical models, and GIS vector data. The dataset was successfully converted to multiple simulation runtime formats by industry partners participating in OBW. All data used or created for this dataset is based on publicly available data and was therefore classified as “DISTRIBUTION A. Approved for public release: distribution unlimited.”

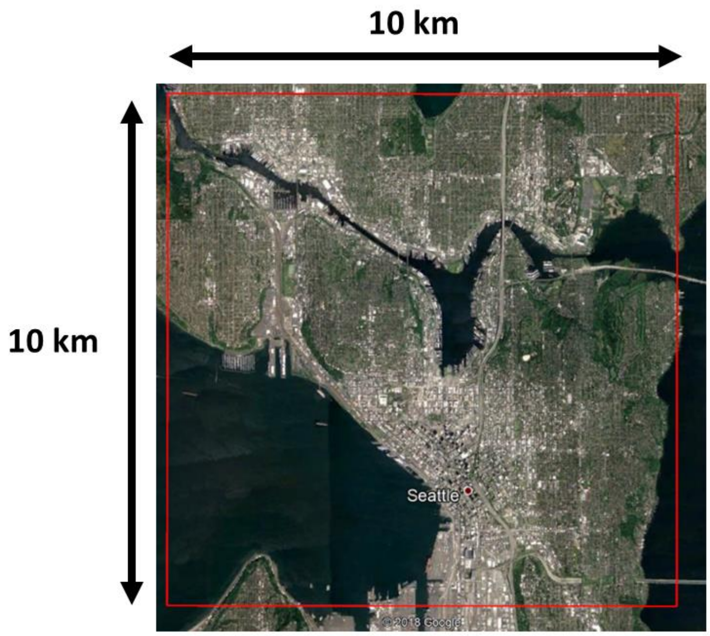


Figure 18. Location and Size of SNE Core Emerald City dataset.

Due to hardware resource limitations, it is not feasible to generate an entire dataset from those represented in the Emerald City dataset. Therefore, only several terrain tiles (measuring 1x1km) of the dataset will be generated, but care will be taken to ensure that the required factors are represented in those selected terrain tiles.

SNE Generation Software Applications

This research will leverage the Conform™ terrain generation software by GameSim to generate the required SNE representations. This research will also implement the Unreal Engine 4 (UE4) visualization software due to its growing popularity across military simulations (Allen, 2011; Prasithsangaree, Manojlovich, Hughes, & Lewis, 2004; Shen & Zhou, 2006; R. Smith, 2006;

Veziridis, Karampelas, & Lekea, 2017). In order to take full advantage of Conform™ coupled with the UE4 level editor's 64-bit and scalable Multi-Machine Build capabilities, a powerful computer workstation is required. Table 17 provides a detailed breakdown of the hardware utilized for terrain generation in this research.

Table 17. SNE Generation PC Hardware Specifications

Component	**Description**
CPU	Intel Core i7-7700HQ (2.8 GHz)
Graphics Processing	NVIDIA GeForce GTX 1070
Memory	16.0 GB
Operating System	Microsoft Windows 10

Since this research will also require the manipulation of 3D objected represented in the SNE (buildings, trees, etc.), AutoDesk Maya™ will be utilized for 3D model creation and editing.

SNE Visual Aesthetic Quality Factor Manipulation

The research will utilize Conform to generate a number of different SNE representations by modifying each of the factors affecting SNE VAQ identified in the Delphi based on the selected experimental design. Conform offers the ability to modify all aspects of SNE representation through its user interface and the ability to upload multiple GIS datasets. This research provides several notional examples for factor manipulation are provided since formal discussion on this topic cannot be conducted until completion of the Delphi.

Conform terrain generation is accomplished through layering of GIS vectors, cultural data, aerial imagery, and terrain elevation. Conform's user interface provides an efficient way to manage and order these layers, much like commercial photo editing software (Adobe Photoshop, etc.). The user can toggle these layers on and off by the user to experiment how different setting affect the terrain scene. Through this feature, it is possible to build a single Conform project file for this research that contains multiple levels of each of the SNE aesthetic quality factors. The user can toggle the appropriate layers on or off corresponding to the specific SNE design in the generation of a terrain file. A researcher can repeat this process for each of the subsequent SNE designs required to generate all terrain files needed for the conjoint study.

Selection of SNE Area of Interest

Since all of the SNE aesthetic quality factors will not occur in the same localized region of the terrain representation, the research must select area for output that contains an example of each factor. The size of this area will be largely unknown until the SNE is developed, but notionally a 2km x 2km are will be recommended. A randomization process to identify a random coordinate for an area of the SNE in order to reduce bias in the experiment. This point will be the center-point for the area of interest for use in the experiment. The researcher will analyze this area to confirm it contains the required SNE VAQ factors and will export a 2km x 2km area surrounding this center point to an FilmBox (.FBX) format for runtime engine conversion.

Runtime Simulation Generation - Unreal Engine 4

Unreal Engine is a suite of game development tools released by Epic Games in 1998 (cite). The latest offering, Unreal Engine 4 (UE4) was launched in 2014 and offered many advanced improvements over previous iterations. The Unreal Engine is made up of several components that work together to develop game content. Its massive system of tools and editors allows developers to organize assets and manipulate them to create the content for gameplay. UE4 components include a sound engine, physics engine, graphics engine, input and the Gameplay framework, and online module (Epic Games, 2020). The UE4 graphics engine will be used ibn this research. The researcher will make heavy use of the UE4 Material Editor. Shaders and Materials give objects its unique color and texture. Unreal Engine 4 makes use of physically-based shading. This material pipeline gives developers greater control over the look and feel of an object. Physically-based shading has a more detailed relationship of light and its surface. This theory binds two physical attributes (microsurface detail and reflectivity) to achieve the final look of the object (Vries, 2020). UE4 implements a concept known as “Landscapes” to create large outdoor spaces. UE4 provides sculpting and painting tools through the Landscape system to help developers. An efficient level of detail (LOD) system and memory utilization allows large scaled terrain shaping. There is also a Foliage editor to apply grass, snow, and sand into the outdoor environment.

UE4 also provides powerful capabilities for lights and shadows through a set of basic lights that could be easily placed in a game rendering. They are Directional Light, Point Light, Spot Light, and Sky Light. This lighting will be used to provide realistic environmental lighting for SNEs developed under this research. The UE4 Matinee Editor will also be used to create video fly0throughs of each SNE for use in follow-on survey design.

Phase Four: End User-based Conjoint Analysis

Phase four of this research will focus on the execution of a traditional full-profile conjoint analysis study involving a population of end users. Since other variations of conjoint (i.e. choice-based and adaptive) are not able to be easily adapted to SNE generation and representation, this research will utilize a traditional full-profile conjoint method. The other Conjoint Analysis methods only focus on several factors per task and would require a significant amount of SNE representations per survey in order to account for all factor configuration variations. Additionally, adaptive conjoint does not allow for pre-built SNE representations, since they generate surveys 'on the fly' based on participant's answers throughout the study. This application of full-profile conjoint will be unique since traditional studies since researchers typically conduct this method with a text-based matrix that display all product configurations at once for a user to rank/rate. This study will instead present the participant with a series of graphical SNE representation to rank without explicably knowing the factor and level composition of each –they will therefore choose solely based on the visual aesthetic of the SNE and not the inherent composition of SNE factor/level choices. This will be accomplished through a Graeco-Latin square survey design. Additionally, price/cost will not play a factor in this conjoint study.

Conjoint Design

As established, a traditional-full profile conjoint analysis will be implemented utilizing a screening-design informed by factors identified from the Delphi study. Participants will therefore rank order each of the SNE choices based on their preferred visual aesthetic look and feel for an interoperable, multi-domain use case. *Ranking* of SNE choices is preferred to *rating* the choices since it will be easier to constrain the calculated utilities to conform with expectations (Sawtooth Software, 2002). The challenge with using a rating scale is that a novice population may have no frame of reference for a virtual SNE environment and will make comparisons of aesthetic quality based on their own frame of reference (i.e. video games or Google Earth.). This frame of reference may be of much higher aesthetic quality than the SNE choices provided to them, thereby assigning a low rating to all choices. Therefore, the ranking of SNE choices will force participants to consider choices relative to all other choices thus enforcing constraints. The challenge of using a ranking system is that it will require more time and concentration by the participant since they will have to manage moving between many SNE example products instead of ranking based on a traditional text-based conjoint study.

The conjoint analysis phase of this research will also utilize Survey Monkey™ to design and conduct the conjoint analysis study. Survey Monkey includes several methods for designing ranking statements through its user interface. One method is to include an option for a number of radio buttons for each SNE design. Another option is to utilize drop-down selection boxes to select the desired rank for a given configuration (SurveyMonkey, 2016a). Each option only allows the participant to select only one of any given rank. Figure 19 and Figure 20 provide several examples of Survey Monkey ranking design. A common option in conjoint surveys is the ability to select ‘None’ or ‘N/A’ as a desired response. This research will not utilize this option since in an interoperable scenario, confederated virtual simulators may have very dissimilar image generation capabilities between them; some high fidelity and some low fidelity and user’s will be forced to work through this disparity. The previously discussed frame of reference problem is also justification for exclusion of the null response option.

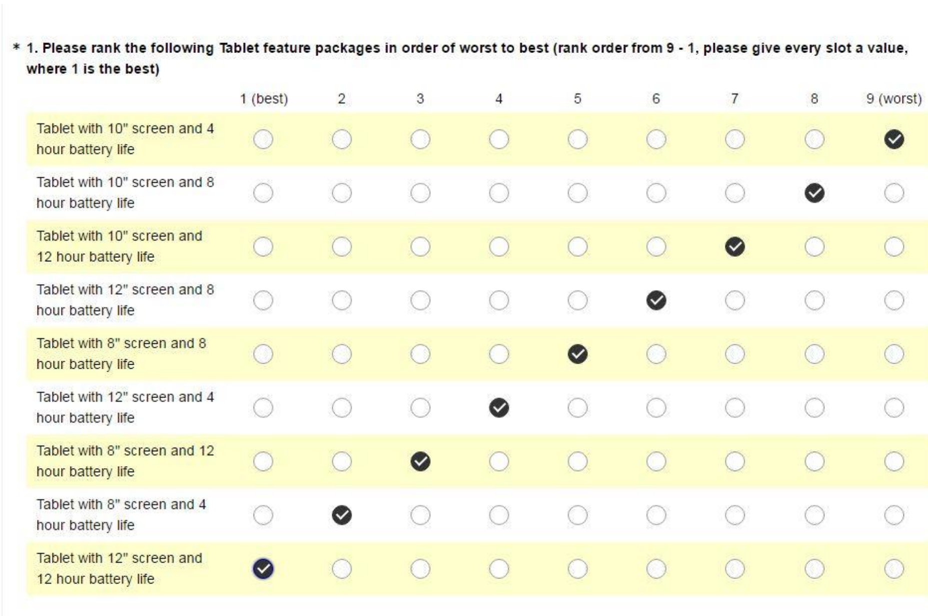


Figure 19. Ranking Survey with Radio Buttons

Source: (SurveyMonkey, 2016a)

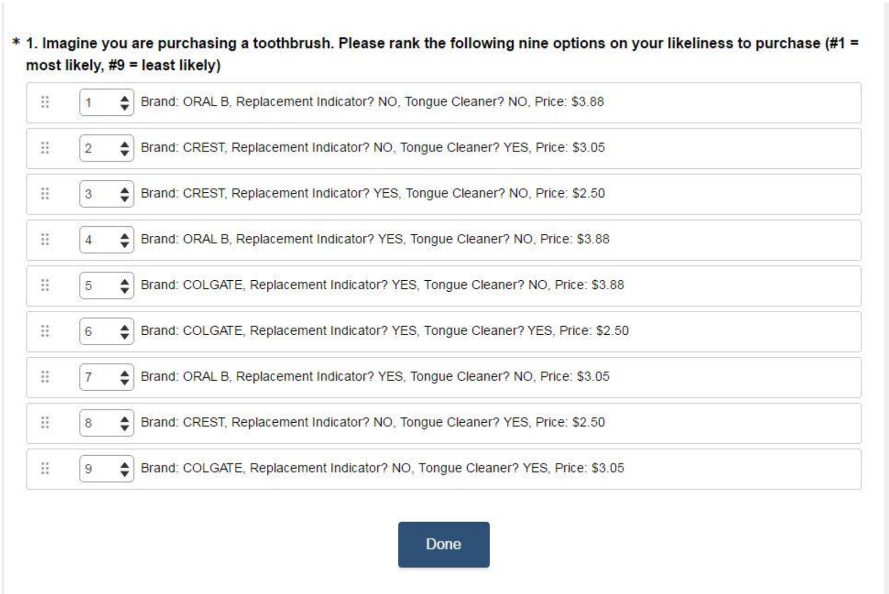


Figure 20. Ranking Survey with Drop-Down Selection

Source: (SurveyMonkey, 2016a)

Survey Monkey also enables the use of video files in its surveys through embedded links using the YouTube Player API for iframe embeds. Thus, the researcher can include videos to the survey without taking up large amounts of real estate on the questionnaire and respondents can view them in the same survey window with full playback controls, such as pause, play, and loop.

Unlike the Delphi, there is no need for users to provide their contact information or additional respondent-specific details since there will not be a follow-up conjoint study or feedback round. Figure 20 provides a mock-up of the planned conjoint analysis survey interface. A YouTube video will be embedded into the Survey Monkey webpage. The video will be a quad-screen video that will present four different video fly-throughs of four different SNE designs identified from the Delphi study. Participants will then be asked to rank each of the four SNEs presented in the video. Ties for rankings will not be allowed. Once complete, the participant will move to the next video presenting the next set of SNEs.

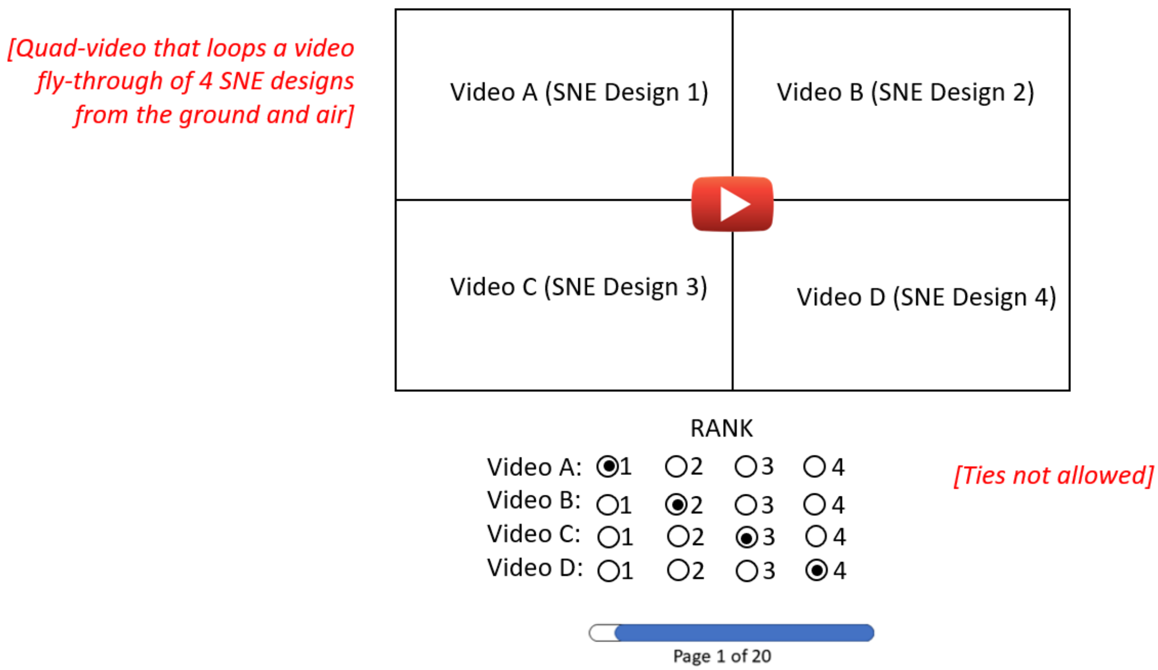


Figure 21. Proposed Conjoint Analysis Survey Design for SNE Visual Aesthetic Quality

Conjoint Participant Selection

The target of the conjoint analysis study will be the SNE end-user population. This population will include both SNE simulation operators (i.e. trainers) and SNE users (i.e. trainees). As in the Dephi study previously, the preferences of users from individual domain use-cases are also of interest. Through the Conjoint Analysis demographic response questionnaire, users will indicate their domain-specific SNE background as follows: Aviation, Ground, or Multi-Domain Operations.

Sample sizes for conjoint analysis studies are a commonly debated topic and there is no definitive guidance on the correct size (Curry, 2001; B. K. Orme, 2010c; Vilikus, n.d.). Some experts agree that at least 70 to 100 respondents make the results stable, but a common rule of thumb is to have a ration of between 5 and 10 respondents per conjoint parameter (Curry, 2001). The number of parameters is equal to the total number of levels across all factors minus the total number of factors plus one.

This research will gather participants from a variety of sources. Since the primary focus of this research is military M&S, the research will draw a majority of participants from this community. The primary source of user participants will be from the actual end user/operator Army Integrated Training Environment (ITE) and Synthetic Training Environment (STE) communities. Through the Program Executive Office for Simulation, Training, and Instrumentation's (PEO STRI) Field Operations office, the research will request actual simulation operator contractors to participate in the conjoint survey. These operators are geographically located across the globe, thus the electronic Survey Monkey proctoring method is well suited. This research will also draw participants from the Army National Simulation Center (NSC) at Fort Leavenworth –the accreditation authority for the U.S. Army ITE virtual collective simulation systems. Further, this research will recruit participants from US Army Centers of Excellence (COEs): the Maneuver COE at Fort Benning and the Aviation COE at Fort Rucker.

In accordance with UCF IRB policy, all participants will be required to complete an experiment consent form. The intention is to provide this consent form as an attachment to the electronic survey for electronic completion. Appendix D provides an approved version of this consent form along with the corresponding UCF IRB approval form in Appendix E.

Piloting the Conjoint Study

Much like the Delphi, piloting the Conjoint Analysis study will be critical in identifying error and time constraints in the conjoint design. This pilot study will also validate the selection of Ranking over rating as a preference method. If pilot user feedback identifies ranking to be too complex a task for this study, the research will instead utilize *rating* and the research will revise data analysis methods to reflect interval data rather than ordinal.

The research will draw 5 users from a pool of trusted end users in both novice and advanced user groups to participate on the conjoint pilot. The research will select the majority of these trusted users from local simulation contractors and stakeholders to facilitate direct feedback communication.

Conducting the Conjoint Study

As addressed, this research will conduct a conjoint study electronically via Survey Monkey. The first page of the survey will provide an overview of the research methodology. The first page of the research will provide respondents with an overview of the interoperability use-case on which SNE respondents will evaluate. The next page of the survey will act as the UCF Experiment Consent Form. The page will outline the experiment consent language and will provide a required multiple-choice ‘yes or no’ selection along with a text box for electronic signature. A ‘yes’ answer will move forward in the survey and a ‘no’ will present a disqualification page.

The third page of the survey will provide a short demographics questionnaire in order to gauge the user’s domain background. The questionnaire will also identify whether a participant is military or civilian.

The next pages of the survey will serve as the primary Conjoint Analysis content. These pages will each include a single embedded YouTube video displaying a flythrough of four-different SNE designs, informed by the Delphi study results. Participants will then be asked to rank each of the four SNE presented in the video based on a provided training scenario. This scenario will be the same across each video and set of SNE. Ties will not be allowed for each ranking task. Each of these ranking pages will also include an optional comment box for participants to justify or expand upon their rankings, if they desire. Once all paged are complete, users will submit results via a button on the survey. Additionally, each page of the survey will include an option for participants to withdraw from the survey. This is in accordance with UCF IRB procedures.

Phase Five: Data Analysis and Recommendations

Phase five of this research will focus on the data analysis of the respondent-based conjoint analysis as well as the analysis between the SNE Core status quo SNE issue methodology versus the Delphi and Conjoint Analysis paradigm presented in this research.

Conjoint Statistical Analysis

This study will implement a combination of regression models within the JMP statistical package to estimate factor importance and major factor interactions appropriate for the selected experimental design.

Domain-specific Use-Cases and Additional Data Analysis

Once the interoperable use-case of SNE visual aesthetic quality factors are analyzed, additional analysis can conducted to identify the preference for the domain-specific use-cases (i.e. aviation, ground, and multi-domain). Furthermore, comparisons can be made between the initial expert forecasts conducted through the Delphi in phase one of this research against the actual user preferences identified by the conjoint analysis completed in this final phase of research. The significance of this final comparison is to highlight the gaps between expert perceptions of VAQ factors forecasted in the abstract, which may be associated with the conceptual phase, versus real-world user preferences after SNE generation. Understanding the level of correlation between the two groups assessments of VAQ will lend a level of confidence to usefulness of using the Delphi to determine VAQ factors. If consistency exists between expert-forecasted VAQ factors and end user VAQ factors, Quality Function Deployment (QFD) may be able to be utilized to abstract the correlated set of expertly accepted and user validated primary SNE VAQ factors into a series of SNE generation process improvements to influence a new SNE VV&A paradigm.

Delphi Study versus Traditional SNE Core VV&A

Additionally, this research will utilize the SE Core method of SNE analyzing user preferences as the M&S status quo-like methodology. During formal V&V events, Subject Matter Experts and terrain stakeholders record issues against an SE Core generated SNE in the form of Discrepancy Reports (DRs). The SMEs then assign a severity to each DR based on the issue's impact to training. Assess a SNE for impact on training is beyond the VAQ scope of this dissertation but analysis may prove insightful into correlation between abstract VAQ, end-user VAQ, and training impact issues. Table 18 summarizes training impact severity levels. Once assigned a severity, SE Core developers log the DRs into an issue tracking software system to facilitate efficient closure of the issue once corrected. Prior to generating a new SNE, SE Core will analyze issues plaguing past SNEs in an effort to more rigorously catch these issues during initial V&V.

Table 18. SNE Core Issue Severity Definitions

Priority	**Description**
1	Safety issue or prevents continuation of an operational or mission essential capability or testing
2	Adversely affects the accomplishment of an operational or mission essential capability and no work-around solution acceptable to the government is known
3	Adversely affects the accomplishment of an operational or mission essential capability but a work-around solution is known
4	Results in user/operator inconvenience or annoyance but does not affect a required operational or mission essential capability
5	Any other effect (e.g., documentation error)

This DR report can be ‘data mined’ to bin DRs into like-issue categories. The researcher will then rank these categories based on the number of DR instances that fall within each high-level issue category. This ranking represents the user priority of this issue category.

Conjoint Analysis versus Traditional SE Core VV&A

Once the researcher categorizes and ranks the historical SNE Core DRs, the research will assign a raw importance rating to each high-level category by dividing the total number of DRs within that category by the average severity rating of all DRs within the same category. A lower DR severity ranking indicates a more important DR; therefore, division is used over multiplication to obtain an importance ranking. The researcher can then use the issue categories and associated importance rankings to address hypothesis 3 through comparison of VAQ factor importance levels calculated during the traditional full-profile conjoint analysis study in phase four of this research. Therefore, differences in user preferences as identified by VAQ levels of importance and part-worth utilities will lead to a failure to accept the null-hypothesis associated with hypothesis 2.

Methodology Summary

Chapter three of this research provided a detailed research plan to identify the significant factors affecting the visual aesthetic quality of synthetic natural environments. This chapter detailed an application of the Delphi study to forecast the significant SNE VAQ factors among the SNE expert community. The top VAQ factors may be used to design an optimized Resolution III fractional factorial screening design. These factors are then used a basis for SNE generation in phase three in order to produce diverse SNE prototypes to be used in a subsequent Phase 4 conjoint analysis study. Through this traditional full-profile conjoint study, the research identifies end user preferences to determine the importance of SNE VAQ factors and factor interactions then compares these importance levels to VAQ factors identified abstractly during the Delphi study. This will provide insight to the utility of Delphi in developing VAQ factors for a SNE during the conceptual design phase. Further, if abstract and end-user assessments are consistent, Quality Function Deployment (QFD) can be utilized to abstract the correlated set of expertly accepted and user validated primary SNE VAQ factors into a series of SNE generation process improvements to influence a new SNE VV&A paradigm. In addition, abstract and end-user VAQ factors will be analyzed with respect to historical training impacts. The following Chapter will provide an overview of the results of this experiment.

CHAPTER FOUR: EXPERIMENTAL EXECUTION AND DATA ANALYSIS

Phase One Results: Delphi Study of SNE Quality Factors

During Phase One, 62 experts in SNE's of one or more simulator platforms were invited to take part in the research. Taking advantage of game development experience and success, selection emphasized developers rather than balance between the three sub-communities. Invitation selection emphasized nominations by pilot test group members, professional network contacts, and publication record. This study also implemented recommended strategies used to enhance expert panel response rate and motivation as proposed by Gill et al. which proved to be extremely beneficial for maximizing panel engagement and communication (Gill et al., 2013).

Participants were contacted and advised of their rights in accordance with protocols approved by the University of Central Florida (UCF) Institutional Review Board (IRB).

This research utilized the web-based Survey Monkey platform to implement the Web-based Delphi survey rounds. Appendix A of this document provides a sample of the Round 1 and Round 2 Delphi Surveys that were distributed as part of this research. Each survey round was sent directly to participants through email and remained open for a period of two full weeks. Reminder email messages to non-respondents occurred after seven days and 24-hours before the round closed. After the data analysis and typically a week later, currently active survey participants were provided a feedback report and a link to the next survey round. Appendix B provides a sample feedback report from this Delphi survey.

Delphi Method Round One: Open-Ended Survey

The Open-Ended Survey provided participants with an overview of the entire study, along with the research goal and background information on the Delphi Method. Additionally, a demographics questionnaire collected: (1) contact information and (2) self-identification with community and domain-specific backgrounds (i.e. aviation, ground, dismounted, etc.). The heart of the survey contained: (1) a detailed overview of the SNE Visual Aesthetic Quality (VAQ) research; (2) survey completion instructions; (3) research questions with a simulation training scenario for real-world contextual framing (Table 19); (4) ten blank fields for experts to provide their own open-ended identification of VAQ factors; (5) a specific training or context example, and (6) recommended parameters for measuring the quality of each factor. VAQ factors sought were those that could potentially impact realism. Participants were each asked to identify at least five VAQ factors.

Table 19. Research question and training scenario provided in round one survey.

Research Question	*What are the primary visual appearance considerations when designing synthetic environments for a virtual distributed interactive simulation exercise?*
Research Training Scenario	*The researcher acknowledges that training tasks and target simulator platforms will ultimately influence visual importance considerations for terrain. Therefore, this research will focus on the use case of an integrated air and ground virtual simulation exercise consisting of one or more rotary-wing aircraft simulators and a mix of armored, wheeled, and dismounted soldier simulators.*

The Open-Ended Survey yielded a total of 189 SNE VAQ factors, and associated parameters. Each proposed factor was reviewed and placed into a specific category. Prominent categories included run-time visual rendering, synthetic environment effects, 3D models, visual textures, and cultural terrain features. The factors with the highest similarity count from each category were selected for the round two survey. A word cloud, shown in Figure 22 based on the frequency of common topics identified in the open-ended feedback was also generated to support this analysis. Similar factors were rewritten into 24 lettered statements (*1* to *24*) for clarity and standardization regarding SNE VAQ as seen in Table 20.

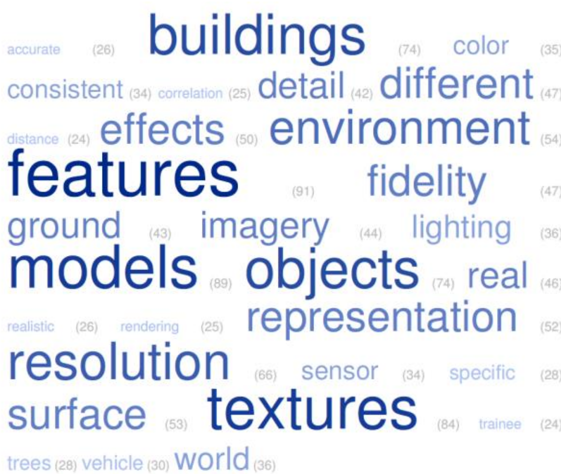


Figure 22. Word cloud generated through the open-ended survey responses from Round One of the Delphi. Numbers in parentheses indicate the frequency of occurrence of each word. Non-SNE related words were excluded from analysis and illustration.

Table 20. Synthesized final statements from the Open-Ended Survey. A simplified descriptor for each factor is also provided for efficient analysis purposes later.

Statement	SNE Visual Aesthetic Quality Statements	Simplified Visual Aesthetic Quality Factor Descriptor
1	Realistic sensor representations (i.e. accurate thermal or infrared picture) for the simulated weapon platform is important to the visual quality of virtual simulation for the training scenario.	Realistic sensor representations
2	Accurate visual representation of dense urban terrain and megacities is important to the visual quality of virtual simulation for the training scenario.	Dense urban terrain and megacities
3	Addressing simulation fidelity trade-offs upfront with trainees to better accept unrealistic representations and/or simulation limitations is important to the	Addressing simulation fidelity trade-offs upfront with trainees

Statement	SNE Visual Aesthetic Quality Statements	Simplified Visual Aesthetic Quality Factor Descriptor
	visual quality of virtual simulation for the training scenario.	
4	Consistent time of day representation across all simulations/simulators is important to the visual quality of virtual simulation for the training scenario.	Time of day representation
5	Physics-based damage and battle-worn textures applied to terrain features within the synthetic world are important to the visual quality of virtual simulation for the training scenario.	Damaged and battle-worn object and environment textures
6	Removal of scintillation and other visual anomalies in the scene (e.g. flickering of pixels) is important to the visual quality of virtual simulation for the training scenario.	Removal of scintillation anomalies
7	Consistent use of a common artistic theme and palette across elements in the synthetic world is important to the visual quality of virtual simulation for the training scenario.	Common artistic theme and palette across elements in the virtual world
8	Accurate object position and orientation relative to other content in the synthetic world (e.g. houses facing sidewalks/streets) is important to the visual quality of virtual simulation for the training scenario.	Accurate object position and orientation
9	Inclusion of features providing spatial relationship cues in the rendered scene (i.e. features that provide geometric relationships to help in seeing and judging size and distance) is important to the visual quality of virtual simulation for the training scenario.	Inclusion of features providing spatial relationship cues
10	Consistent color and contrast rendition across the visual scene and connected image generators is important for visual quality of virtual simulations for the training scenario.	Consistent color and contrast rendition across the visual scene
11	Vegetation density and fidelity is important for fair fight and visual quality of virtual simulation for the training scenario.	Vegetation density and fidelity
12	Realistic atmospheric, environmental, and weather effects, such as clouds, fog, dust, precipitation, and aerosol disbursement, is important to the visual quality of virtual simulation for the training scenario.	Realistic atmospheric, environmental, and weather effects
13	Realistic and high-fidelity ground texture/imagery, that is free of visual artifacts, is important to the visual quality of virtual simulations for the training scenario.	High-fidelity ground texture/imagery
14	Accurate and organic representation of natural terrain objects and features (i.e. rivers flowing downstream, steep cliff faces, natural vegetation) in the synthetic world is important to the visual quality of virtual simulation for the training scenario.	Organic representation of natural terrain objects
15	Accuracy and density of lighting (points, pools, textures, etc.) across the synthetic world is important for the visual quality of virtual simulations for the training scenario.	Accuracy and density of environmental lighting

Statement	SNE Visual Aesthetic Quality Statements	Simplified Visual Aesthetic Quality Factor Descriptor
16	Using a single geospatial source provider across all simulations/simulators is important to the visual quality of virtual simulation for the training scenario.	Using a single geospatial source provider
17	Density and variation of 3D features across the synthetic world is important to the visual quality of virtual simulations for the training scenario.	Density and variation of 3D features
18	Accurate representation and rendering of transportation networks in the synthetic world is important to the visual quality of virtual simulation for the training scenario.	Accurate transportation networks
19	Accurate depiction and placement of geospecific locations to the real-world is important for the visual quality of virtual simulation for the training scenario.	Accurate geo-specific locations
20	Common Level of Detail (LOD) transition of the out-the-window view (not sensor view), across connected image generators, is important to the visual quality of virtual simulation for the training scenario.	Common Level of Detail (LOD) transition
21	Realistic shadows and feature shading across the synthetic world scene is important to the visual quality of virtual simulation for the training scenario.	Realistic shadows and feature shading
22	Inclusion of environmental clutter (i.e. benches, signage, power poles, rubble) throughout the synthetic world is important to the visual quality of virtual simulation for the training scenario.	Environmental clutter
23	Accurate 3D representation of the terrain surface elevation, both with the real-world and across connected simulations, in the synthetic world is important to the visual quality of virtual simulation for the training scenario.	Accurate terrain surface elevation
24	Consistent texture density and resolution across the visual scene is important for the visual quality of virtual simulations for the training scenario.	Consistent texture density and resolution

Delphi Method Rounds Two, Three, and Four

The fore mentioned 24 statements were presented to all the experts in round two as VAQ factors. For example, VAQ factors such as scintillation and anomalies were presented to the expert within the factor statement, "Removal of scintillation and other visual anomalies in the scene (e.g. flickering of pixels) is important to the visual quality of virtual simulation for the training scenario." The experts rated the level of importance of each factor statement using a seven-point

Likert-type scale rating shown in Table 21. The seven-point scale allowed for greater data granularity to support analysis of the responses (Gill et al., 2013). Statements were not revised between subsequent rounds though some experts raised private questions or expressed open comments about the statements. Private questions were clarified individually. None of the private questions warranted revision of the VAQ statements.

Table 21. Likert-type scale response: Level of Importance.

Weighting	**Level of Importance**
1	Not at all important
2	Low importance
3	Slightly important
4	Neutral
5	Moderately important
6	Very important
7	Extremely important

Delphi Method: Panel Feedback Report

At the conclusion of Rounds Two and Three, experts shared open comments and rating justifications with other panel members for each VAQ statement. Feedback was captured and passed along to other panel members within the Panel Feedback Report. The Report included a one page summary of all factor medians and rating distributions as recommended by Hsu and Sanford (Hsu & Sanford, 2007) and a one page compilation of panel feedback for each statement (factor). The compilation contained the median statement rating by the panel along with a graphical representation and comments for, against, and neutral (ambivalent) about a given statement. Additionally, experts were given their own prior responses in comparison to the panel's group response.

Delphi Method: Participant Response

Table 22 reports the actual number of respondents in each round. In compliance with the literature, panel members who did not respond in a given round were not invited to the subsequent round. The low of 28 responses in Round Four far exceeded the recommended minimum of 15 participants for a Delphi panel. The highest response rate was Round Three with 94.74%. The Trainer community proved to be the most vigilant throughout this study with only one drop-out and that in final Round Four. The developer community had the highest number of respondents in all rounds, emphasizing the developer perspective more than the manager or trainer perspective. While Keeney *et al.* indicates online questionnaire-based research is often plagued with low response rates below 50% (Keeney, Hasson, & Mckenna, 2006), this study never dropped below 77%. The unrelated exit survey had a response rate below 50% as it was administered several weeks after the conclusion of the Delphi study and did not have any impact on the assessment of the Delphi study results.

Table 22. Response rate between each Delphi study round

	Number Sent	**Number Received by Community**				**Response Rate**
		Developer	*Manager*	*Trainer*	**Total**	
1 (open ended)	62	N/A*	N/A*	N/A*	34	54.84%
2	62	19	12	7	38	61.29%
3	38	18	11	7	36	94.74%
4	36	13	9	6	28	77.77%
Exit Survey	62	N/A**	N/A**	N/A**	15	24.19%

*Population distribution was not known in Round 1
**The exit survey was anonymous and did not require participants to response with identification

Delphi Method: Data Analysis

For analysis purposes, any statement that falls below the fore mentioned 80% threshold is considered for removal from further consideration. Additionally, or alternatively, the Wilcoxon Signed-Rank statistical test may compute whether the expert panel's rating for each statement was statistically different from ambivalence. If a given statement is not significantly different from ambivalence after the fourth round, then the statement may also be considered for removal. Further, the Kruskal Wallis (KW) test, the nonparametric form of the single sample ANOVA, determines if there is a statistical difference in rank between statements. For this research, each statement was assumed to be independent from one another since experts did not provide feedback indicating dependencies between statements between rounds, thus satisfying a critical assumption of the Kruskal Wallis test. The mean of the ratings calculated by the KW became the importance rankings for each statement. If the Kruskal Wallis test found differences in any round, a pairwise Mann-Whitney U test, the nonparametric form of a two-sample t test, identified statements that were statistically different from each other in importance.

To obtain rank order and level of consensus, Delphi panel members, over a period of four months, went through three Delphi rounds analyzing and commenting on the 24 Statements and feedback from other panel members.

Table 23 provides the final rankings of each statements based on the Kruskal Wallis test's mean rank calculation along with the measured consensus among the expert panel for each statement's median rating. APPENDIX F: VAQ STATEMENT IMPORTANCE RANKING CHARTS FROM DELPHI STUDY provides a graphical representation of these VAQ statement rankings across each round. The Kruskal Wallis test confirmed that the median rating across each statement were statistically different across each of the three rounds. The measure of consensus greater than or equal to 80% agreement within a single Delphi round was calculated by the Equation below.

Equation Set: Calculation of Delphi consensus for a given statement within a single round

$$Consensus\ (C_{ø,\Theta}) = Maximum\left(\frac{\sum_{i=L_\alpha}^{L_\beta} Ri_{ø\Theta}}{N_{ø\Theta}}\right)$$

$$Delphi\ Consensus\ Achievement\ Level = If\left(C_{ø,\Theta}\ \geq 80\%\right), then\ yes, otherwise\ no.$$

Where:

$ø$= survey round, 2 to 4

Θ = statement *a* to *x*

$Ri_{ø\Theta}$= total expert ratings across survey round ø and statement Θ within i^{th} Likert category, where *i goes from 1 to 7.*

L_α, L_β = two adjacent rating categories (points) on a 7-point Likert scale where $\alpha = i$ *and* $\beta = \alpha+1$

$N_{ø\Theta}$= the total number of ratings on a 7-point Likert scale across survey round ø and statement Θ

Table 23. Rank Order from top to bottom based on mean for the indicated Delphi round. Shaded cells indicate the Statement did not meet the 80% consensus level.

Round 2			Round 3			Round 4		
Statements listed in KW Mean rank order	**KW Mean Rank**	**Round Consensus Level**	**Statements listed in KW Mean rank order**	**KW Mean Rank**	**Round Consensus Level**	**Statements listed in KW Mean rank order**	**KW Mean Rank**	**Round Consensus Level**
Realistic sensor representations	629.54	76.32%	Realistic sensor representations	628.78	88.89%	Realistic sensor representations	527.21	96.43%
Accurate terrain surface elevation	548.09	68.42%	Accurate terrain surface elevation	622	86.11%	Inclusion of features providing spatial relationship cues	491.59	89.29%
Accurate object position and orientation	534.71	73.68%	Inclusion of features providing spatial relationship cues	617.25	88.89%	Accurate terrain surface elevation	478.88	89.29%
Inclusion of features providing spatial relationship cues	525	78.95%	Accurate object position and orientation	576.35	86.11%	Accurate object position and orientation	464.75	92.86%
Vegetation density and fidelity	511.25	76.32%	Addressing simulation fidelity trade-offs upfront with trainees	533.74	69.44%	Dense urban terrain and megacities	427.2	82.14%
Addressing simulation fidelity trade-offs upfront with trainees	500.04	57.89%	Vegetation density and fidelity	526.4	91.67%	Addressing simulation fidelity trade-offs upfront with trainees	406.5	71.43%
Accurate geo-specific locations	499.01	71.05%	Realistic atmospheric, environmental, and weather effects	526.4	91.67%	Vegetation density and fidelity	396.8	89.29%
Consistent texture density and resolution	495.05	76.32%	Dense urban terrain and megacities	521.25	83.33%	Realistic atmospheric, environmental, and weather effects	390.75	92.86%
Removal of scintillation anomalies	491.92	68.42%	Accurate geo-specific locations	461.31	83.33%	Accurate geo-specific locations	379.68	78.57%
Dense urban terrain and megacities	486.89	73.68%	Removal of scintillation anomalies	456.18	83.33%	Removal of scintillation anomalies	342.18	92.86%
Accurate transportation networks	479.24	81.58%	Time of day representation	425.07	83.33%	Time of day representation	338.63	96.43%
High-fidelity ground texture/imagery	476.12	68.42%	Common Level of Detail (LOD) transition	412.4	66.67%	High-fidelity ground texture/imagery	316.77	96.43%

Round 2			Round 3			Round 4		
Statements listed in KW Mean rank order	**KW Mean Rank**	**Round Consensus Level**	**Statements listed in KW Mean rank order**	**KW Mean Rank**	**Round Consensus Level**	**Statements listed in KW Mean rank order**	**KW Mean Rank**	**Round Consensus Level**
Time of day representation	471.43	63.16%	Organic representation of natural terrain objects	411.36	88.89%	Common Level of Detail (LOD) transition	315.34	67.86%
Common Level of Detail (LOD) transition	454.54	65.79%	High-fidelity ground texture/imagery	410.64	91.67%	Accurate transportation networks	311.54	92.86%
Consistent color and contrast rendition across the visual scene	450.86	68.42%	Accurate transportation networks	400.85	91.67%	Organic representation of natural terrain objects	279.82	82.14%
Organic representation of natural terrain objects	417.88	68.42%	Density and variation of 3D features	381.99	88.89%	Environmental clutter	276.36	85.71%
Environmental clutter	406.83	71.05%	Consistent texture density and resolution	380.07	91.67%	Consistent texture density and resolution	255.21	85.71%
Realistic atmospheric, environmental, and weather effects	403.84	63.16%	Environmental clutter	371.65	88.89%	Using a single geospatial source provider	250.98	64.29%
Density and variation of 3D features	397	73.68%	Consistent color and contrast rendition across the visual scene	300.71	77.78%	Consistent color and contrast rendition across the visual scene	247.75	92.86%
Common artistic theme and palette across elements in the virtual world	386.39	52.63%	Accuracy and density of environmental lighting	298.25	83.33%	Density and variation of 3D features	245.71	85.71%
Damaged and battle-worn object and environment textures	363.24	63.16%	Damaged and battle-worn object and environment textures	285.96	80.56%	Accuracy and density of environmental lighting	242.16	89.29%
Accuracy and density of environmental lighting	358.18	71.05%	Realistic shadows and feature shading	283.14	77.78%	Damaged and battle-worn object and environment textures	233.07	85.71%
Realistic shadows and feature shading	339.03	63.16%	Using a single geospatial source provider	282.35	50.00%	Common artistic theme and palette across elements in the virtual world	228.84	89.29%
Using a single geospatial source provider	329.91	36.84%	Common artistic theme and palette across elements in the virtual world	265.92	77.78%	Realistic shadows and feature shading	228.29	78.57%

As seen in Table 24, the Wilcoxon Signed-Rank test indicates Statement 1 is the most distant Statement from neutrality for all Statements in every round. Statement 16, highlighted below, was not statistically different from a 'Neutral' rating across the three analyzed survey rounds.

Table 24. Results of Wilcoxon Signed-Rank Test visa via ambivalence for each structured Delphi survey round. Shaded cells indicate statements with no statistical difference from a neutral rating, or ambivalence.

Statement	Round Two	Round Three	Round Four
	p-value	*p-value*	*p-value*
1	6.2739E-08	3.3756E-08	7.0063E-07
2	0.000013	2.1358E-07	0.000004
3	0.000023	5.7917E-07	0.000015
4	0.000002	1.5304E-07	0.000002
5	0.000046	0.000005	0.000129
6	4.9427E-07	1.1856E-07	0.000003
7	0.007685	0.000181	0.000005
8	2.1969E-07	7.2622E-08	0.000001
9	1.9071E-07	3.6009E-08	0.000001
10	0.000028	3.6009E-08	0.000003
11	2.9383E-07	9.9754E-08	0.000004
12	0.00042	9.9754E-08	0.000005
13	0.000001	8.322E-08	0.000002
14	0.00001	4.0631E-07	0.000111
15	0.000055	0.000006	0.000002
16	0.47996	0.412509	0.141742
17	0.00001	3.6035E-07	0.000004
18	2.0064E-07	7.3437E-08	0.000002
19	0.000006	4.2555E-07	0.000008
20	0.00035	0.000032	0.000329
21	0.003656	0.000101	0.000486

22	0.000055	2.7754E-07	0.000007
23	0.000003	2.1923E-07	0.000001
24	0.000001	1.2158E-07	0.000018

Figure 23 through Figure 25 illustrate the Mann-Whitney pairwise relative levels of importance for statistically significant different statements for Rounds Two, Three, and Four respectively. After the elimination of Statement 16, Figure 23 reveals *Statement 1* dominates in Round Two seven Statements: *5, 7, 12, 15, 17, 21, 22*. Round Three indicates further domination of other Statements but only *Statements 5, 15, and 21* remain dominated from those dominated in Round Two. The divergence observed in Round Four coupled with the low participation rate and high dropout rate undermines the value of Round Four for statistical analysis purposes.

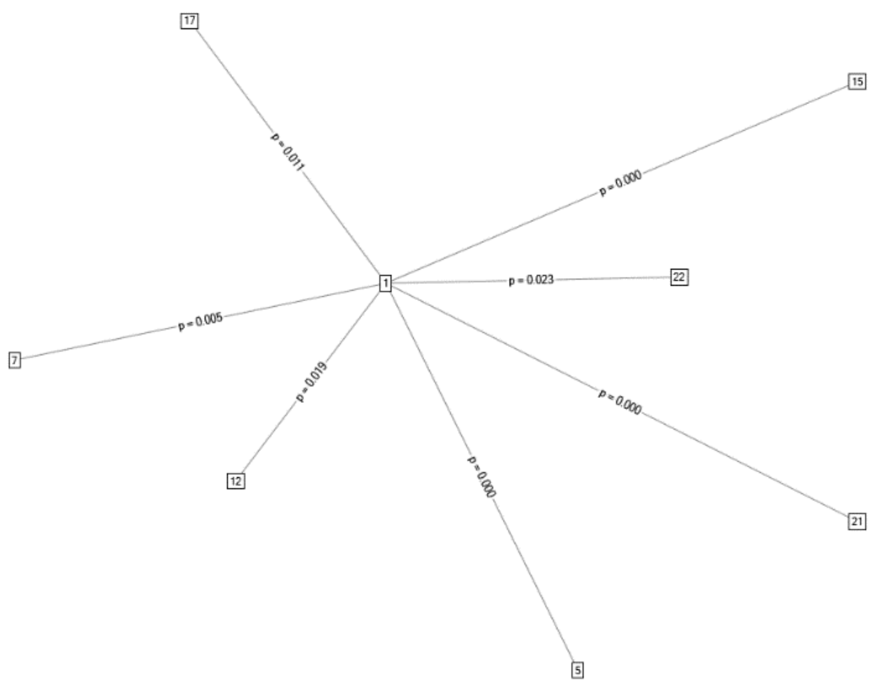


Figure 23. Round Two Pairwise Graph indicates the most important Statement 1 is statistically more important than Statements 5, 7, 12, 15, 17, 21, 22

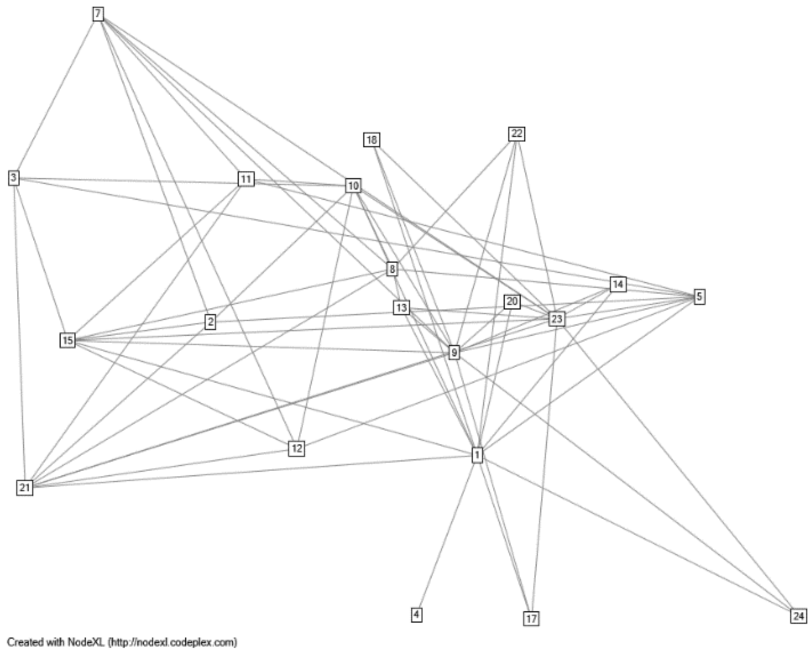


Figure 24. Round Three Pairwise Graph indicates convergence of relevant importance with Statements 5, 7, 15, 21 of the previous set now having six or more statistically significant differences with other Statements

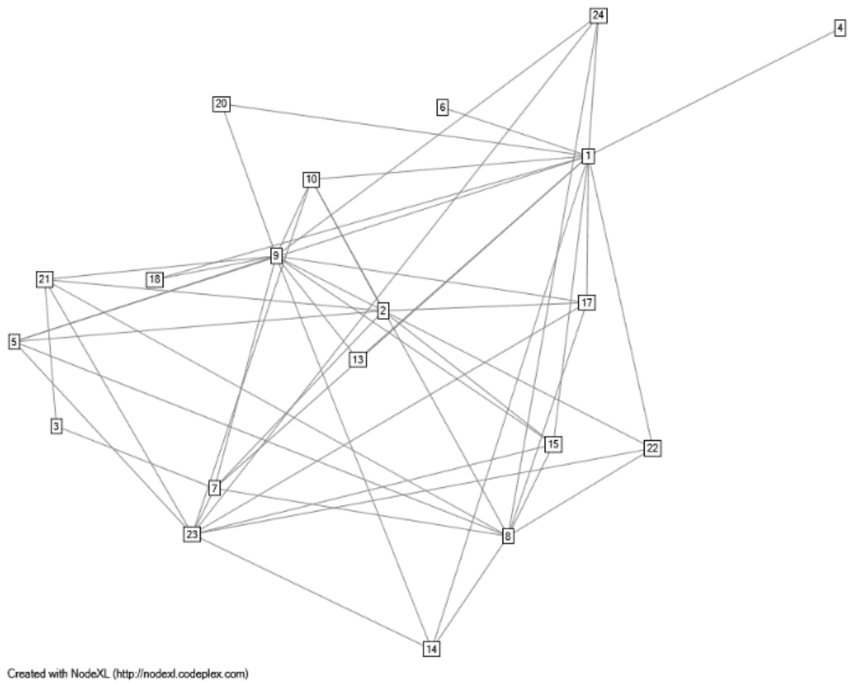


Figure 25. Round Four Pairwise Graph indicates the dominance of Statement 1 but also has inconsistency with relevant importance of other Statements identified in Round Three

Expert Panel Convergence

The reduction in response rate between Rounds Three and Four negatively impacted the utility of Round Four results. Further, *Statement 16* moved up five positions in terms of ranked importance while never obtaining 80% consensus (see

Table 23). *Statement 19* lost consensus while having a higher ranking than *Statements 7* and *10* that gained consensus yet remained near the bottom on rank order of importance. These and other changes may be due to the unequal proportional loss of respondents in the *Manager* and *Developer* categories. The differences between the trajectory toward convergence from Rounds Two and Three and the shifts and inconsistency between ranking and consensus observed in Round Four indicates divergence rather than convergence and undermines the singular use of an 80% consensus measure to throw out or retain a statement. Round Three had the highest percentage of participants, highest consistency between the consensus measure and median ranking, and avoids the fore mention divergence and inconsistency in Round Four. Therefore, Round Three was used as the basis for statement pruning.

The Wilcoxon Paired Signed-Ranks Test, Kruskal-Wallis one-way analysis of variance, and Mann–Whitney pairwise U test aided statement pruning (Fleiss, 1971; Kalaian & Kasim, 2012; Shah & Kalaian, 2009; Viera & Garrett, 2005). These non-parametric approaches address the lack of uniformity of Likert scale measurement between and within expert judgments. Using the Wilcoxon Signed Rank Test alone only eliminates *Statement 16* previously identified by the consensus method. Adding the Mann-Whitney pairwise U test identifies three more Statements – *5*, 7, and *21* - that are consistently in the bottom five with *21* also identified by the consensus measure. Table 25 provides the resulting importance rankings among VAQ statements, excluding statements dropped due to ambivalence or lack of relative importance.

Table 25. Final Ranking of Visual Aesthetic Quality Statements based on Kruskal Wallis mean rankings and Consensus from Round Three of the Delphi

Rank	Statement
1	Realistic sensor representations (*Statement 1*)
2	Accurate terrain surface elevation (*Statement 23*)
3	Inclusion of features providing spatial relationship cues (*Statement 9*)
4	Accurate object position and orientation (*Statement 8*)
5	Addressing simulation fidelity trade-offs upfront with trainees (*Statement 3*)
6	Vegetation density and fidelity (*Statement 11*)
7	Realistic atmospheric, environmental, and weather effects (*Statement 12*)
8	Dense urban terrain and megacities (*Statement 2*)
9	Accurate geospecific locations (*Statement 19*)
10	Removal of scintillation (*Statement 6*)
11	Time of day representation (*Statement 4*)
12	Common Level of Detail (LOD) transition (*Statement 20*)
13	Organic representation of natural terrain objects (*Statement 14*)
14	High-fidelity ground texture/imagery (*Statement 13*)
15	Accurate transportation networks (*Statement 18*)
16	Density and variation of 3D features (*Statement 17*)
17	Consistent texture density and resolution (*Statement 24*)
18	Environmental clutter (*Statement 22*)
19	Consistent color and contrast rendition across the visual scene (*Statement 10*)
20	Common artistic theme and palette across elements in the virtual world (*Statement 7*)

Expert Panel Divergence

The Delphi technique traditionally focuses on achieving convergence and consensus among an expert panel population; however, divergent viewpoints and competing priorities can also be analyzed through Delphi study results where the expert panel is known to include participants from heterogenous populations. Table 26 illustrates community divergence in this study through the rating of *Statement 16* across each survey round. Upon further analysis, the two largest populations, Developers and Managers, consistently rated *Statement 16* as ‘neutral’, whereas the

Trainer population ranked this statement near “Moderately Important’ across later survey rounds. As the Developer and Manager response rates dropped across each survey, the Trainer response rate stayed relatively static and caused *Statement 16* to considerably increase ranking in Round 4. Therefore, it can be inferred that the simulation Trainer population believes that “Using a single geospatial source provider across all simulations/simulators is important to the VAQ of virtual simulation for the training scenario” more so than the simulation Developer or manager community. A similar divergence phenomenon among populations is observed in *Statement 17* between Rounds Three and Four.

Table 26. Comparison of Median Ranking and Response Rate for Statement 16 (single geospatial source provider) across each Delphi survey Round.

Community	**Round 2**		**Round 3**		**Round 4**	
	Median Rating	*Response Rate*	*Median Rating*	*Response Rate*	*Median Rating*	*Response Rate*
Developer	4.63	N/A*	4.11	94.74%	4.23	72.22%
Manager	3.83	N/A*	4.18	91.67%	4.44	81.82%
Trainer	4.00	N/A*	4.86	100.00%	5.33	85.71%

Social psychology concerns: mimicking and groupthink

Critics of the Delphi method claim the process enables Mimicking and promotes Groupthink or can be used to unknowingly shape responses to a certain favorable position (Hsu & Sanford, 2007). "Mirroring" is a process where one individual "mirrors" another. While appropriate in some situations - a student dancer "mirroring" the dance technique of a dance instructor -

"mirroring" can be detrimental in a conversation if one is “mimicking” what is expected of them without standing up for truths they have formulated from the realm of their experience and education. This research was unable to track mimicking. Groupthink is a phenomenon that occurs when the desire for group consensus overrides people's common-sense desire to present alternatives, critique a position, or express an unpopular opinion. To track Groupthink in this survey, two approaches were implemented.

Since the expert panel was comprised of experts across varying simulation backgrounds, a questionnaire was provided in the exit survey to gauge respondent behavior related to the various categories of VAQ statements that were rated in the Delphi surveys. The results of this questionnaire can be seen in Figure 26.

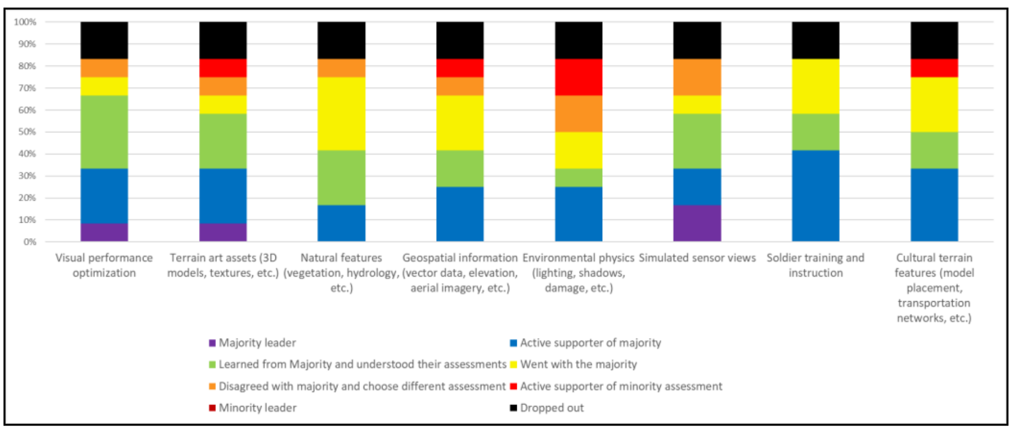


Figure 26. Results of the mirroring versus mimicking questionnaire from the Delphi exit survey.

As indicated by the yellow (“Went with the majority”) responses present, Groupthink appears most prominent in ‘Natural Features’, ‘Geospatial Information’, ‘Soldier Training and Instruction’, and ‘Cultural Terrain Features’.

As part of the exit survey, another Groupthink questionnaire was constructed using common indicators for Groupthink (Manktelow et al., 2018). Respondents were asked to rate their agreement against a series of statements about their behavior as the Delphi surveys progressed. Figure 27 provides the results of this Groupthink questionnaire.

The only potential indicators of Groupthink are seen in statements 4 and 5 indicated on Figure 27. Rationalizing is when members convince themselves that despite evidence to the contrary, the decision or alternative being presented is the best one (Manktelow et al., 2018). A moderate agreement to rationalization could be a slight indicator of Groupthink. Likewise, the illusion of unanimity identified in *Statement 5* may also indicate the slight presence of Groupthink. The illusion of unanimity is often what feeds the Groupthink and causes it to spiral out of control (Manktelow et al., 2018).

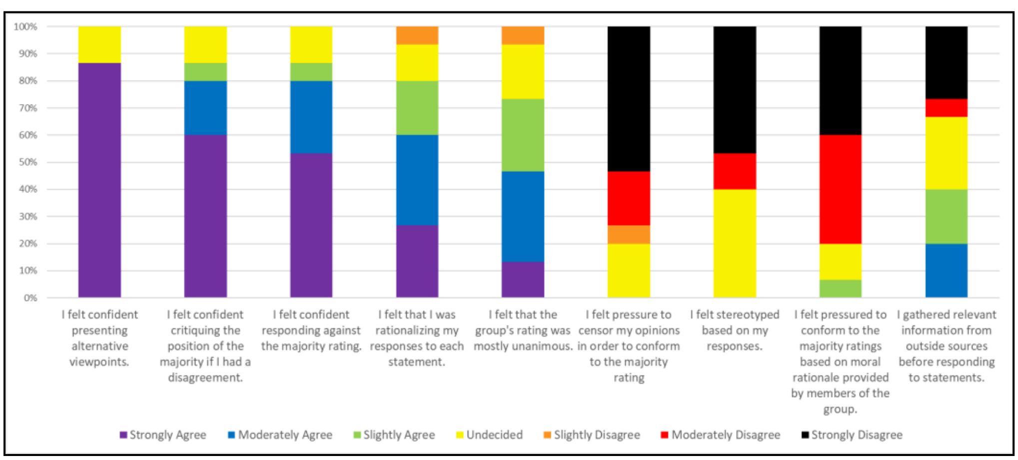


Figure 27. Results of the groupthink questionnaire from the Delphi exit survey.

Phase Two Results: Selected Experimental Design and Factors

A Resolution III screening experiment enables one to screen large numbers of primary VAQ factors while reducing the number of SNE designs considered (Montgomery, 2000). Given limited resources, this research conducted a nine-factor, 2-levels per factor Resolution III screening of 16 SNE designs. In this application, a "factor" corresponds to the presence or absence of a "level" of the VAQ contained in a given SNE design. The JMP™ statistical software package was used to generate the design and validate the power and orthogonality of this design. When grouped in sets of 4 SNE designs, twenty trials enabled all SNE designs to be compared once and only once with each other.

To reach 9 factors of the 20 factors available, the research team considered in turn each top-ranked factor for inclusion in the experiment. Removed from consideration were: (1) *Statement 1 Realistic sensor representations*, as it is specific to a simulated weapon system or vehicle platform rather than the general synthetic environment creation (Baca & Proctor, 2017; Evangelista, Darken, & Jungkunz, 2013; Jacobs, 1999; Kooi & Toet, 2005; Toet, 1998); (2) *Statement 3* as it focuses more on the trainer-trainee relationship and overarching simulation psychological perceptions rather than a discrete factor that can be readily modified within SNE design; (3) *Statement* 6, *Removal of scintillation*, is a well-known quality factor in the SNE community(Andrei, 2012; Persson, 2012; Sellers et al., 2013) involving "Z-fighting"(Piorkowski, Mantiuk, & Siekawa, 2017). This anomaly can normally be addressed with LODs (Statement *20*) or several modern visual rendering engines are able to automatically identify potential Z-fighting issues (Khrnos Group, 2018).

Of the nine factors considered, *Statement 11* was further modified to only focus on "vegetation density" rather than fidelity. 3D model fidelity was addressed through *Statement 2*.

Table 27 lists the final nine factors and levels manipulated to build the SNE designs used in Conjoint Analysis. All have at least 80% consensus between two points on a seven-point Likert scale in their rankings within a single round. The factors associated with the remaining 15 Delphi statements were present and kept constant across each SNE design. Table 27 maps each selected factor to the original Delphi Statement, described in a short one- or two-word DOE factor name for ease of discussion. Table 28 provides a factor and level breakdown of the Resolution III screening experiment chosen for this research across the 16 SNE designs.

Table 27. Synthetic Environment Visual Aesthetic Quality Factors from the Delphi Method to the Conjoint Analysis Screening Design

DOE Factor Name	**Level 1**	**Level 2**	**Simplified VAQ Factor Descriptor**
Elevation	30m DEM	3m DEM	Accurate terrain surface elevation (Statement *23)*
Spatial Cues	Actual Size Trees	2x Size Trees	Inclusion of features providing spatial relationship cues (Statement *9*)
Building Orientation	Actual orientation	10-degree rotate CW	Accurate object orientation (Statement *8*)
Weather Effects	None	Rain/Storm effects	Realistic atmospheric, environmental, and weather effects (Statement *12*)
Vegetation Density	100% density	25% density	Vegetation density (Statement *11*)
Building Complexity	Extruded buildings	Complex building models	Realistic dense urban terrain (Statement *2*)
Geo-specific Locations	Accurate Space Needle Placement	Inaccurate Space Needle Placement	Accurate geospecific locations (Statement *19*)
Time of Day	Day (noon)	Dusk	Time of day representation (Statement *4*)

Levels of Detail	Zero LOD Transitions	Two LOD Transitions	Common Level of Detail (LOD) transition (Statement *20*)

Table 28. Resolution III Fractional Factorial Design used for Synthetic Environment Conjoint Analysis

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
	Elevation	**Spatial Cues**	**Orientation**	**Weather Effects**	**Vegetation Density**	**Building Complexity**	**Geospecific Locations**	**Time of Day**	**LOD**
SNE Design 1	30m DEM	Normal Size Trees	Normal	None	75% reduction	Complex	Inaccurate	Daytime	Two LOD
SNE Design 2	30m DEM	Normal Size Trees	Normal	Rain/Storm effects	Normal Density	Extruded	Accurate	Daytime	No LOD
SNE Design 3	30m DEM	Normal Size Trees	10-degree CW	None	Normal Density	Complex	Inaccurate	Dusk	No LOD
SNE Design 4	30m DEM	Normal Size Trees	10-degree CW	Rain/Storm effects	75% reduction	Extruded	Accurate	Dusk	Two LOD
SNE Design 5	30m DEM	2x Size Trees	Normal	None	75% reduction	Extruded	Inaccurate	Dusk	No LOD
SNE Design 6	30m DEM	2x Size Trees	Normal	Rain/Storm effects	Normal Density	Complex	Accurate	Dusk	Two LOD
SNE Design 7	30m DEM	2x Size Trees	10-degree CW	None	Normal Density	Extruded	Inaccurate	Daytime	Two LOD
SNE Design 8	30m DEM	2x Size Trees	10-degree CW	Rain/Storm effects	75% reduction	Complex	Accurate	Daytime	No LOD
SNE Design 9	3m DEM	Normal Size Trees	Normal	None	75% reduction	Complex	Accurate	Dusk	No LOD
SNE Design 10	3m DEM	Normal Size Trees	Normal	Rain/Storm effects	Normal Density	Extruded	Inaccurate	Dusk	Two LOD
SNE Design 11	3m DEM	Normal Size Trees	10-degree CW	None	Normal Density	Complex	Accurate	Daytime	Two LOD
SNE Design 12	3m DEM	Normal Size Trees	10-degree CW	Rain/Storm effects	75% reduction	Extruded	Inaccurate	Daytime	No LOD
SNE Design 13	3m DEM	2x Size Trees	Normal	None	75% reduction	Extruded	Accurate	Daytime	Two LOD
SNE Design 14	3m DEM	2x Size Trees	Normal	Rain/Storm effects	Normal Density	Complex	Inaccurate	Daytime	No LOD
SNE Design 15	3m DEM	2x Size Trees	10-degree CW	None	Normal Density	Extruded	Accurate	Dusk	No LOD
SNE Design 16	3m DEM	2x Size Trees	10-degree CW	Rain/Storm effects	75% reduction	Complex	Inaccurate	Dusk	Two LOD

Phase Three Results: SNE Generation

This section describes the development of the 16 'Emerald City' SNEs utilized in this research along with the implementation and representation of each of the nine visual aesthetic quality factors and factor levels selected for the screening design. These factors were implemented through a combination of geospatial information data, 3D model content, and runtime effects through manipulation of the selected Image Generator (IF), Unreal 4. Table 29 provides a summary of all software and information used to generate the 16 SNE designs.

Table 29. Software Applications and Plugins used to generate the 16 SNEs for the Conjoint Analysis phase of the research

Application Name	**Application Category**	**Purpose**
QGIS™	Software	Geospatial data visualization and manipulation
Conform™	Software	3D Geospatial and content visualization and SNE generation.
Unreal Engine 4 (UE4)	Software	SNE visualization; image generator software
GOOD FX: Rain	UE4 Plugin	Plugin to procedurally generate dynamic rain and fog weather effects
Open World Demo Collection	UE4 Plugin	Open-source collection of 3D natural models; trees, rocks, etc.
Rain Drops	UE4 Plugin	Plugin to create dynamic water-droplet effects on the first-person view camera.
Ultra Dynamic Sky	UE4 Plugin	Plugin to create dynamic atmospheric effects such as lighting, sun and moon positions, and procedural clouds.

Factor 1: Elevation

This research implemented two levels of terrain surface elevation resolution: Low (Level 1) and High (Level 2). The Low-level resolution terrain surface elevation was obtained from the United States Geological Survey's (USGS) 3D Elevation Program (3DEP). 3DEP data serve as the elevation layer of The National Map, and provide basic elevation information for Earth science studies and mapping applications in the United States (United States Geological Survey, 2020a). This Low-level fidelity resolution elevation data was collected at resolutions of 1 arc-second (approximately 30 meters). The High-level resolution terrain surface elevation was obtained from the United States Geological Survey's (USGS) National Elevation Dataset (NED). NED is an elevation dataset that consists of seamless layers and a high-resolution layer. Each of these layers are composed of the best available raster elevation data of the conterminous United States, Alaska, Hawaii, territorial islands, Mexico and Canada (United States Geological Survey, 2020b). This High-level elevation data was collected at a resolution 1/9 arc-second (approx. 3 meters). Table 30 provides a summary of the Terrain Surface Elevation data implemented within this research.

Table 30. Comparison of digital elevation datasets used for SNE Factor 1 of the screening design

Level	Resolution	Publisher	Dataset(s)
Level 1 – Low Fidelity	1 arc-second (30 meters)	U.S. Geological Survey	• USGS NED 1 arc-second n48w123 1 x 1 degree IMG 2018
Level 2 – High Fidelity	1/9 arc-second (3 meters)	U.S. Geological Survey	• USGS NED ned19_n47x75_w122x25_wa_cederriverbasin_2010 1/9 arc-second 2011 15 x 15 minute IMG • USGS NED ned19_n47x75_w122x25_wa_puget_sound_2000 1/9 arc-second 2012 15 x 15 minute IMG • USGS NED ned19_n47x75_w122x50_wa_puget_sound_2000 1/9 arc-second 2012 15 x 15 minute IMG

Figure 28 and Figure 29 show the subtle differences in these two terrain surface resolutions as visualized through Conform's 3D elevation visualization tool.

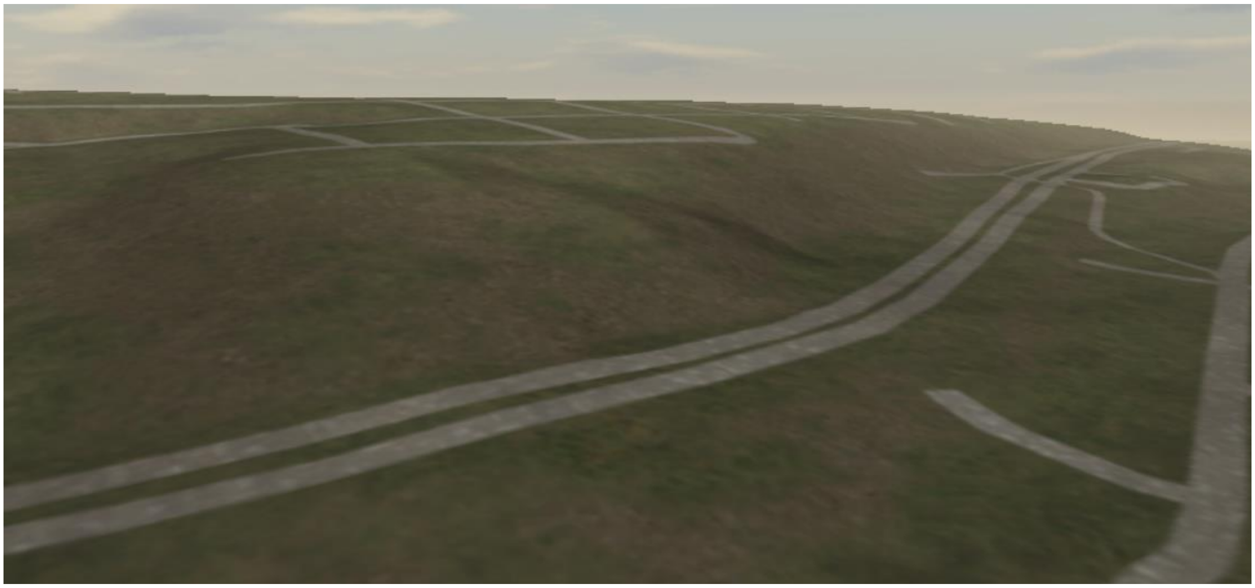


Figure 28. Visualization of low-resolution (30-meter) elevation used for Factor 1 of the SNE design.

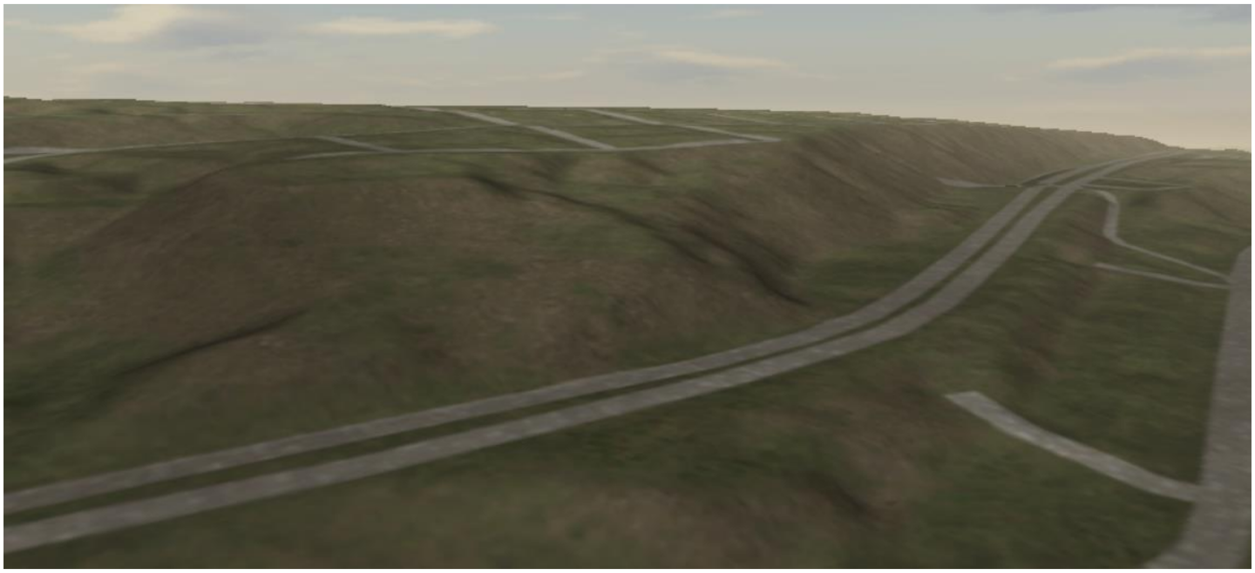


Figure 29. Visualization of high-resolution (3-meter) elevation used for Factor 1 of the SNE design. Notice the increased terrain relief and shadows near the center of the image.

Factor 2: Spatial Cues

In order to replicate the factor of "*Inclusion of features providing spatial relationship cues*", trees were chosen as an everyday object with a known size relative to other natural and manmade features for the selected geographic area. A normal sized tree model (~10-meters tall) was implemented as Level 1 across the entire SNE and a tree model double the normal size (~20-meters tall) was implemented as Level 2. The 3D tree model used was obtained from the UE4 "Open World Demo Collection" of open 3D art assets. Figure 30 and Figure 31 provide a comparison of these two spatial cue levels, respectively. This research implemented the point-feature vector layer of trees used during the OBW 2019 exercise to place trees within the virtual scene.

Figure 30. Level 1 of Factor 2 (Spatial Cues) depicted through an ~10-meter tall tree model. Image rendered in UE4.

Figure 31. Level 2 of Factor 2 (Spatial Cues) depicted through an ~20-meter tall tree model. Image rendered in UE4.

Factor 3: Orientation

Building Orientation (Factor 3) was represented through the manipulation of underlaying geospatial information vector data. The all vector data representing building footprints were rotated exactly 10-degrees clockwise from normal using QGIS™ software to provide the effect of erroneous building placement or alignment. The selection of 10-degrees was made based on trial and error since even a slight misalignment of buildings can be noticeable based on their large scale. Figure 32 provides a comparison of the normal vector data against the modified building footprint orientation. Figure 33 and Figure 34 provide a 3D representation of these two levels, respectively.

Figure 32. Vector data comparison of Factor 3 (Building Orientation) levels. Blue shapes represent the default building orientation (Level 1) and Red shapes represent the incorrectly oriented buildings (Level 2). Data depicted within QGIS software.

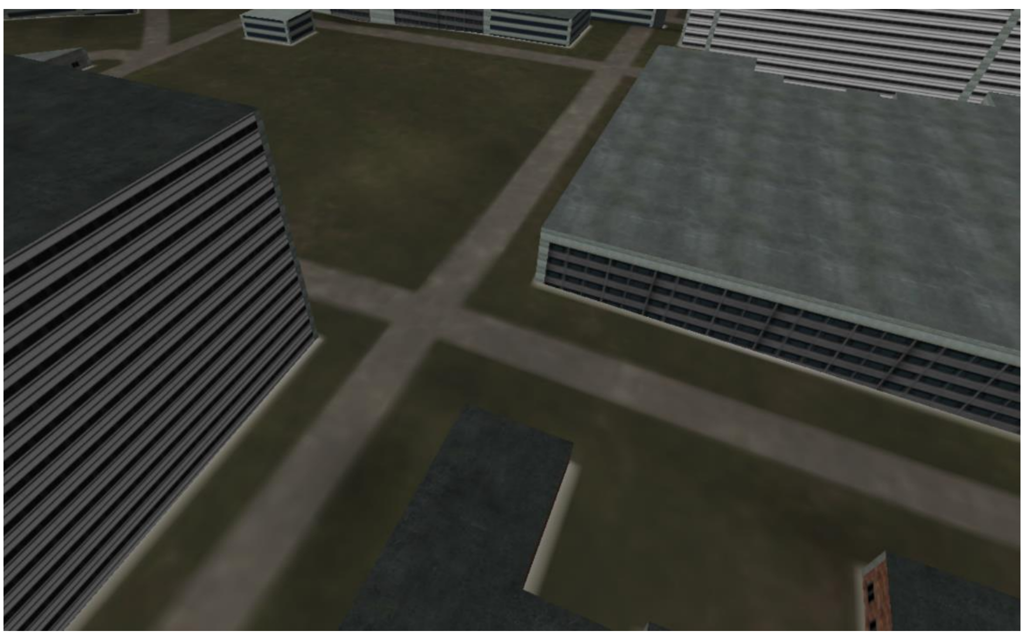


Figure 33. Correct Building Orientation (Level 1) depicted in Conform™ 3D visualization software

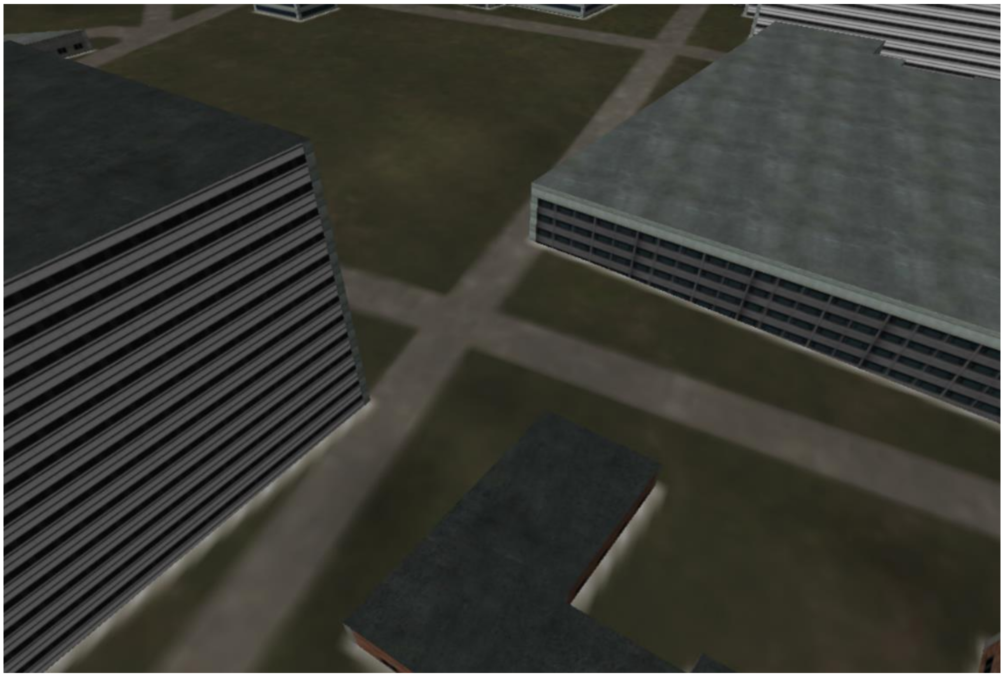


Figure 34. Incorrect Building Orientation (Level 2) depicted in Conform™ 3D visualization software

Factor 4: Weather Effects

Dynamic weather effects were implemented to represent Factor 4 within the Unreal 4 Engine using third-party plugins *GOOD FX: Rain* and *Rain Drops*. Level 1 did not include any weather effects, whereas Level 2 added atmospheric effects to the scene, such as overcast skies, fog, rain, and water droplets on the first-person eye-point within the scene. Level 1 of Factor 4 (Weather Effects) is depicted in Frame ‘A’ of Figure 35 and Level 2 is represented in frame ‘B’ of Figure 35.

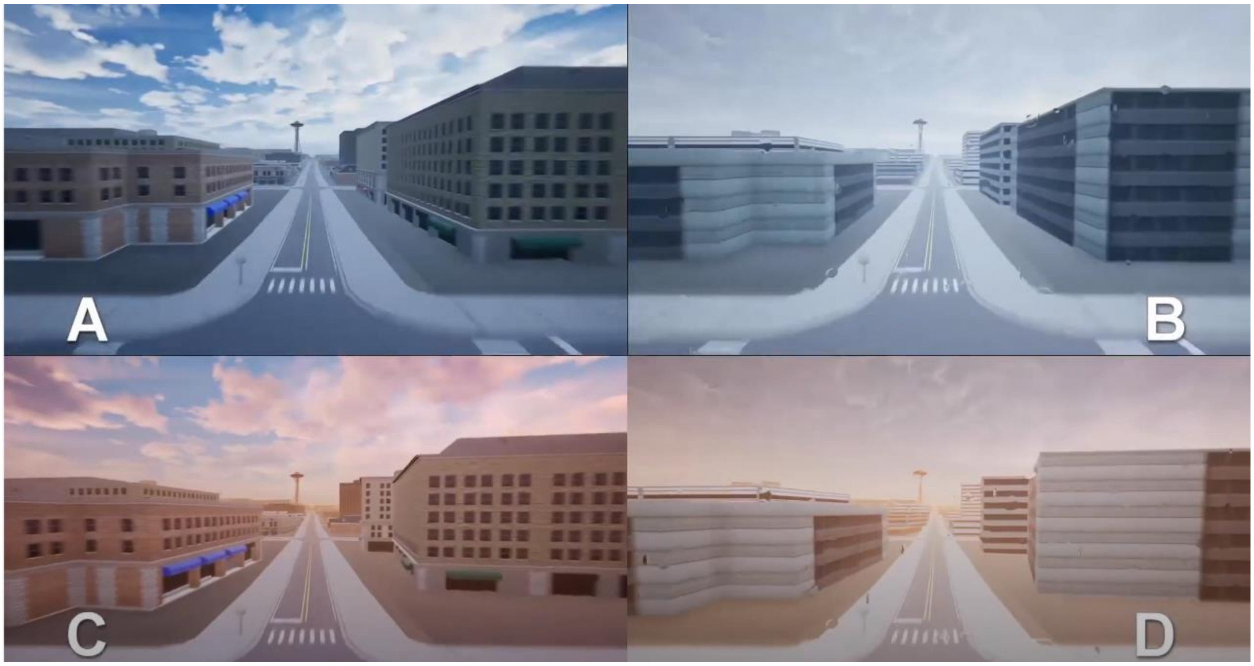


Figure 35. Comparison of Factor 4 (Weather Effects) and Factor 8 (Time of Day). Frame ‘A’ displays a daytime scene with no weather effects. Frame ‘B’ displays a daytime scene with weather effects enabled (note water droplets on the screen). Frame C displays a dusk scene with no weather effects. Frame D displays a dusk scene with weather effects enabled.

Factor 5: Vegetation Density

Vegetation Density was represented in the SNE through two-levels: the standard density used during the OBW 2019 exercise (Level 1) and a 75% reduction in tree density (Level 2). This reduction factor was selected as to visually establish a change while still maintain vegetation content in the scene. Figure 36 illustrates the scale of the tree reduction across the two factor levels.

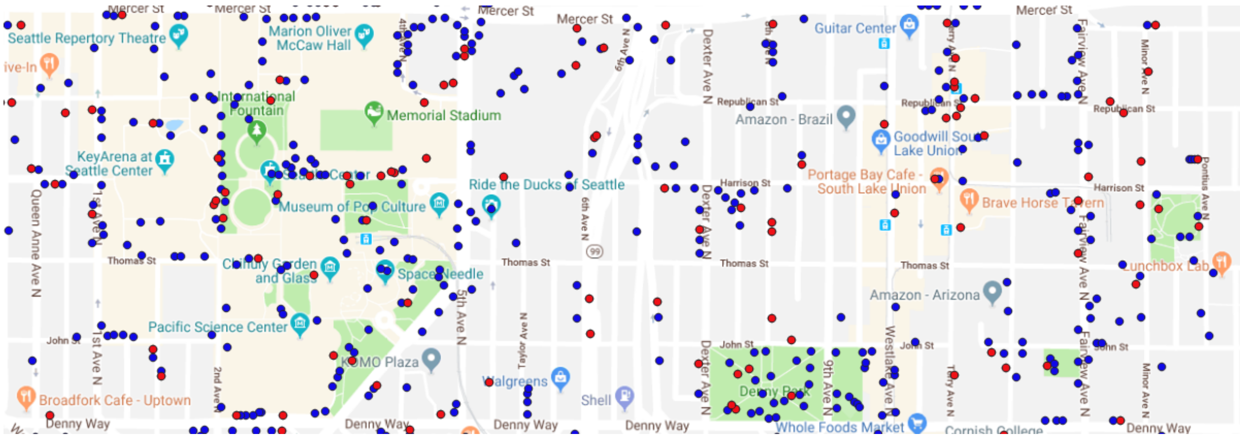


Figure 36. Comparison of Vegetation Density levels as depicted through geospatial vector features. The Blue dots represent the normal density of trees (Level 1) while the Red dots represent the remaining trees after a 75% reduction in geospatial point features (Level 2).

Factor 6: Building Complexity

Building Complexity was represented through two levels within the SNE visual scene: Low-fidelity "Extruded" type buildings (Level 1) and high-fidelity "geotypical" complex building models (Level 2). Extruded-type models are created by creating a simple polygon structure around the geospatial vector footprint at a set height based on source data attribution and applying a generic art texture to the polygon. These extruded models are depicted in Figure 37. Geotypical models are models created specifically to be used in a specific geographic region or location and often contain more complex polygonal structured than extruded models. Figure 38 illustrates these complex geotypical models. This research integrated geotypical models and model textures created for the OBW 2019 exercise.

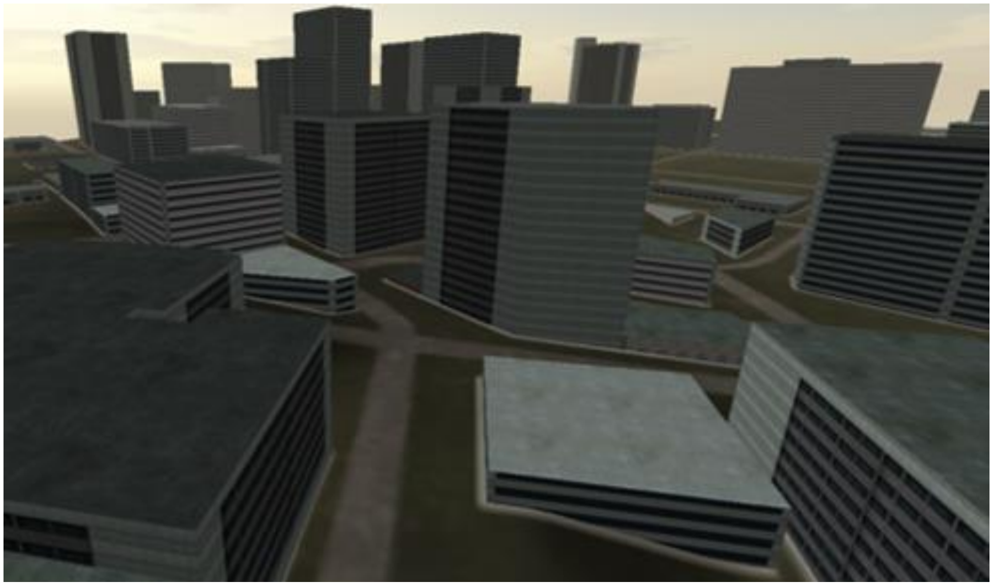


Figure 37. Virtual cityscape of Seattle, WA depicting "extruded" models (Level 1). Note the simple repeating textures and simple building shapes. The visual scene is rendered through the Conform™ geospatial visualization tool.

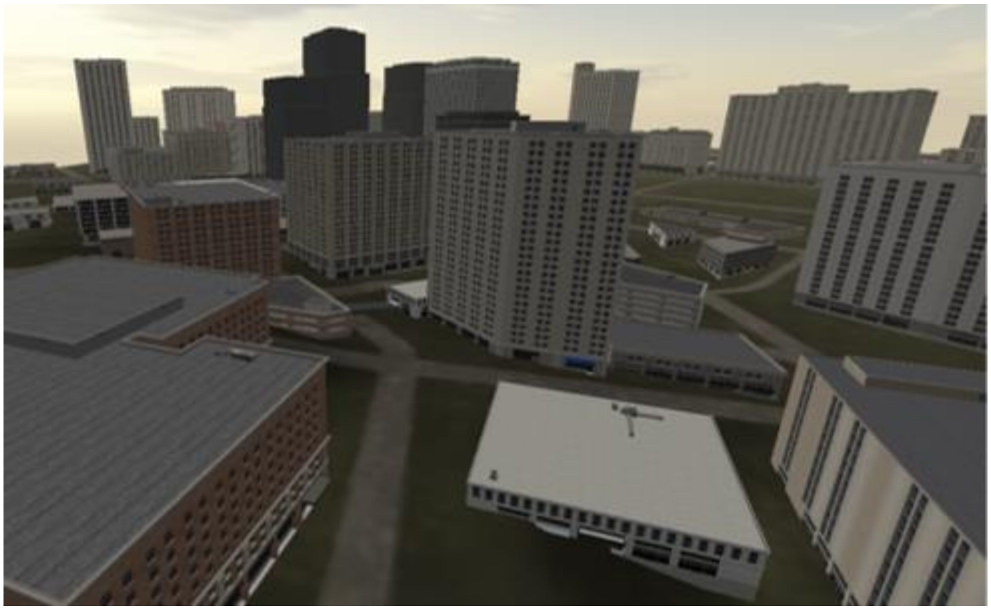


Figure 38. Virtual cityscape of Seattle, WA depicting complex "geotypical" models (Level 2). Note the enhanced geometry to include windows, building overhangs, roof ledges, and other building features. The visual scene is rendered through the Conform™ geospatial visualization tool.

Factor 7: Geospecific Locations

In order to depict geospecific locations, a well-known landmark was selected and altered. The Seattle Space Needle was selected as a geospeifc location for this research sue to the geocarpic location of the selected SNE extents. An unmodified Space Needle was selected as factor Level 1. The correct placement and depiction of the Space Needle is illustrated in Figure 39. Instead of modifying the physical location of the landmark, a fictitious lake was placed under the Space Needle to represent Level 2 – or incorrect geospecific depiction. This incorrect depiction is shown in Figure 40. This research used the Space Needle geospeicifc 3D model developed for the OBW 2019 exercise.

Figure 39. Accurate placement and depiction of the Space Needle (Level 1). Visual scene rendered within Unreal Engine 4.

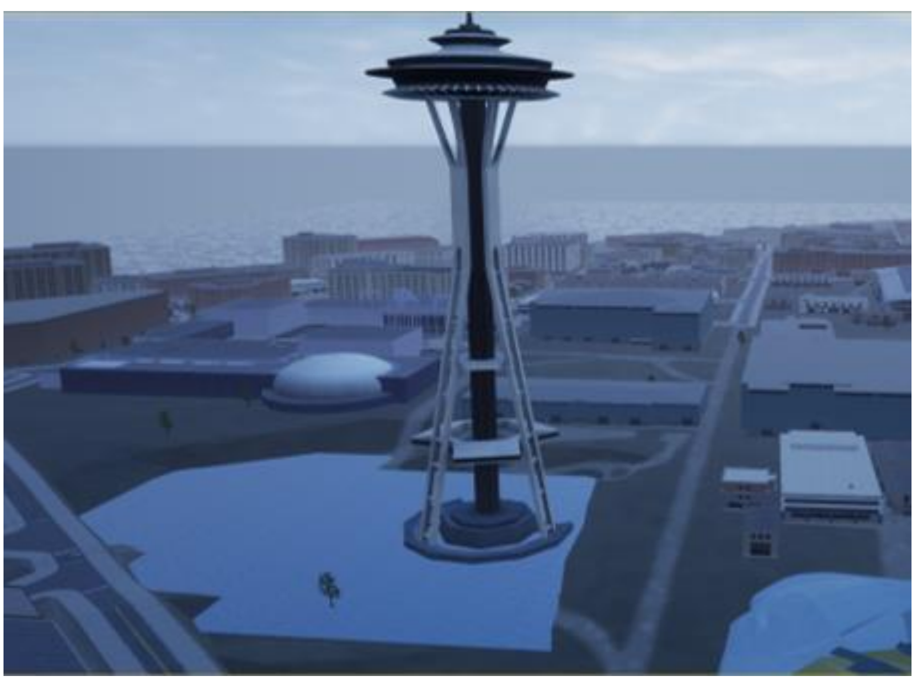


Figure 40. Inaccurate placement and depiction of the Space Needle (Level 2). Note fictitious lake underneath the model. Visual scene rendered within Unreal Engine 4.

Factor 8: Time of Day

Time of day was represented through two distinct levels in the SNE visual scene: Daytime – or noon (Level 1) – and Dusk – or ~5:00pm (Level 2). Time of day was manipulated through the UE4 plugin *Ultra Dynamic Sky* which adjusted the sun position and lighting according to time of day. Level 1 of Factor 8 (Time of Day) is depicted in Frame 'A' of Figure 35 and Level 2 is represented in frame 'C' of Figure 35.

Factor 9: Level of Detail (LOD)

Two representations of Level of Detail (LOD) were implemented across the SNE designs. Level 1 did not include any LOD transitions within the scene, which means all models and textures were rendered at the highest level of details, no matter their distance from the eye point. Level 2 included two LOD transitions with a 1500-meter transition range. This allows for optimized performance since further away features are rendered in less detail than those closest to the observer's eye point. Figure 41 depicts the scene with all content rendered at the highest LOD within the scene and Figure 42 depicts multiple LODs rendered. The lower LOD can be observed in the upper section of Figure 42.

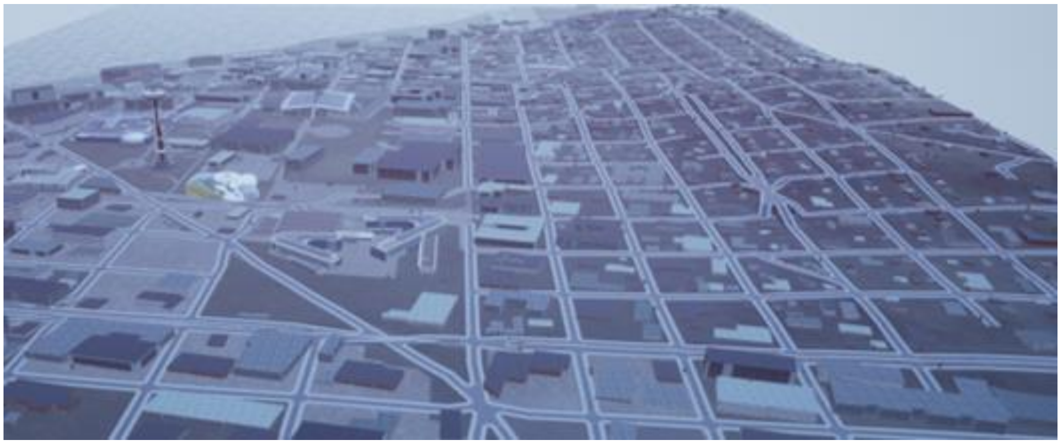


Figure 41. Depiction of visual scene rendered entirely at the highest LOD. Scene rendered within UE4.

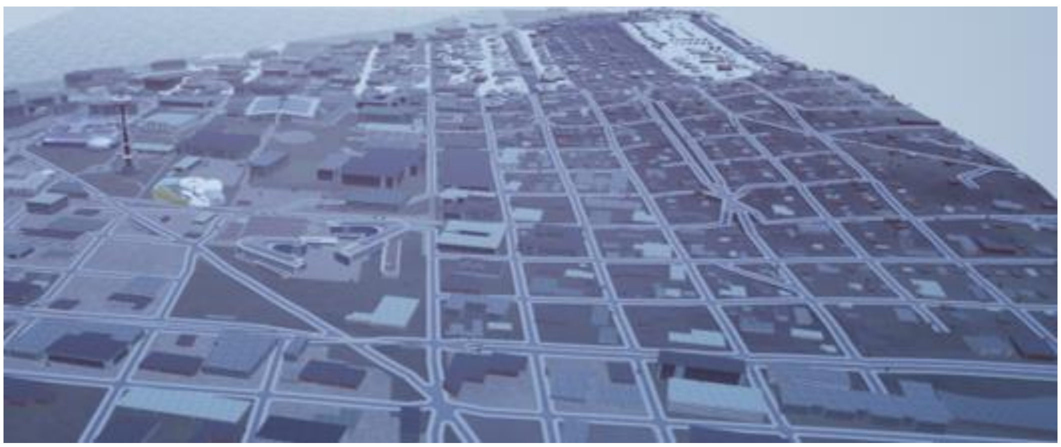


Figure 42. Depiction of multiple LODs within the visual scene. The lower LOD can be observed in the top section of the figure when compared to the above figure. Scene rendered in UE4.

Phase Four Results: End User-based Conjoint Analysis

Conjoint Analysis Design

Graeco-Latin Square Design

Graeco-Latin Square (GLS) designs are particularly useful in reducing the number of trials (combinations) for a complete design (Nielsen & Schmidt, 1990). A GLS was implemented as the basis for the conjoint analysis to provide experimental “blocking” variable in order to streamline participant comparison of multiple SNE designs. Blocking reduces unexplained variability and arranges the SEs in groups (blocks) that are like one another, but not in an order of primary interest to this research. This enables all SEs to be compared once and only once when grouped in sets of 4 SNE designs. Table 31 provides the structure of the GLS conjoint analysis design implemented in this research. This GLS could have been replicated to confirm participant preferences, but the research was concerned with survey drop-out due to doubling the length of the conjoint analysis survey.

Table 31. Conjoint Analysis Survey Design using the combination of a fractional factorial screening design and Graeco-Latin Squares. SNE Designs 1, 2, 3 and 4 are highlighted to show that each SNE combination is only ever compared once.

	Option A	**Option B**	**Option C**	**Option D**
Trial 1	Design 1	Design 2	Design 3	Design 4
Trial 2	Design 5	Design 6	Design 7	Design 8
Trial 3	Design 9	Design 10	Design 11	Design 12
Trial 4	Design 13	Design 14	Design 15	Design 16
Trial 5	Design 1	Design 5	Design 9	Design 13
Trial 6	Design 2	Design 6	Design 10	Design 14
Trial 7	Design 3	Design 7	Design 11	Design 15
Trial 8	Design 4	Design 8	Design 12	Design 16
Trial 9	Design 1	Design 6	Design 11	Design 16
Trial 10	Design 2	Design 5	Design 12	Design 15

	Option A	Option B	Option C	Option D
Trial 11	Design 3	Design 8	Design 9	Design 14
Trial 12	Design 4	Design 7	Design 10	Design 13
Trial 13	Design 1	Design 7	Design 12	Design 14
Trial 14	Design 2	Design 8	Design 11	Design 13
Trial 15	Design 3	Design 5	Design 10	Design 16
Trial 16	Design 4	Design 6	Design 9	Design 15
Trial 17	Design 1	Design 8	Design 10	Design 15
Trial 18	Design 3	Design 6	Design 12	Design 13
Trial 19	Design 2	Design 7	Design 9	Design 16
Trial 20	Design 4	Design 5	Design 11	Design 14

Electronic Survey Design

As with the Delphi Method, the Conjoint Analysis phase of this research was conducted through Survey Monkey™ in a web-based environment. Unlike the Delphi survey, participant information was completely anonymous except for limited demographic information collected for participant roles (i.e. civilian, military) and training domain experience (i.e. air, ground, etc.) Whereas the Delphi survey targeted SNE experts, the Conjoint Analysis survey targeted virtual simulation users, operators, and trainers. This population represents the target end-user for SNEs.

In this survey, each participant was presented with a series of four (4) synthetic environments depicted through a correlated flythrough video presented as an embedded YouTube link[1]. Each SNE was depicted from a ground and air perspective. Subsequently each participant simultaneously viewed the same four SNE designs from the perspective of a drone flying through the respective SNE at approximately 380-meters above ground level (AGL). Both ground and aerial perspectives traveled through the SNE at a simulated speed of approximately 40 miles per hour. Each video could be viewed as many times as desired. Along with each set of SEs, the participant was presented with a training task to evaluate the SNE against. This training task was the same across all SNE comparisons and participants - *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).* Based on the presented SEs and training task, the participants were asked to rank each SNE relative to the other choices presented based on its visual aesthetic quality to meet that training task. Figure 43 and Figure 44 illustrate the layout of this survey and the synthetic environments presented. A more detailed view of this survey can be found in Appendix C. In accordance with the GLS design in Table 31, participants continued this rating across the 20 trials – or survey questions.

[1] Example Synthetic Environment video for the Conjoint Analysis survey: https://youtu.be/GJy8iBMQuRs

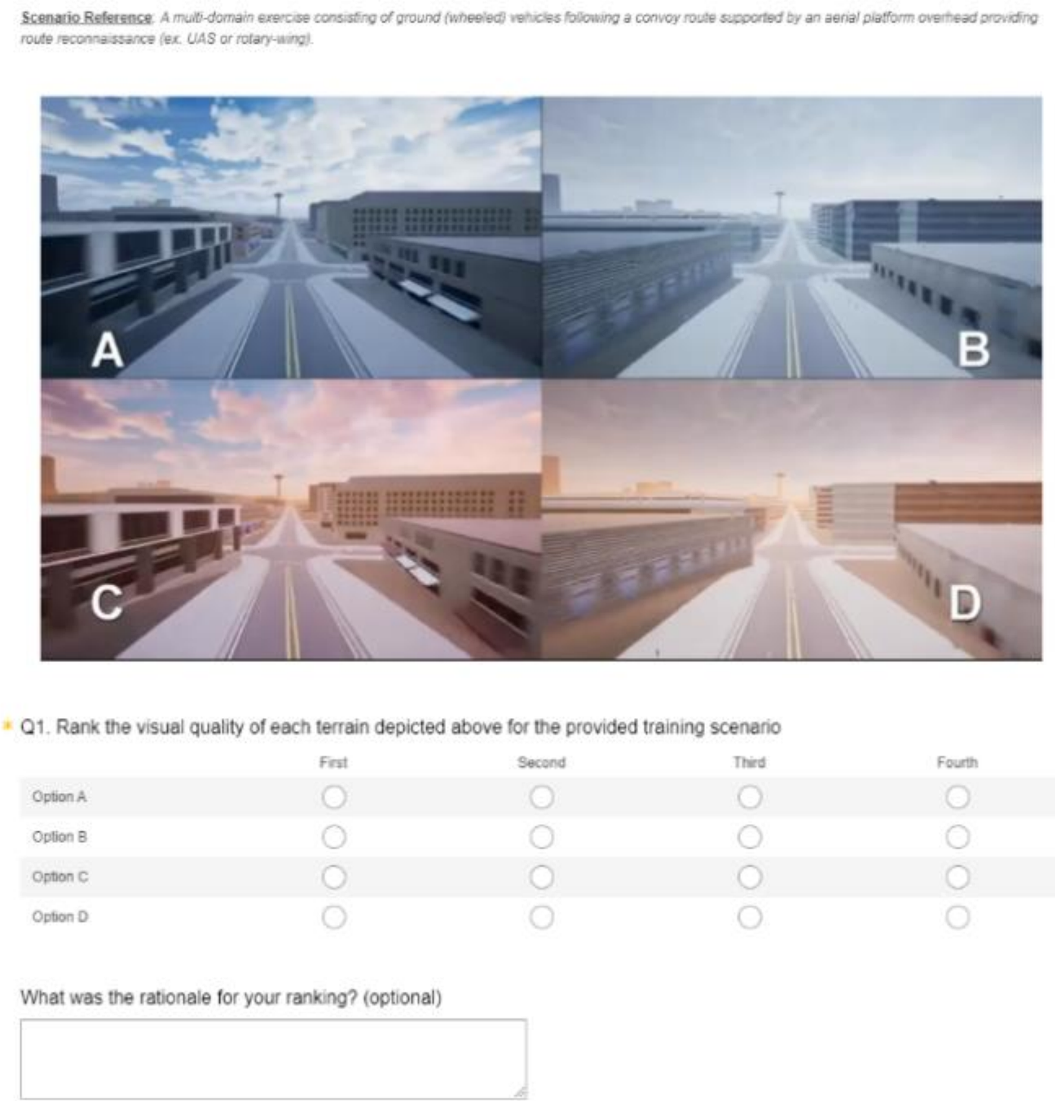


Figure 43. Layout of a single "trial" for the Conjoint Analysis survey. Option A represents SNE Design 1, Option B is SNE Design 2, Option C is SNE Design 3, and Option D is SNE Design 4. The video depicts the SNE from the ground perspective.

Figure 44. This image depicts the same SNE as seen in Figure 1 but further in the video, presenting an aerial view of the SNE.

Conjoint Analysis Survey Execution

The research team invited 83[2] military and civilian virtual simulation expert users and trainers to participant. In accordance with UCF IRB approved protocols, invited experts were advised of their rights as participants in the experiment. Participant information was completely anonymous except for limited demographic information collected for participant roles (i.e. civilian, military) and training domain experience (i.e. air, ground, etc.).

[2] This number is higher as some participants sent the survey to their peers as well. The researcher does not have direct insight into these second-order survey participants.

The total number of parameters in this experiment is nine (16-8+1). Based on the rule of thumb ratio for Conjoint Analysis participation, this would suggest a minimum of 45 respondents would be necessary to produce stable results. A total of 58 experts started the assessments, with a total of 48 completing the assessments in their entirety, resulting in an 83% completion rate, but fulfilling the rule of thumb guidelines. Of the 48 completed assessments, there were 17 responses from active-duty military experts, 30 civilian experts, and 1 academic expert. These demographics were further broken down by specific simulation domain. Of the 48 participants, 11 had expertise in Aviation simulation, 2 in Dismount simulations, 7 in ground simulations, and 28 had experience in combined-arms simulations. The assessment web site was open for a total of 85 days. The assessment took an average of 48 minutes to complete.

Conjoint Analysis Data Analysis

Empirical Validation of Main Effects only through Conjoint Analysis Screening Design

For consistency, this research implements a 0.2 level of statistical significance for Conjoint Analysis based on the previously implemented 80% level of consensus during the Delphi survey

Rankings for each SNE design option were assigned a numeric value (First = 4, Second = 3, Third = 2, and Fourth = 1). At the conclusion of the assessment, all values for each option across the twenty comparison trials in the survey were summed to a total value for each. These total values serve as the response variable (Y) for this experiment and represent the "SNE Visual Quality Score". Since this VAQ response variable is based on ranked, or ordinal, there is disagreement within literature whether the sum of this data can be considered continuous for use in parametric statistical procedures (Glass, Peckham, & Sanders, 1972; Jamieson, 2004; D. R.

Johnson & Creech, 1983; Norman, 2010; Sullivan & Artino, 2013; Zumbo & Zimmerman, 1993). Since this research is primarily focused on identifying the overall significance of each factor over quantifying a parameter estimation of the factors, the VAQ response variable was considered to be continuous, and regression coefficients for each factor were only used to estimate the magnitude of the relationship of each factor to the SNE VAQ response.

The Resolution III design used to construct this SNE quality experiment means that main effects are not confounded (aliased) with any other main effects, but main effects are aliased with two-factor interactions and two-factor interactions may be aliased with each other. Therefore, the significance of each main effect factor affecting SNE VAQ can be readily estimated through a standard least squares model.

Table 32 presents the effects summary, importance ranking, and statistical significance of these factors through a standard least-squares regression model.

Table 32. Summary of Fit, Analysis of Variance, and Parameter Estimates of the regression model considering only main effects for Combined Expert Responses to the SNE Conjoint Analysis. Bold numbers indicate statistically significant findings at the 0.2

Combined Expert Response (Main Effects Only)			
Sample Size	*n = 48*		
R-squared	*0.971532*		
F-Ratio	*22.7511*		
Prob > F P-value	**0.0006**		
Factor Importance	**Factor**	**Regression Coefficient**	**P-Value**
1	Building Orientation [L2: 10-degree CW]	-88.25	**0.00006**
2	Weather Effects [L1: None]	59	**0.00053**
3	Building Complexity [L2: Complex]	58.625	**0.00055**
4	Time of Day [L1: Day]	27.625	**0.01991**
5	Elevation [L1: 30m]	14.75	**0.14398**
6	Vegetation Density [L1: 100%]	9	0.34489
7	Spatial Cues [L2: 2x Size Trees]	-2.875	0.75446
8	Geo-specific Locations [L1: Accurate]	-2.5	0.78542
9	Levels of Detail [L1: Zero LODs]	-1.875	0.83797

The R-squared value indicates that 97.1% of the variance in SNE VAQ can be predicted from the nine factors investigated in the model. Further, the p-value associated with the F-ratio indicates that these factors can reliably predict the VAQ of the SNE as represented by the expert rankings.

Empirical Validation of Main Effects and 2-Factor Interaction Conjoint Analysis Screening Design using Lenth's Pseudo-Standard Error (PSE) Method & Monte-Carlo simulation

Factor interactions or the combination of specific visual factors by developers may have non-deliberate affects on SNE VAQ (Kang et al., 2015). The JMP™ Screening Platform can be used identify the main effects and interactions in a regression model. Combining Lenth's Pseudo-Standard Error (PSE) Method with 10,000 Monte-Carlo simulations on the JMP™ Screening Platform estimated p-values for un-replicated screening designs. Lenth's method constructs an estimate of the residual standard error using effects that appear to be inactive ("Lenth's Analysis of Unreplicated Factorial Experiments," 2011; Lenth, 1989; Proust, 2018). Table 33 provides the results from the Screening Platform based on a Lenth's PSE of 4.3125. Candidate Main Effects and Interactions for the model are highlighted in blue. While previously identified main effects did not change in terms of significance, the interaction between *Building Orientation* and *Building Complexity* is estimated to be statistically significant by this method. This 2-factor interaction is also aliased with several other 2-factor interactions, but Lenth's PSE successfully estimates this interaction as significant based on the magnitude of the associated main effects through multiple simulations. The significance of the interaction is validated through an analysis of key words used by survey participants in the feedback comments of each survey option ranking shown in Figure 45. Terms associated with "Building Orientation" and "Building Complexity" were used significantly more than terms associated with the aliased 2-factor interactions.

Table 33. Analysis of Main Effects and 2-Factor Interactions through the JMP Screening platform using Lenth's Pseudo-Standard Error (PSE). Bold numbers indicate statistically significant findings at the 0.2 level. Orange numbers indicate a p-value<0.01

Term	Contrast	Lenth t-Ratio	Individual p-Value
Building Orientation	-88.2500	-20.46	**<.0001**
Weather Effects	59.0000	13.68	**<.0001**
Building Complexity	58.6250	13.59	**<.0001**
Time of Day	27.6250	6.41	**0.0011**
Elevation	14.7500	3.42	**0.0127**
Vegetation Density	9.0000	2.09	**0.0590**

Term	Contrast	Lenth t-Ratio	Individual p-Value
Spatial Cues	-2.8750	-0.67	0.5315
Geo-specific Locations	-2.5000	-0.58	0.5966
Levels of Detail	-1.8750	-0.43	0.6890
Building Orientation*Building Complexity†	-19.8750	-4.61	**0.0041**
Building Orientation*Time of Day‡	-7.1250	-1.65	**0.1127**
Building Complexity*Time of Day	-3.5000	-0.81	0.3913
Building Orientation*Elevation	-0.0000	-0.00	1.0000
Building Complexity*Elevation	1.6250	0.38	0.7269
Time of Day*Elevation	1.3750	0.32	0.7643

†Aliased with other 2-factor interactions of Vegetation Density*Spatial Cues, Time of Day*Geospecifc Locations, and Elevation*Levels of Detail

‡Aliased with other 2-factor interactions of Elevation*Spatial Cues, Building Complexity*Geospecifc Locations, and Vegetation Density*Levels of Detail

Figure 45. Word cloud of key words used throughout the comments of the combined SNE conjoint analysis survey responses. The numbers to the right of each word indicate frequency that each word was used.

The results from the JMP™ Screening Platform analysis can be used to construct a more robust regression model using the estimated main effects and 2-factor integrations, while discarding inactive effects and interactions. Table 34 provides analysis and statistical significance of factors in this enhanced regression model.

Table 34. Summary of Fit, Analysis of Variance, and Parameter Estimates of the regression model considering main effects AND 2-factor interactions for Combined Expert Responses to the SNE Conjoint Analysis. Bold numbers indicate statistically significant finding

Combined Expert Response (Main Effects and 2-Factor Interactions)			
Sample Size	*n = 48*		
R-squared	*0.997857*		
F-Ratio	*407.5191*		
Prob > F P-value	*<.0001*		
Factor Importance	**Factor**	**Regression Coefficient**	**P-Value**
1	Building Orientation [L2: 10-degree CW]	-88.25	**<.0001**
2	Weather Effects [L1: None]	59	**<.0001**
3	Building Complexity [L2: Complex]	58.625	**<.0001**
4	Time of Day [L1: Day]	27.625	**0.00003**
5	Building Orientation [L2: 10-degree CW] * Building Complexity [L2: Complex]	19.875	**0.0003**
6	Elevation [L1: 30m]	14.75	**0.00198**
7	Vegetation Density [L1: 100%]	9	**0.02499**
8	Building Orientation [L2: 10-degree CW] * Time of Day [L1: Day]	7.125	**0.0152**

Comparing Synthetic Environment User Groups

The JMP™ Screening Platform was also used to identify the importance of factors and 2-factor interactions among various SNE user groups based on responses collected from the demographic questionnaire provided with the conjoint analysis survey. Table 35 provides a comparison of SNE VAQ preferences across military and civilian simulation users and trainers who participated in the survey. Table 36 provides a further breakdown of user preferences based on simulation domain background and experience of each participant. The results indicate that most user groups placed the highest importance on “Building Orientation”, but Military and Ground Simulation users identified “Weather Effects” as the most important factor. “Building Orientation”, “Building Complexity”, and “Weather Effects” were each consistently ranked the three highest importance factors among all user groups in this study. The factors “Levels of Detail” and Geospecifc Locations were not identified as significantly statistical factors across any of the user populations. The factor “Spatial Cues” was only identified as statistically important by the Dismount Simulation user group.

Table 35. Comparison of Synthetic Environment VAQ factor importance rankings between military and civilian users queried through the conjoint analysis survey.

Factor Importance	**Military Simulation Users**			**Civilian Simulation Users**	
	n = 17			*n = 30*	
	R-squared: 0.986671			*R-squared: 0.97014*	
	Factor	**P-Value**		**Factor**	**P-Value**
1	Weather Effects	<.0001		Building Orientation	<.0001
2	Building Orientation	<.0001		Building Complexity	0.00001
3	Building Complexity	<.0001		Weather Effects	0.00003
4	Time of Day	0.00028		Time of Day	0.0031
5	Elevation	0.0065		Building Orientation*Building Complexity	0.00963
6	Building Orientation*Building Complexity	0.01849			

Table 36. Comparison of Synthetic Environment VAQ factor importance rankings between users of various simulation-domain backgrounds queried through the conjoint analysis survey.

Factor Importance	**Aviation Simulation Users**		**Dismount Simulation Users**		**Ground Simulation Users**		**Multi-Domain Simulation Users**	
	n = 11		*n = 2*		*n = 7*		*n = 28*	
	R-squared: 0.975832		*R-squared: 0.936433*		*R-squared: 0.94277*		*R-squared: 0.991191*	
	Factor	**P-Value**	**Factor**	**P-Value**	**Factor**	**P-Value**	**Factor**	**P-Value**
1	Building Orientation	<.0001	Building Orientation	0.00003	Weather Effects	<.0001	Building Orientation	<.0001
2	Building Complexity	<.0001	Weather Effects	0.00007	Building Orientation	0.00001	Building Complexity	<.0001
3	Weather Effects	0.00001	Vegetation Density	0.0006	Building Complexity	0.00368	Weather Effects	<.0001
4	Elevation	0.00072	Building Complexity	0.0006	Time of Day	0.01483	Time of Day	0.00002
5	Time of Day	0.00635	Spatial Cues	0.06501			Building Orientation*Building Complexity	0.00008
6	Building Orientation*Building Complexity	0.00824						

Comparison of Findings Against Historical Data

The VAQ importance results of both the Delphi Method and Conjoint Analysis can be compared to SNE quality findings from traditional VV&A methods. Over 900 Discrepancy Reports (DRs) captured from three different US Army SE Core user assessments were analyzed by sorting DRs into common categories based on their descriptions. The three SE Core user assessment analyzed focused on terrain from: 1) Fort Hood, TX, 2) Joint Base Lewis-McChord, WA, and 3) Germany. Each category was than assigned an overall importance ranking based on the total number of DRs included within the category divided by the average severity rating of all DRs within that category. The SE Core DR Reports were used with permission from the US Army but cannot be reproduced within this dissertation due to controlled data distribution restrictions. Table 37 provides the overall rankings of these categories developed through the historical analysis and compares this with VAQ factor importance rankings identified through the Delphi Method and Conjoint Analysis. A word cloud derived from the raw comments provided from users/SMEs across these three SE Core SNE user assessments is provided in Figure 46. Inspection of the frequency of common terms within this world cloud validates the rankings of this historical VV&A data.

Table 37. Comparison of historical SNE VV&A VAQ Findings from the US Army SE Core program with the implemented research methodology. Shaded cells indicate matching ranks across methodologies.

Rank	Historical Analysis (SE Core)	Delphi Method w/ KW Mean Rank Order	Conjoint Analysis with Graeco-Latin Square Comparisons (Main Effects + 2-Factor Interactions)
1	Geospecifc Features*	Elevation Fidelity	Building Orientation
2	Transportation Networks	Spatial Cues	Weather Effects
3	Vegetation Fidelity	Building Orientation	Building Complexity
4	Surface Textures	Weather Effects	Time of Day
5	Elevation Fidelity	Vegetation Fidelity	Building Orientation*Building Complexity
6	Building Complexity	Building Complexity	Elevation
7	Airfield Fidelity	Geospecifc Features	Vegetation Fidelity
8	Building Orientation	Time of Day	Building Orientation*Time of Day
9	Visual Anomalies	Levels of Detail	-
10	Urban Areas	-	-
11	Levels of Detail	-	-
12	Spatial Cues	-	-

*Possible outlier due to SME/user familiarity with the geographic regions and locations under evaluation.

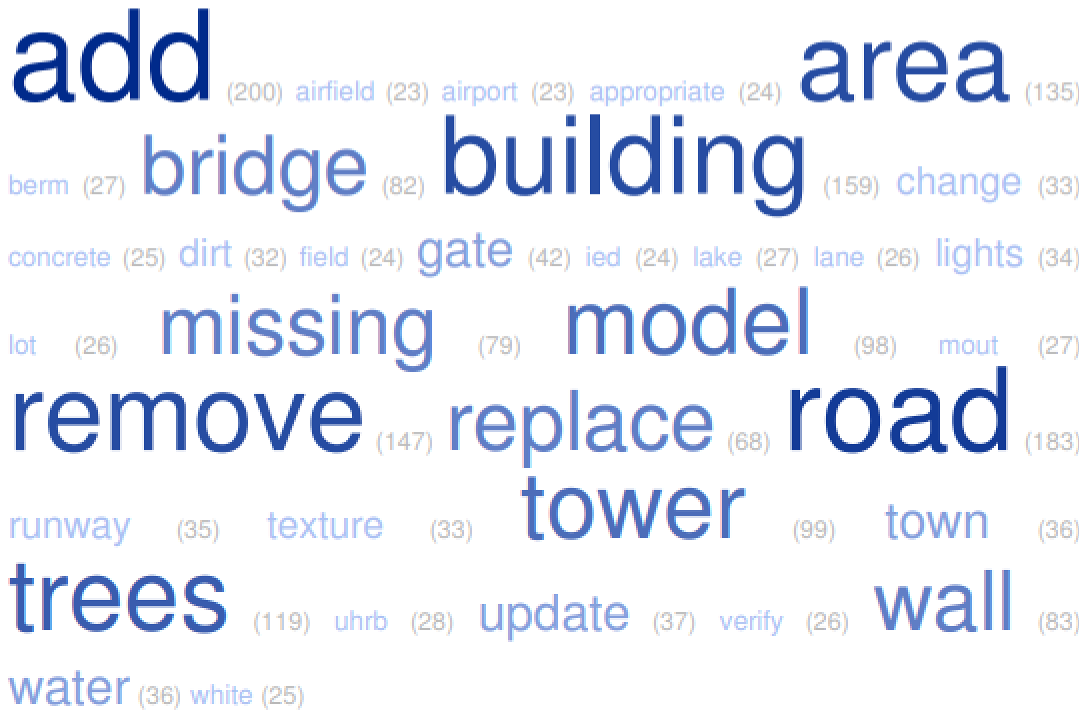


Figure 46. Word cloud analysis of 931 SE Core discrepancy reports across three user assessments. The number in parentheses indicates the frequency of the word used across all discrepancy reports.

CHAPTER 5: DISCUSSIONS, LIMITATIONS, FUTURE RESEARCH, AND CONCLUSIONS

Traditional techniques used for verification, validation, and accreditation (VV&A) of Synthetic Natural Environments for DoD applications are time consuming, subjective, and often costly. Each Synthetic Natural Environment (SNE) varies widely in appearance and use case affecting greatly the significance of the representation of common visual elements – elevation posts, imagery, features, vegetation, buildings, roads, etc. Based on past user accreditation decisions, these elements often also vary greatly between SNE's in terms of visual aesthetic quality (VAQ). Early identification in the SNE development process of which elements or factors, particularly VAQ factors, contributes greatest to a given SNE quality and reduces VV&A issues downstream and informs future development. The question emerges, assuming an "interoperability" case perspective taken previously by Purdy & Goldiez (1995), what is the power of forecasting SNE visual aesthetic quality factor in predicting user preferences for individual virtual environment use-cases and the simulation interoperability community at large?

The research findings suggest that using an online Delphi Method enables identification of subjective factors affecting the SNE VAQ can be accomplished early in the life cycle. Further, the research indicates that SNE VAQ may be quantified for specific military training tasks. This requires the careful application of committee-based expert and user methodologies, within the bounds and limitations of this research, specifically related to the fitness for use of training tasks and geographic extents. Limitations with the proposed technique, as well as recommendations for additional research are provided to further refine the parameters associated with these subjective factors to increase the efficiency and application of the proposed approach.

This chapter provides critical analysis of the findings identified through the implemented research techniques from Chapter Four and seeks to answer the primary research questions and hypotheses associated with subjective factors affecting SNE Visual Aesthetic Quality (VAQ).

The multiple exploratory hypotheses posited at the start of this research are addressed through discussion of the data analysis results.

Findings relevant to each specific hypothesis is discussed in detail below.

> *Hypotheses 1:* For a given synthetic natural environment representation across dissimilar simulators within a multi-domain simulation exercise, there exists a correlated set of expertly accepted and user validated primary VAQ factors that affect overall realism and training utility in the virtual domain.

The null hypothesis may be largely accepted as discussed below and based on the results presented in Table 33. The expert panel queried through the Delphi Method was able to identify a set of VAQ factors. The discussion of validation of the various levels of importance of VAQ factors to acceptance of the SNE by users is left largely to the Conjoint Analysis section below.

Data Analysis Discussion

Each phase of this research provided critical insights into the understanding of subjective VAQ factors affecting a SNE deployment. Each of these phases contributed to a larger framework to effectively identify and verify these factors in a real-word application. The discussion below addresses each of these phases in turn: Delphi Method, Conjoint Analysis, and a Comparison of Research Techniques. Further, this discussion addresses SNE VAQ factors for implementation within Virtual Training, addresses research questions and hypotheses posited at the start of the research, and finally provides lessons learned, limitations, and future investigation topics related to this research.

Delphi Method Discussion

The research of using an online implementation of the Delphi method had at least three observations – participant attrition, VAQ factor identification implications, and ranking vs the Delphi Method - worthy of future VAQ researcher awareness.

First unlike a Delphi Study involving a team of individuals committed throughout the project, implementation of the Delphi Methods online and using volunteers had negative consequences in terms of participant attrition. As identified in Table 22, Round Four of the Delphi saw more than a 22% decrease in expert panel response from Round Three. The exit survey indicated that participation may have been more consistent between rounds if there was less time between rounds, if statements were less ambiguous, and if they were reminded more frequently to complete the survey. This is consistent with Keeney *et al. who also* indicates online questionnaire-based research is often plagued with low response rates. Keeney points out that the Delphi technique asks much more of respondents than a simple survey, therefore potential for low response rates increases drastically (Keeney et al., 2006). Another phenomenon identified in the literature that may contribute to response drop off was the perspective that a member felt he or she was not truly "partners" in the study and subsequently lost interest in the topic (Keeney et al., 2006). The use of median may contribute to that perspective. The median is an excellent choice for estimating the importance of a VAQ factor due to its stability against outliers, as seen in this study where most statements maintained a median of 5 or higher for most rounds. In contrast, as a feedback mechanism to expert panel members, the median may not provide enough information as the sole statistic, especially if multiple statements within the Delphi are ranked similarly important. When respondents see this “rating” unchanging between rounds, they may simply lose interest and assume that their input is unnecessary.

In terms of factor identification, the results confirm that the Delphi method is a valuable technique to identifying the importance of factors affecting the VAQ of SNEs, with minimal interference imposed by social psychology dynamics. The Delphi Method also allows for the identification between rounds of convergence, stability, and/or divergence of factor priorities among a heterogeneous populations that makeup the expert panel.

Finally, while the Delphi method was ultimately able to derive the order of importance among SNE visual factors through expert ratings, it may be more effective for experts to directly rank-order statements instead of rating them and deriving a ranking as is done in the Delphi method. The action of rating statements resulted in many statements sharing similar median scores throughout each survey, whereas ranking may yield a more definite level of importance among statements; however, ranking statements would most likely yield longer survey times among experts which could negatively impact response rate between surveys.

Hypotheses 2: Conjoint analysis will improve the understanding of the significance and power of identified factors and preferences

There is sufficient evidence to accept the null hypothesis based on data analysis of Conjoint Analysis results below. Table 34 provide statistical evidence of the overall importance of each identified VAQ factor main effects (p-value = 0.0006) and main effects plus two-factor interactions (p-value < 0.0001), respectively. Further, Table 35 Table 36 confirm the ability for Conjoint Analysis to assess VAQ preferences across disparate user populations.

Conjoint Analysis Discussion

Main-Effect Only Screening with GLS Blocking

The application of Conjoint Analysis, widely successful and commercially used technique to enable empirical identification of product salient features, was employed to determine the VAQ of sixteen different SNE designs. A Resolution III Screening and GLS blocking experiment with twenty experimental ground level and twenty aerial trials enabled experts to identify main effects for SNE drive through and fly over scenarios. This technique was able to estimate the statistical significance of all main effect factors for the given scenario and explains 97.1% of the variance in SNE VAQ for VV&A purposes, but was also useful in identifying differences between the theoretical Delphi rankings and the empirical Conjoint Analysis rankings. While in the short term, the differences found between the techniques undermine VV&A confidences, in the long term, awareness of these significant differences raises the question, what other empirical scenarios must be modeled and evaluated to explain the difference in outcome? For example, an empirical three-mph road clearing operations in an asymmetric battlespace may likely identify significantly different VAQ factors than seen in either of the 30-mph aerial operation (Baca & Proctor, 2017). An additional weakness in Resolution III Screening and GLS blocking experiment is that two-factor interactions were not considered.

Based on linear-regression model of SNE VAQ factors, considering main-effects only, the following may be inferred:

- Improper building orientation may negatively affect the VAQ of the SNE (p=0.00006)
- Absence of weather affects may positively affect the VAQ of the SNE (p=0.00053)
- Increased 3D model complexity of buildings may positively affect the VAQ of the SNE (p=0.00055)
- Daytime visual scenes may positively affect the VAQ of the SNE (p=0.01991)
- Lower-fidelity Elevation may positively affect the VAQ of the SNE (p=0.14398)

Main-Effect and 2-Factor Interaction Screening and GLS Blocking

Conjoint Analysis of Main Effect and 2-Factor Interaction Screening using Lenth's PSE Method and Monte-Carlo simulations helps identify for the given scenario the significance of two-factor interactions and explains 99.7% of the variance in SNE VAQ. Had the additional technique of investigating main effects and two-factor interactions not been performed, the interactions of "Building Orientation" and "Building Complexity", and "Building Orientation" and "Time of Day" would not have been discovered. Identification of these interactions are valuable in that manipulation of individual factors by SNE developers may unknowingly cause overall negative quality impacts. For example, an improper building orientation coupled with complex 3D building models may increase the VAQ of the SNE. This relationship is counter intuitive and not supported by the literature nor experience. Further experimentation is required to better understand this relationship. One possible explanation is that the representation of these factors within this experiment are directly coupled within the SNE visual scene. Both factors were applied simultaneously to the 3D building models within the SNE designs. Participants may simply favor the complex 3D models and chose to ignore the irregular orientation of the associated building within that scene. This hypothesis is further supported by analysis of participant comments in Figure 45. Words associated with building complexity such as “detail” and “texture” appear more frequently than words associated with building orientation (i.e. “placement” and “accurate”).

Based on the selected linear-regression model provided in Table 34, the following may be inferred using the regression coefficients as estimates of the magnitude of the relationships to SNE VAQ:

- Building orientation may be the most important factor affecting SNE VAQ and improper orientation will negatively affect the VAQ of the SNE ($p<0.0001$)
- Absence of weather affects will positively affect the VAQ of the SNE ($p<0.0001$)
- Increased 3D building model complexity positively affects the VAQ of the SNE ($p<0.0001$)
- Daytime visual scenes may positively affect the VAQ of the SNE ($p=0.00003$)
- The combination of building orientation and 3D building model complexity is significant[3] ($p=0.0003$)
- The combination of building orientation and Time of Day is significant ($p=0.0152$)
- Lower-fidelity elevation models positively affect the VAQ of the SNE ($p=0.00198$)
- Increased vegetation density positively affects the VAQ of the SNE ($p=0.02499$)

[3] Additional experiments needed to better understand this interaction. Literature and researcher expertise does not support the findings that the combination of improper building orientation and 3D building model complexity would have a positive affect on visual aesthetic quality of the synthetic environment. See “Discussion” section for more analysis of this finding.

Comparison of Research Techniques

This research applied four techniques to aid identification of VAQ factors for a common SNE: (1) Open-Ended Survey; (2) Delphi Method with KW Mean Rank Order and Consensus; (3) a user-based Empirical Validation of Visual Aesthetic Quality Factors through Main-Effect Screening Conjoint Analysis; and 4) a user-based Empirical Validation of Main Effect and 2-Factor Interaction Screening Conjoint Analysis using Lenth's PSE Method and Monte-Carlo simulations. Table 38 compares the importance rankings of identified factors across each analysis technique.

Table 38. Comparison of Synthetic Environment Visual Aesthetic Quality Factor Rankings Across Implemented Methodologies.

	Synthetic Environment Visual Aesthetic Quality Factor Rankings			
	Technique 1: Open-Ended Survey with implicit Word Cloud Rank Order	**Technique 2: Delphi w/ KW Mean Rank Order and Consensus**	**Technique 3: Validation through Conjoint Analysis with Graeco-Latin Square Comparisons (Main Effects only)**	**Technique 4: Validation through Conjoint Analysis with Graeco-Latin Square Comparisons (Main Effects + 2-Factor Interactions)*** **
Importance	**Factor***	**Factor****	**Factor**	**Factor**
1	Feature Density	Elevation	Building Orientation	Building Orientation
2	Building Complexity	Spatial Cues	Weather Effects	Weather Effects
3	Building Orientation	Building Orientation	Building Complexity	Building Complexity
4	Ground Texture Fidelity	Weather Effects	Time of Day	Time of Day
5	Consistent use of Texture	Vegetation Density	Elevation	Building Orientation*Building Complexity
6	Environmental Clutter	Building Complexity	Vegetation Density	Elevation
7	Natural Terrain Features	Geo-specific Locations	Spatial Cues	Vegetation Density
8	Elevation	Time of Day	Geo-specific Locations	Building Orientation*Time of Day
9	Weather Effects	Levels of Detail	Levels of Detail	

*Factor importance measured through qualitative analysis of open-ended responses. See Figure 1 and Table 4. Only the top-nine factors included for comparison with other techniques.
Factor description condensed to match Conjoint Analysis factors. See **Error! Reference source not found.. Excludes "Realistic sensor representations" (Statement *1* from Delphi)

***Exclude inactive main effect factors of Spatial Cues, Geospecific Locations, and Levels of Detail.

Table 38 reveals that none of the four techniques yielded the same VAQ factor rank order. Further neither of the empirical validations - Conjoint Analysis with Graeco-Latin Square Comparison rank order nor Lenth's PSE Method and Monte-Carlo simulation rank order - confirmed the theoretical Open-Ended Survey or Delphi Method rank orders. The lack of user-based empirical validation of expert derived factor prioritization portends SNE users will likely not be satisfied, at least initially, with SNE designs based on equal expert selected factors.

A weakness of both Conjoint Analysis techniques using a Resolution III screening design is that it is scenario dependent and requires follow-on experimentation with higher resolution screening designs to better refine main effect and two-factor interaction parameters. Since all two-factor interactions are aliased in a Resolution III design, critical interactions may be missed or mistaken as a significant main effect. This may be particularly true for VV&A of SNE designed for Mission Rehearsal or post Mission Scenario analysis (M. D. Proctor & Paulo, 1996). In this research, a Resolution III screening design identified significant results that indicate lower-resolution terrain elevation (a main effect factor) as having a positive effect on overall SNE VAQ. The most logical explanation for this finding is related to the geometric representation of 3D buildings models used within the SNE designs. The two types of 3D building models used in the Conjoint Analysis SNE designs to represent "Building Complexity" were extruded and procedurally generated models. Extruded buildings are formed by simply creating and raising a basic polygon shape based on the underlaying "footprint" of the geospatial building data, creating a continuous geometry from the lowest elevation point of the building footprint.

Procedurally Generated models are created through a precise algorithm that matches the geospatial footprint and attributes with known architectural designs of the geographic location to create a set of extremely detailed 3D models. These models are then placed as geospatial "point features" in the terrain at the center of the corresponding geospatial footprint. Depending upon the magnitude of the elevation changes surrounding the complex models, they could be perceived as "floating" since parts of the model may hang of the edge of a hill or steep incline. Therefore, a higher fidelity elevation such as a 3-meter Digital Elevation Model (DEM) may intensify the visual floating anomaly more than a lower fidelity elevation model, such as the 30-meter DEM. This anomaly is depicted in Figure 47. A solution to this would be to create a "skirt" for each procedurally generated model which provides the building with extra geometry under the building to extend through the terrain mesh and prevent the perception of hovering or floating off the ground. Therefore, there is most likely a two-factor interaction between Building Complexity and Terrain elevation, but the screening design lacks sufficient resolution to properly identify.

Simple “Extruded” 3D Building Model

Complex “Procedurally Generated” 3D Building Model

Figure 47. Depiction of the "floating" building visual anomaly with complex 3D building models. Images depict a high-resolution 3-meter Digital Elevation Model (DEM).

Addressing Synthetic Natural Environment (SNE) Visual Aesthetic Quality (VAQ) Factors for Virtual Training

The results of the Conjoint Analysis suggest an importance of factors related to dense urban areas and megacities as “Building Orientation” and “Building Complexity” were highly-ranked, with users favoring properly oriented buildings and more complex 3D building models. These findings support the US Army’s current trajectory of increased virtual training within dense urban environments (Alderton, 2019). This research therefore posits that implementers of the US Army’s OWT paradigm should focus on the VAQ of urban environments, specifically urban buildings, when designing common SNE for use across multiple training domains and disparate simulators.

The Conjoint Analysis yielded other interesting results seen through the identification of the significant 2-factor interaction between “Building Orientation” and “Building Complexity”. The results indicate that improper building orientation coupled with complex 3D building models will increase the VAQ of the SNE. This relationship is not supported by literature nor experience and further experimentation is required to better understand this relationship. One possible explanation is that the representation of these factors within this experiment are directly coupled within the SNE visual scene. Both factors were applied simultaneously to the 3D building models within the SNE designs. Participants may simply favor the complex 3D models and chose to ignore the irregular orientation of the associated building within that scene. This hypothesis is further supported by analysis of participant comments in Figure 45. Words associated with building complexity such as “detail” and “texture” appear more frequently than words associated with building orientation (i.e. “placement” and “accurate”).

Another interesting finding from the Conjoint Analysis is the positive effect of lower-resolution terrain elevation on SNE VAQ. The most logical explanation for this finding is related to the geometric representation of 3D buildings models used within the SNE designs. This anomaly is discussed in the above section and highlighted in Figure 47.

The results of this research also confirm and extend the findings of Purdy and Goldiez in their early SNE simulation study. In their complex scene study, they found that the ‘Size’ and ‘Position’ of 3D features were the most important visual factors within a SNE (Purdy & Goldiez, 1995). The research results also partially confirm this conclusion through our identification of accurate object position/orientation and 3D model complexity as statistically significant visual quality factors affecting SNEs. The factor 'features providing spatial relationship cues' was determined to be important based on the expert panel consensus received through the Delphi, but was proven to be inactive in a real-world application through Conjoint Analysis within the context of the identified training task and geographic region.

A shortcoming, and divergence from the Purdy and Goldiez findings, of the of the proposed VAQ factor identification techniques is the failure to identify incorrect Geo-specific Locations when present within the visual scene. The VAQ factor "Geo-specific Locations" was never identified as a significant factor across any applied technique which means participants accepted the incorrect placement of the Seattle Space Needle in the middle of an artificial lake as depicted within the Conjoint Analysis trials. The historical analysis of SNE user evaluations, presented in Table 37, identifies Geo-specific Features as the most important factor identified by users of SNE that have been generated from traditional development processes, therefore additional techniques may need to be applied outside those presented in this research to better identify deviations from real-world, or Geo-specific, feature content within the SNE.

<u>Discussion of Additional Research Hypotheses</u>

Hypotheses 3A: A Delphi study using a panel of experts will forecast the same VAQ factor considerations as Conjoint Analysis of end user assessments.

Based on the results presented in Table 33, there is sufficient evidence to partially accept the null hypothesis. The expert panel queried through the Delphi Method was able to identify a set of VAQ factors that were each verified at various level of importance by SNE users through

Conjoint Analysis. The null hypothesis cannot be completely accepted since the VAQ importance rankings differed across the Delphi Method (Techniques 1 and 2) and Conjoint Analysis (Techniques 3 and 4), nor did the Delphi Method successfully predict the two-factor interactions discovered through Conjoint Analysis of the screening design.

Hypotheses 3B & 3C: Data mining of historical SNE issue reports will identify the same level of importance of VAQ factors as users reviewing SNE representations through a Conjoint Analysis AND Delphi panel expert forecasts.

There is sufficient evidence to reject the null hypothesis within the bounds of this research. The results presented in Table 37 identify similar VAQ factors across historical findings with those discovered through this research; however, there are differences in importance rankings. This may be due to a multitude of factors, such as user or expert backgrounds, familiarity with specific geographic location or area of interest being assessed, and/or the maturity of SNE at the time of assessment. Identification of the rationale for these importance differences is out of scope for this research but may be investigated in future research activities.

Hypotheses 4: Quality Function Deployment (QFD) can be utilized to abstract the correlated set of expertly accepted and user validated primary SNE VAQ factors into a series of SNE generation process improvements to influence a new SNE VV&A paradigm.

Additional research is required before a conclusion can be made against this hypothesis. Instead, several recommendations on immediate QFD applications of research results and future research concepts are provided.

The findings of this research could be used conjointly with QFD to validate assumptions made about the Voice of the Customer (VOC), attribute importance levels, and attribute correlations during the Product Definition and Product Development phases of QFD. Figure 48 provides an example QFD House of Quality (HOQ) implementation of the product design phase for SNE Core terrain development activities with the One World Terrain (OWT) activity provided as a competitor. In this example vignette, the QFD product team did not identify a correlation between Feature Orientation and Feature Complexity or Feature Orientation and Environmental Lighting, as identified by the red circles in the roof of the HOQ in Figure 48. Had this example QFD product planning team leveraged the proposed research methodology and findings as historical background they would have identified a positive correlation between Building Orientation and Building Complexity and the negative correlation between Building Orientation and Time of Day. Further, they may have forgone the selected relationship weightings chosen for Levels of Detail, due to this research's findings on the low importance of LODs.

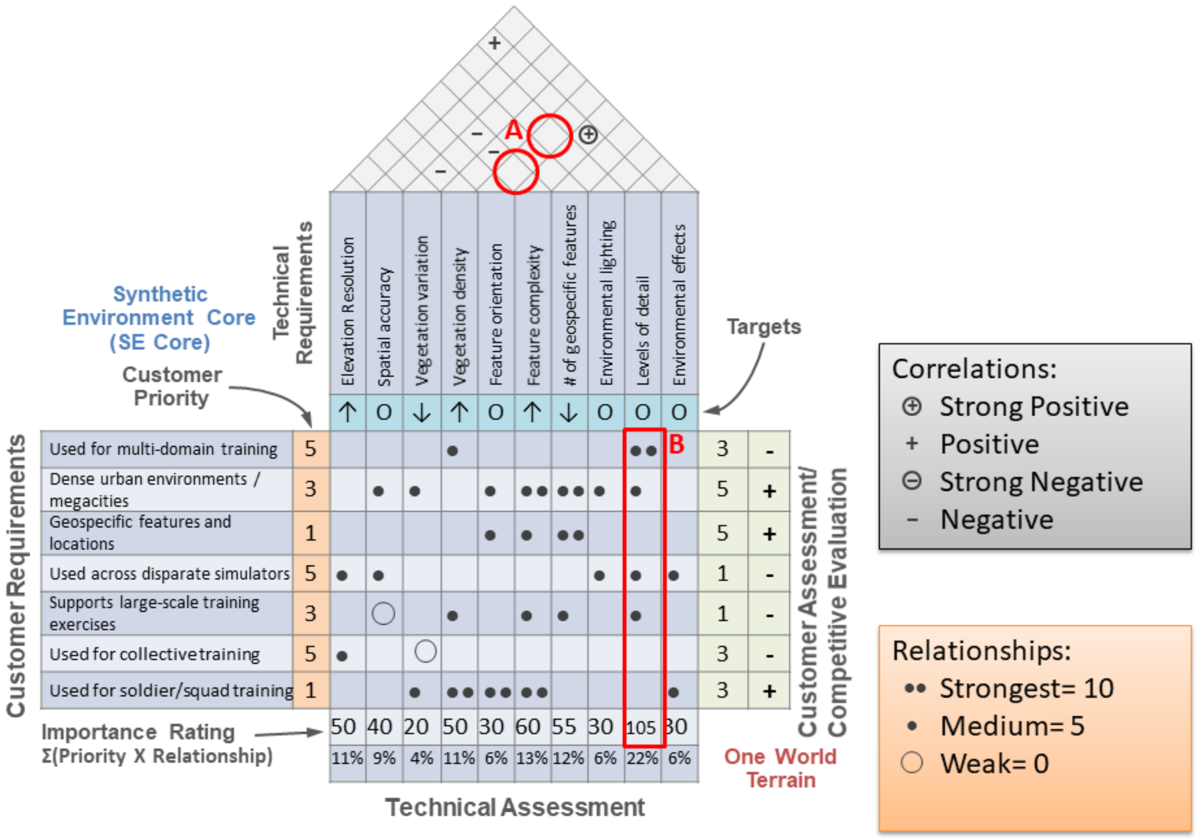


Figure 48. Example QFD House of Quality implementation for the US Army SNE Core program. The section highlighted in 'A' and 'B' show example incorrect assumptions that may have been revised based on findings highlighted through this research.

These results can be further harnessed as reference to reinforce the QFD House of Quality paradigm as depicted in Figure 49. Reinforcing the central superstructure of the QFD House of Quality through the implemented Research Methodology. First, the “foundation” of the HOQ could be reinforced through the identification of technical factor parameters validated through the conjoint analysis screening designs (i.e. factor levels) applied across this research. These technical parameters can be further refined through multiple, subsequent screening experiments of higher resolution (i.e. Resolution III and IV). Second, historical data obtained from this research methodology can be used to reinforce the “core” of the HOQ – Inter-relationships. This would provide validation for QFD practitioners to determine how consumers will judge the SNE’s value for a particular training task and may provide a basis for cross-functional agreement and departmental “buy in” on user requirements for the SNE among managers, developers, and other critical stakeholders when weighing relationships between customer and technical requirements.

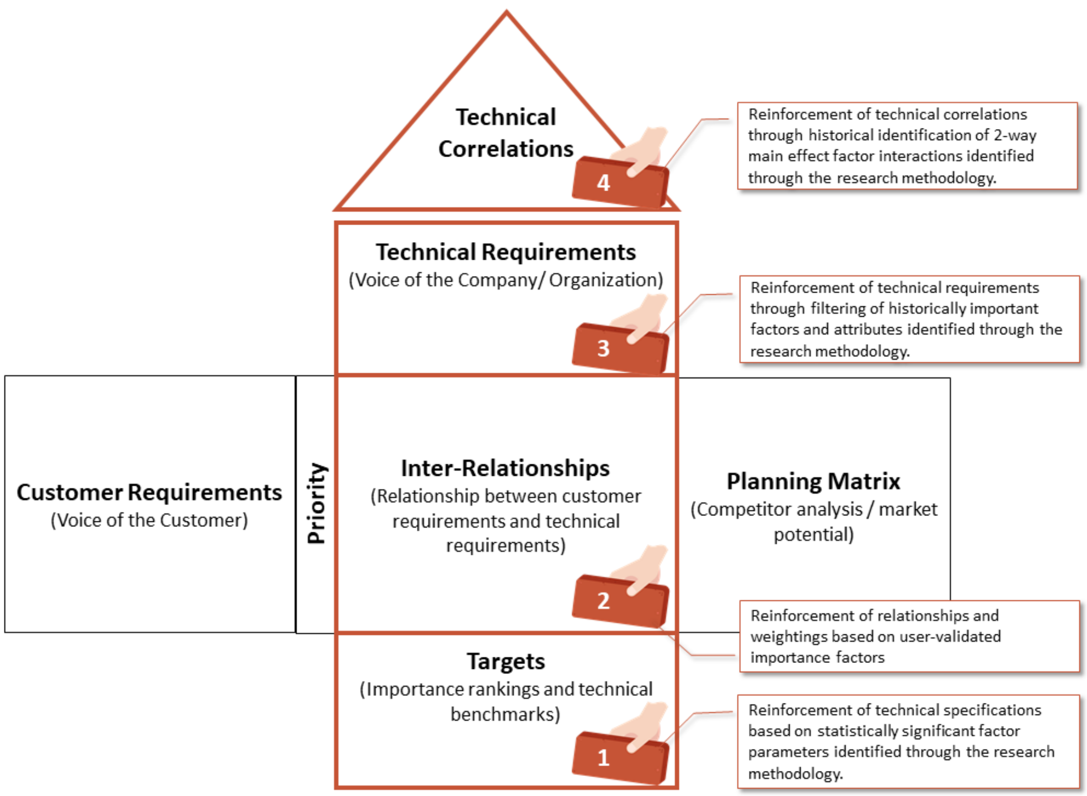


Figure 49. Reinforcing the central superstructure of the QFD House of Quality through the implemented Research Methodology

Next, the “attic” of the HOQ can be reinforced through careful identification of Technical Requirements that truly influence customer satisfaction. These technical requirements, or factors, can be verified or inferred through factor importance rankings identified through expert rankings through the Delphi methodology. Finally, the “roof” of the HOQ can be reinforced through validating technical factor correlations by comparing assumptions with historical data of 2-way factor interactions identified through the conjoint analysis phase of this research methodology. The research has shown the ability to identify both positive and negative 2-way interactions at varying levels of magnitude, as required for the HOQ. In summary, the primary superstructure of the QFD HOQ can be reinforced through the careful application of the proposed research methodology. This reinforcement provides quantitative stability to the traditionally qualitative process of building the HOQ.

Under this HOQ reinforcement paradigm, QFD practitioners would still be required to appropriately gather the VOC. Further research could be conducted to gather VOC through traditional QFD mechanisms and conduct parallel SNE development activities using QFD and the methodology presented in this research to better identify gaps and further opportunities for improvement across both the mature QFD approach and the novel methodology presented in this research.

Limitations

This research does not seek to provide definitive answers on improving the visual aesthetic quality across all synthetic environments, but instead provides a methodology and techniques for identifying importance of commonly subjective SNE visual aesthetic quality factors. While this research obtained statistically significant results within the SNE training task and geographic parameters presented, it is the researcher's hope that these results can be generally applied as a foundation across the SNE development community.

Lesson Learned

There are several lessons learned throughout this research that may benefit future practitioners of the techniques discussed throughout this research. The first relates to the selection and ordering of factors used within the screening design applied to the conjoint analysis. The original factors and factor levels depicted in Table 27 could have been better organized to allow for more in-depth data analysis and conclusions. Table 39 provides an alternative screening design that would have placed all "Low Fidelity" SNE factor levels in Level 1 (L1) and their "High Fidelity" counterpart in Level 2 (L2). This deliberate approach to ordering factor levels may have provided a clearer preference of SNE VAQ factors by users during the conjoint analysis trials.

Table 39. Re-Structured Synthetic Environment VAQ Factors from the Delphi Method to the Conjoint Analysis Screening Design to Reflect Low-Fidelity and High-Fidelity Factor levels. Factor Levels in blue have been updated from the original research screening design.

DOE Factor Name	Low Fidelity (L1)	High Fidelity (L2)	Description from Delphi Statements
Elevation	30m DEM	3m DEM	Accurate terrain surface elevation (Statement *23)*
Spatial Cues	Up to 2x Size Trees	Precise Size Trees	Inclusion of features providing spatial relationship cues (Statement *9*)
Building Orientation	10-degree building orientation error relative to the road	Precise Orientation	Accurate object orientation (Statement *8*)
Weather Effects	None	Rain/Storm effects	Realistic atmospheric, environmental, and weather effects (Statement *12*)
Vegetation Density	75% density error	100% density	Vegetation density (Statement *11*)
Building Complexity	Extruded buildings	Complex building models	Realistic dense urban terrain (Statement *2*)

Geo-specific Locations	Wrong feature placement (Lake and Space Needle)	Accurately place correct Space Needle Placement	Accurate geospecific locations (Statement *19*)
Time of Day	Day (noon)	Dusk	Time of day representation (Statement *4*)
Levels of Detail	Zero LOD Transitions	Two LOD Transitions	Common Level of Detail (LOD) transition (Statement *20*)

This research could have also benefited from an initial survey to query real-world SNE users on perceived important VAQ factors, like the traditional QFD approach of identifying the "voice of the customer" upfront and early in the process. Instead, this research first queried a combination of SNE users, developers, and managers. An upfront user-only survey could have been an initial baseline by which to compare the user feedback results identified from the applied conjoint analysis. This initial survey may have also led to better integration and application of research results with QFD.

Future Research

Further data analysis can be performed to understand the correlation of VAQ preferences across various training backgrounds and domains through the applied framework of techniques. The Resolution III design used for the Conjoint Analysis was beneficial for analyzing many factors and identifying significant main effects, but further experimentation should be conducted across a smaller number of factors to better explore and understand the interactions of these significant factors. A Resolution IV design can therefore be implemented to avoid confounding of main effects and 2-factor interactions. This follow-up experiment may also yield recommended parameters associated with each significant factor, as a fifth technique to identify quality tradeoff decisions. This could further help SNE developers to automate visual verification & validation software tools or tailor SNE rendering for specific user communities.

When analyzing the results of the Delphi Method, the technique appears to generate factors that are related to functional realism rather than reflective of user demand for physical and photo realistic SNE representations. Additional research can be conducted to validate and better understand this phenomenon in terms of Ferweda's realism definitions (Ferwerda, 2003).

Additional research can also be conducted to understand the efficiency and accuracy of the this proposed methodology against more traditional product and quality planning methodologies, such as Quality Function Deployment (QFD) and Total Quality Management (TQM).

Conclusions

The findings suggest that the subjective factors affecting the visual aesthetic quality of a synthetic natural environment can be quantified for specific military training tasks through the careful application of committee-based expert and user methodologies. The acceptance of these findings is caveat within the bounds and limitations of this research, specifically related to the fitness for use of training tasks and geographic extents. The researcher's hope is that the findings within this research will serve as a foundation for future synthetic natural environment modernization initiatives and inform developers on visual aesthetic quality considerations of multi-domain and combined arms synthetic environments. This research demonstrates that the structured application of both the Delphi and Conjoint Analysis methodologies can accurately identify significant importance rankings of subjective factors, as demonstrated through our vignette of visual aesthetic quality factors of synthetic environments. This research may also serve as basis to develop automated quality tools, VV&A processes, or provide the basis for SNE product planning. Further, the presented framework can be implemented by others looking to refine these initial findings or explore the significance of alternative subjective quality factors.

APPENDIX A: ELECTRONIC DELPHI SURVEY

Research Introduction

Research Overview

Thank you in advance for your participation in this Delphi survey. As an expert in the defense Modeling and Simulation industry, your input and feedback will be extremely important to this success of this research.

The goal of this research is to identify the primary visual aesthetic quality factors, factor priorities, and factor interactions that affect the correlation of a virtual Synthetic Natural Environment (SNE) and its representation across dissimilar simulators in a distributed interoperable simulation. Once identified and validated through a follow-on conjoint analysis study, understanding of these visual quality factors and associated parameters can be used to inform the SNE generation process upstream in order to reduce the cost and schedule of SNE Verification and Validation downstream. Earlier and improved upstream quality will likely result in SNE products being a much higher quality at the conclusion of their generation process and may alleviate many of the issues identified once they reach the user

Quality can be a subjective term for which each person, sector, or community may have its own definition. One may describe the degree to which a SNE reflects the real world natural environment in terms of fidelity, quality, acceptance, or some combination of the three. Unless otherwise described, this research will focus on quality in the sense of human-based, visual aesthetic of a SNE or Visual Aesthetic Quality (VAQ). This research defines VAQ as the visually pleasing appearance of a product in accordance to a customer needs and wants.

Given the quality linkage, a military-based case study is extensible beyond the military use-case. An important aspect of this research rests in the general process of identifying a user's needs and wants and incorporating them into the design and production of a product.

This research seeks to extend the complex scene study experiment first conducted by Purdy & Goldiez (1995). This research leverages the Purdy & Goldiez case-based interoperability problem definitions for consideration of SNE VAQ factors. Table 1 provides a summary of each case. This research will focus primarily on identifying the purely visual SNE factors affecting Case 3 and Case 4, which address the correlation and representation of a virtual SNE across dissimilar simulators.

Table 1: Purdy & Goldiez Interoperability Problem Definitions

Case Number	Case Title	Case Description
Case 1	Interaction/virtual incompatibilities	This type of interoperability problem arises when a characteristic or behavior of one type of virtual world entity is not recognized or is incompatible with the same characteristic or behavior of another entity.
Case 2	Low interactive fidelity	Virtual world entities should interact with each other and with the environment according to a set of rules such as the laws of physics.
Case 3	Non-uniform levels of simulation fidelity	This interoperability problem occurs when two simulators model objects or behavior at differing levels of fidelity. Such differences skew interaction results and degrade training efficacy.
Case 4	Differences in the virtual environment	An assumption made when connecting simulators is a single and common virtual environment. When the environment is not common, problems such as intervisibility, floating tanks and subterranean aircraft degrade the realism of the training scenario, skew the interaction outcome, and degrade training efficacy.

The Delphi Method

Developed in the early 1960's by the RAND corporation, the Delphi process is a decision-making technique that relies on the judgment of experts to achieve a convergence of opinion on a specific real-world issue. Researchers found that traditional round-table style discussion groups with the object of achieving a group consensus were plagued by negative psychological impediments, such as dominant personalities or those who had the tendency to want to appease all parties, a la "the loudest voice rather than the soundest argument may carry the day".

The Delphi reduces psychological factors, by eliminating the physical meeting of the committee altogether. Direct debate is replaced by a series of carefully planned and sequential questionnaires or surveys intermixed with feedback derived from computed consensus from earlier iterations. Often, researchers request members of an expert panel to provide reasons for their response, which is then subject to critique be fellow experts. The Delphi attempts to improve a committee's interaction by subjecting the views of an individual expert to group expert in an anonymous fashion in order to avoid the stigma of face-to-face confrontation. During this controlled debate, more often than not, an expert panel moves towards a consensus; but in the event that this does not occur, the reasoning for dissimilar opinions is made obvious.

During the this Delphi survey, you will first be asked to provide open-ended feedback responses to what you believe to be important SNE visual quality factors for consideration of a combined arms maneuver interoperable exercise consisting of two or more of the above virtual simulation system. The primary investigator will then analyze these responses and synthesize them into statements on SNE visual quality factors whereby you will be asked to rate your agreement to each stamen. This will be continued through three rounds or until a consensus is reached on the top eight SNE visual quality statements.

Ethics and Security Information

Participation in this Delphi survey is purely voluntary and you may withdraw your participation at any time. Please note however, that withdrawal from any phase of the Delphi process will disqualify you from further participation is future rounds of the Delphi.

The power of the Delphi technique comes from the anonymity it provides between expert peers, thus overcoming the psychological barriers of face-to-face committee discussions. Therefore, this research will protect respondent identity between other panel members; however due to the need for the researcher to provide feedback the responses will not be full anonymous. Only the primary researcher will be able to link participants to their responses since the researcher will strip any personally identifiable information (PII) from the feedback and will only contain a unique ID assigned to each participant. This research will treat participant's PII as confidential.

Researcher access to the survey data will be password protected and hosted on SurveyMonkey's secure servers. SurveyMonkey implements a combination of physical and digital security measures to protect customer data on these servers. Some of these are outlined below, but more in-depth coverage can be found at SurveyMonkey's Security Statement.

User Security

Authentication: User data on our database is logically segregated by account-based access rules. User accounts have unique usernames and passwords that must be entered each time a user logs on. SurveyMonkey issues a session cookie only to record encrypted authentication information for the duration of a specific session. The session cookie does not include the password of the user.

Passwords: User application passwords have minimum complexity requirements. Passwords are individually salted and hashed.

Single Sign-On: For our Team Collaboration accounts, SurveyMonkey supports SAML 2.0 integration, which allows you to control access to SurveyMonkey across your organization and define authentication policies for increased security. For more details, visit our SSO help page.

Data Encryption: Certain sensitive user data, such as credit card details and account passwords, are stored in encrypted format.

Data Portability: SurveyMonkey enables you to export your data from our system in a variety of formats so that you can back it up, or use it with other applications.

Privacy: We have a comprehensive privacy policy that provides a very transparent view of how we handle your data, including how we use your data, who we share it with, and how long we retain it.

Data Residency: All SurveyMonkey user data, to include Wufoo, TechValidate, SurveyMonkey Intelligence, is stored on servers located in the United States. For FluidSurveys and FluidReview, all data is stored in Canada.

Physical Security

All SurveyMonkey information systems and infrastructure are hosted in world-class data centers. These data centers include all the necessary physical security controls you would expect in a data center these days (e.g., 24×7 monitoring, cameras, visitor logs, entry requirements). SurveyMonkey has dedicated cages to separate our equipment from other tenants. In addition, these data centers are SOC 2 accredited. For more information, visit SuperNAP and InterNAP. If you are looking for FluidSurvey or FluidReview information, please contact us directly.

Power: Servers have redundant internal and external power supplies. Data centers have backup power supplies, and are able to draw power from the multiple substations on the grid, several diesel generators, and backup batteries.

Network Security

Firewalls: Firewalls restrict access to all ports except 80 (http) and 443 (https).

Access Control: Secure VPN, 2FA (two-factor authentication), and role-based access is enforced for systems management by authorized engineering staff.

Logging and Auditing: Central logging systems capture and archive all internal systems access including any failed authentication attempts.

Encryption in Transit: By default, our survey collectors have Transport Layer Security (TLS) enabled to encrypt respondent traffic. All other communications with the surveymonkey.com website are sent over TLS connections, which protects communications by using both server authentication and data encryption. This ensures that user data in transit is safe, secure, and available only to intended recipients. Our application endpoints are TLS only and score an "A" rating on SSL Labs' tests. We also employ Forward Secrecy and only support strong ciphers for added privacy and security.

References:

1. Purdy, R., & Goldiez, B. (1995). Interoperability Issues Associated With the Use Of Dissimilar Simulators. In The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC). Retrieved from http://www.iitsecdocs.com/download/1995/335466Q256571403

Next

Delphi: SNE Visual Aesthetic Quality for Interoperable Use Case

Instructions

This research implements a case study approach to understanding the visual aesthetic quality (VAQ) factors affecting the correlation and representation of a virtual synthetic natural environment (SNE) across dissimilar simulators.

This Delphi survey will consist of no more than four rounds of questionnaires and feedback. The goal of this Delphi is to reach expert consensus for at least eight VAQ factors or considerations for designing/representing a SNE for a virtual interoperable training scenario. Each round of the survey will remain open for a period of two-weeks. During that time, three reminder emails will be sent out to participants who have not yet completed the survey. Participants who are not able to complete a round of the survey will be removed from all future survey rounds in order to maintain consensus building.

Use Case

In this survey, you will be asked to identify the primary visual considerations that affect the aesthetic, or visually pleasing, quality of an interoperable SNE representation for use across dissimilar virtual Training Aids, Devices, Simulators, and Simulations (TADSS). This research does not seek to address the impact of interoperability to specific training tasks and virtual simulation use-cases.

Demographics Questionnaire

By virtue of your selection on this panel, you are considered by the researcher to be an expert on one or more of the platforms or domains in Table 1. To help the researcher to better understand the domain-specific demographics of this research, you will be asked to fill out a short questionnaire to indicate you level of familiarity with each of the primary virtual simulation training domains. You will also be asked to provide a contact email address to be used to provide survey feedback, a hallmark of the Delphi technique.

Round One of the Delphi Survey

Round one of the Delphi survey will present panel members with fifteen opportunities to provided open-ended responses on visual aesthetic qualify factors and considerations for designing a virtual interoperable SNE representation. For each item of consideration, experts will be asked to provide a brief description of the VAQ factor, a justification for why this visual factor is important for the visual aesthetic of the scene and the interoperable use-case, as well as provide recommended parameters for this factor or consideration in order to improve correlation between dissimilar simulations. While there are fifteen open-ended question, only ten are mandatory for completion for Round one.

Rounds Two through Four of the Delphi Survey

The researcher will carefully analyze each open-ended response from each participant and synthesize a series of statements on VAQ factors and their impact to a virtual interoperable SNE representation and use-case. During Round two of the Delphi, participants will be asked to rate their level of agreement with each of these stamens based on a seven-level Likert type scale rating and provide a justification for their rating. Once complete, the researcher will analyze this data and provide feedback to each participant in the form of summary of Round two that will indicate where their ratings fell in comparison with the panel averages. In this research, consensus on a statement is reached once 80-ercent of panel votes fall within two rating categories of the Likert scale. Using their Round two summary for reference, Round three will ask participants to either revise or maintain their rating towards panel consensus and will again require participant justification for moving towards or against consensus. Again, feedback will be provided in the form of a round three summary report and the process will continue for a fourth round or stop if consensus has been reached on at least eight statements.

Prev Next

Delphi: SNE Visual Aesthetic Quality for Interoperable Use Case

Demographics Information

1. Please enter an email address. This will allow us to provide feedback and track responses. This address will remain confidential and made only accessible to the primary investigator.

2. Please select your primary **virtual** training domain background:

3. Please rate your level of familiartiy with each of the below virtual collective simulator domains.

	Not at all familiar	Slightly familiar	Somewhat familiar (Novice)	Moderately familiar	Extremely familiar (Expert)
Armor (M1 Abrams, M2 Bradley, etc.)	○	○	○	○	○
Ground-wheeled (HMMWV, HEMTT, etc.)	○	○	○	○	○
Dismounted Soldier/Squad/Platoon	○	○	○	○	○
Rotary-wing Aviation (AH-64, UH-60, UH-72, etc.)	○	○	○	○	○
Fixed-wing Aviation (MQ-1 Predator, MQ-1C Grey Eagle, etc.)	○	○	○	○	○
Distributed Interactive Simulation / Live, Virtual, Constrictive, and Gaming (LVCG)	○	○	○	○	○

Please specify other virtual simulation experience outside of the above domains.

Prev Next

Powered by

Delphi: SNE Visual Aesthetic Quality for Interoperable Use Case

SNE Visual Aesthetic Quality Factor Considerations

Referencing the Purdy & Goldiez Case 3 & 4 interoperability problem definitions as well as the interoperable SNE use-case identified on Page 2, use the below open-ended response modules to provide your feedback on the purely visual factors and considerations affecting the aesthetic quality of virtual SNE representation across dissimilar simulators. What are the primary visually pleasing appearance considerations when designing a SNE for a virtual distributed interactive simulation exercise?

EXAMPLE:

Visual Aesthetic Quality Factor Description: *Ground-texture imagery fidelity and resolution*

Justification: *Imagery must be of sufficient detail to allow for aviation navigation, yet must be of high resolution to prevent distraction of the ground simulation user due to imagery pixilation.*

Parameter Consideration: *Developers should strive for sub-meter real world imagery, but down sample the resolution to maintain target simulation system performance. Alternatively, a hybrid or synthetic imagery solution can be implemented.*

4. Please describe a quality factor consideration to achieve a visually appealing virtual SNE representation for an interoperable use-case.

Visual Aesthetic Quality Factor Description:	
Justification:	
Parameter Consideration (if applicable):	
Additional remarks:	

5. Please describe a quality factor consideration to achieve a visually appealing virtual SNE representation for an interoperable use-case.

Visual Aesthetic Quality Factor Description:

Justification:

Parameter Consideration (if applicable):

Additional remarks:

6. Please describe a quality factor consideration to achieve a visually appealing virtual SNE representation for an interoperable use-case.

Visual Aesthetic Quality Factor Description:

Justification:

Parameter Consideration (if applicable):

Additional remarks:

7. Please describe a quality factor consideration to achieve a visually appealing virtual SNE representation for an interoperable use-case.

Visual Aesthetic Quality Factor Description:

Justification:

Parameter Consideration (if applicable):

Additional remarks:

Prev Done

Powered by
SurveyMonkey®
See how easy it is to create a survey.

Delphi: SNE Visual Aesthetic Quality for Interoperable Use Case (Round 2)

---

* 1. SNE Visual Quality Factor Statement 1 - Median: 4 / Std. Dev: 2.01

Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
○	○	○	○	○	○	○

Justification FOR or AGAINST Consensus

* 2. SNE Visual Quality Factor Statement 2 - Median: 4 / Std. Dev: 1.70

Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
○	○	○	○	○	○	○

Justification FOR or AGAINST Consensus

* 3. SNE Visual Quality Factor Statement 3 - Median: 4 / Std. Dev: 1.60

Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
○	○	○	○	○	○	○

Justification FOR or AGAINST Consensus

* 4. SNE Visual Quality Factor Statement 4 - Median: 4 / Std. Dev: 2.06

Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
○	○	○	○	○	○	○

Justification FOR or AGAINST Consensus

* 5. SNE Visual Quality Factor Statement 5 - Median: 4 / Std. Dev: 2.11

Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
○	○	○	○	○	○	○

Justification FOR or AGAINST Consensus

* 6. SNE Visual Quality Factor Statement 6 - Mean: 5 / Std. Dev: 2.11

Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
○	○	○	○	○	○	○

Justification FOR or AGAINST Consensus

* 7. SNE Visual Quality Factor Statement 7 - Mean: 4 / Std. Dev: 2.11

Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
○	○	○	○	○	○	○

Justification FOR or AGAINST Consensus

* 8. SNE Visual Quality Factor Statement 8 - Mean: 4 / Std. Dev: 2.14

Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
○	○	○	○	○	○	○

Justification FOR or AGAINST Consensus

* 9. SNE Visual Quality Factor Statement 9 - Mean: 4 / Std. Dev: 2.10

Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
○	○	○	○	○	○	○

Justification FOR or AGAINST Consensus

* 10. SNE Visual Quality Factor Statement 10 - Mean: 5 / Std. Dev: 2.13

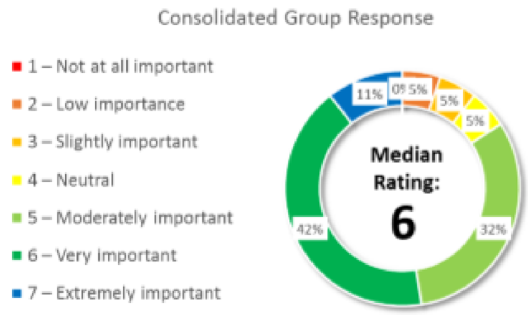
Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
○	○	○	○	○	○	○

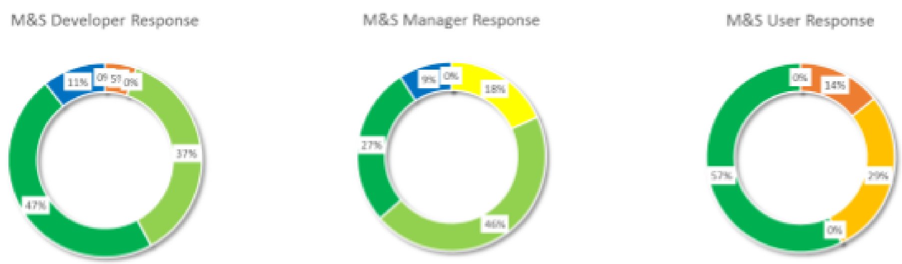
Justification FOR or AGAINST Consensus

Done

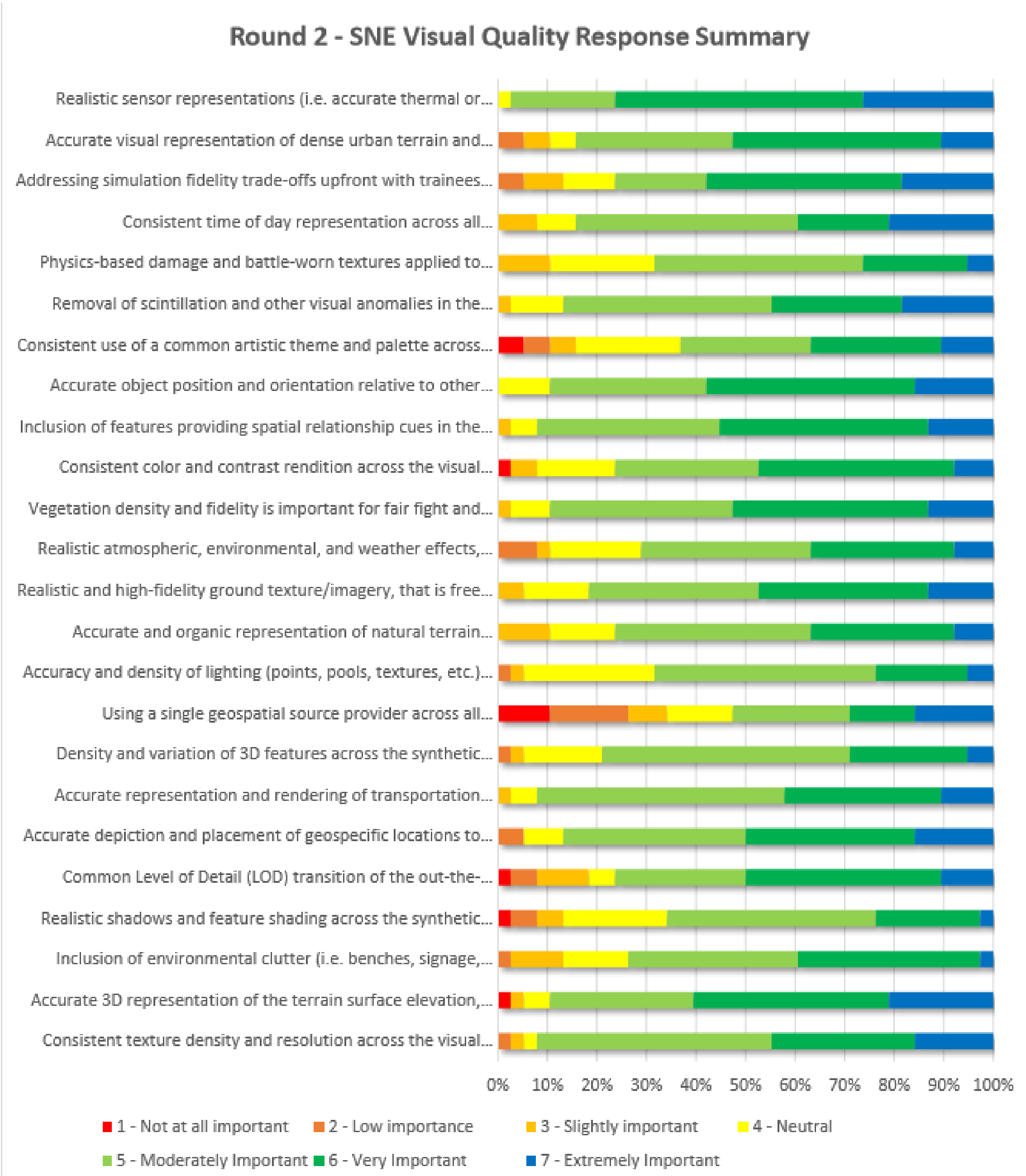
APPENDIX B: SAMPLE INDIVIDUAL DELPHI FEEDBACK REPORT

FACTOR 2: *Accurate visual representation of dense urban terrain and megacities is important to the visual quality of virtual simulation for the training scenario.*





Justification <u>FOR</u> Importance:	• Depends on the training mission. This would be important for mission rehearsal more than a collective training event. • Dense urban terrain where applicable is a key element of modern unit operations.
NEUTRAL Justification	• It should depend on the training objective of the scenario. • It would help to immerse the trainee into the simulation but it adds a large toll on the system resources. • Although this subject is trending now, it is only one element of the full operational domain.
Justification <u>AGAINST</u> Importance:	• Most mounted virtual combat training is not conducted in a jungle or urban environment. Damageable building and ruble are more important to replicate for the purpose of training.



Round 2 - SNE Visual Quality Response Summary
Realistic sensor representations (i.e. accurate thermal or...
Accurate visual representation of dense urban terrain and...
Addressing simulation fidelity trade-offs upfront with trainees...
Consistent time of day representation across all...
Physics-based damage and battle-worn textures applied to...
Removal of scintillation and other visual anomalies in the...
Consistent use of a common artistic theme and palette across...
Accurate object position and orientation relative to other...
Inclusion of features providing spatial relationship cues in the...
Consistent color and contrast rendition across the visual...
Vegetation density and fidelity is important for fair fight and...
Realistic atmospheric, environmental, and weather effects,...
Realistic and high-fidelity ground texture/imagery, that is free...
Accurate and organic representation of natural terrain...
Accuracy and density of lighting (points, pools, textures, etc.)...
Using a single geospatial source provider across all...
Density and variation of 3D features across the synthetic...
Accurate representation and rendering of transportation...
Accurate depiction and placement of geospecific locations to...
Common Level of Detail (LOD) transition of the out-the-...
Realistic shadows and feature shading across the synthetic...
Inclusion of environmental clutter (i.e. benches, signage,...
Accurate 3D representation of the terrain surface elevation,...
Consistent texture density and resolution across the visual...
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
1 - Not at all important
2 - Low importance
3 - Slightly important
4 - Neutral
5 - Moderately Important
6 - Very Important
7 - Extremely Important

APPENDIX C: SNE VISUAL AESTHETIC QUALITY CONJOINT ANALYSIS SURVEY

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Survey Introduction and Consent

Thank you for your participation in this survey. Your expertise and feedback during this study will be critical to shaping the future of U.S. Army Synthetic Natural Environment (SNE) development and test processes.

Research Goal:
The goal of this research is to identify factors, priorities, and interactions that affect the Visual Quality (VQ) of virtual and interoperable Synthetic Natural Environments (SNEs) for use across dissimilar virtual simulators. Quality can be a subjective term describing the degree to which a SNE reflects the real world natural environment for a particular use case or training task. This research defines VQ as the quality and visual appearance of a product in accordance with a customer/user needs and wants

Research Background:
This research leverages and seeks to extend the complex scene study experiment first conducted by Purdy & Goldiez (1995). This research focuses [m1] on factors affecting the representation of a virtual SNE across dissimilar simulators. A summary of the relevant cases from that study is provided below.

Case 3: Non-uniform levels of simulation fidelity - This interoperability problem occurs when two simulators model objects or behavior at differing levels of fidelity. Such differences skew interaction results and degrade training efficacy.

Case 4: Differences in the virtual environment - An assumption made when connecting simulators is the use of a single and common virtual environment. When the environment is not common, problems such as inter-visibility, floating tanks and subterranean aircraft degrade the realism of the training scenario, skew the interaction outcome, and degrade training efficacy.

References:

1. Purdy, R., & Goldiez, B. (1995). Interoperability Issues Associated With the Use Of Dissimilar Simulators. In The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC). Retrieved from http://www.iitsecdocs.com/download/1995/335466Q256571403

Survey Instructions:
In this survey you will be asked to rank the visual quality of **twenty (20)** sets of SNE visual scenes depicting downtown Seattle, WA. A ground vehicle and a UAV perspective on the SNE scenes are provided. The SNE visual scenes are presented as a series of ground-level and aerial fly-through videos that replay to allow continued reviewing of the visual scenes. As a preamble to SNE scenes, a Google Earth fly-through on the same visual scenes is also provided as a means of establishing familiarity with the actual environment and a frame of reference for evaluation the SNE visual scenes. All videos contain full playback controls and allows you to pause, rewind, and fast forward playback. The videos can also be expanded to full-screen to allow for more detailed review. In order to force factor prioritization and importance of interactions, no single SNE visual scene is considered visually perfect. Each scene may contain various environmental effects such as rain, clouds, and lighting artifacts. Visual anomalies and even errors may be present throughout all scenes. Your task will be to rank the visual quality of the scenes presented from the perspective described below. *No ties will be allowed in the ranking and all scenes must be ranked.*

Scenario:
All synthetic natural environment visual quality comparisons will be rated against the following context training scenario.

A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aviation asset overhead providing route reconnaissance (ex. UAS or rotary-wing aviation)

Real-World Terrain Depiction:

A real-world fly-through of the terrain being evaluated is provided below. This video was captured through Google Earth Software. This video is intended only to provide a frame of reference and not as a direct comparison for the visual scenes being ranked. The yellow lines represented in the below video are an artifact of Google Streetview and should be ignored.

PLEASE NOTE: You may start, stop, and return to the survey at any time using the survey link, but your responses WILL NOT save between attempts. I recommend completing this survey in a single attempt if possible.

* Written documentation of consent is not required for this research, but written information describing the research to be performed must be provided to all participants. A copy of the research information form can be found at the following link: Study Consent Form. Please save a copy of this document for your records.

- ◯ Yes. I consent to participate in this survey and have downloaded a copy of the research information form for my records.
- ◯ No. I wish to withdraw my participation from this survey.

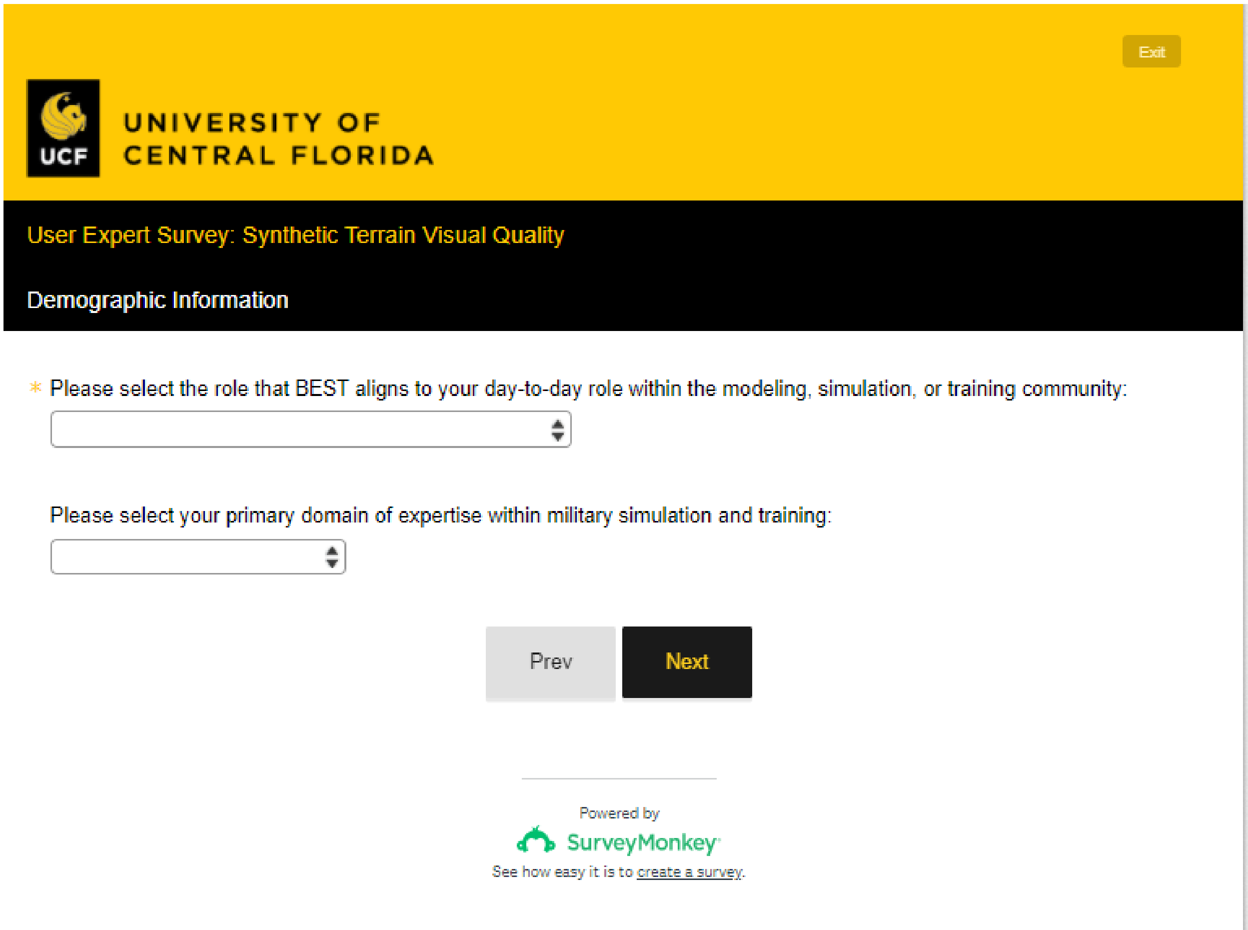
Next

Powered by

SurveyMonkey

See how easy it is to create a survey.

Privacy & Cookie Policy



Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Demographic Information

* Please select the role that BEST aligns to your day-to-day role within the modeling, simulation, or training community:

Please select your primary domain of expertise within military simulation and training:

Prev Next

Powered by SurveyMonkey

See how easy it is to create a survey.

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*

* Q1. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

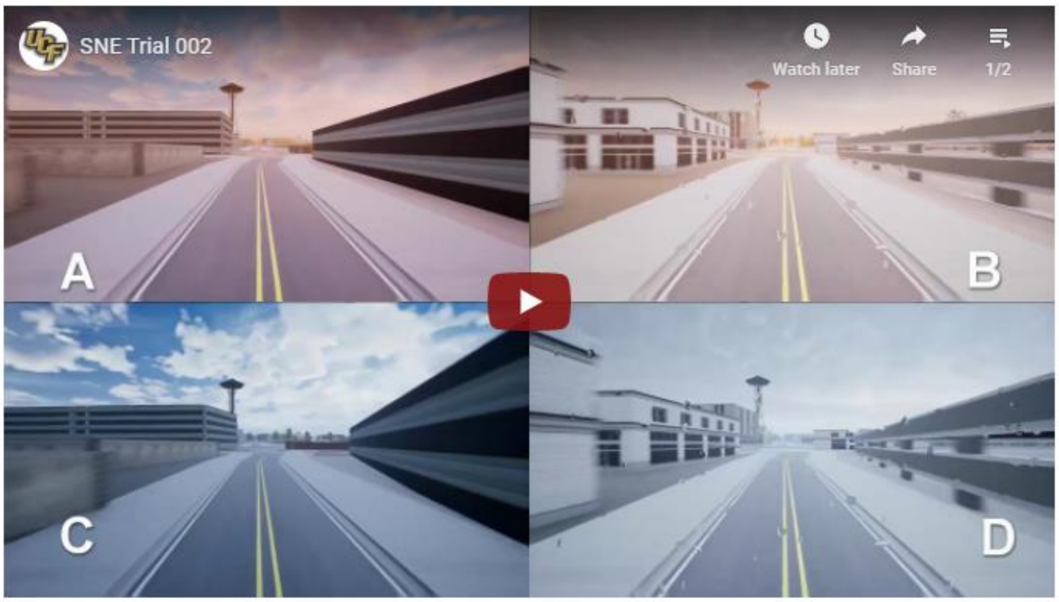
Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*



* Q2. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

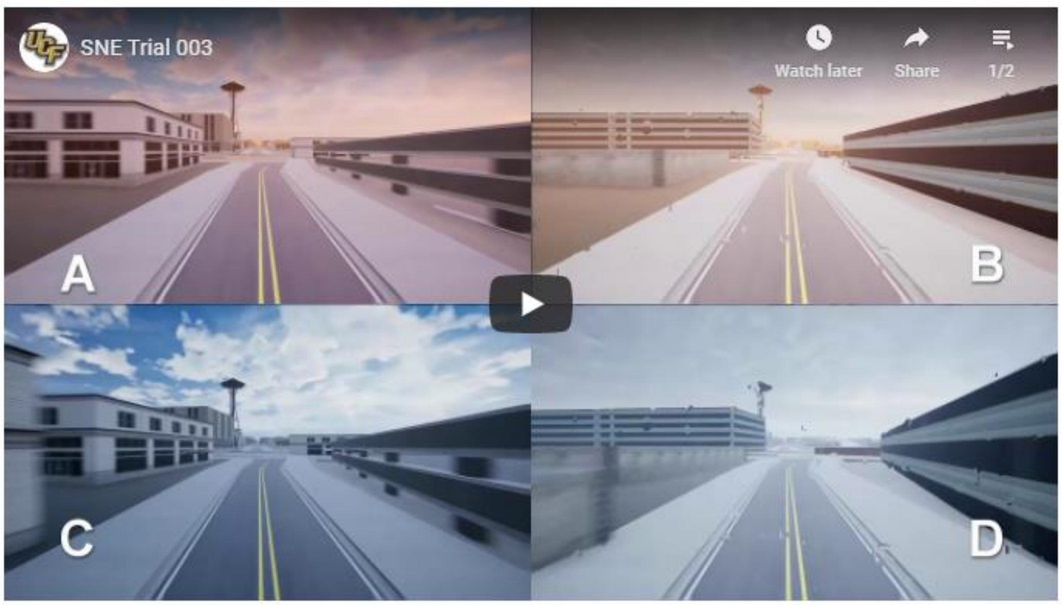
Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

__Scenario Reference__: A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).



* Q3. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*

* Q4. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

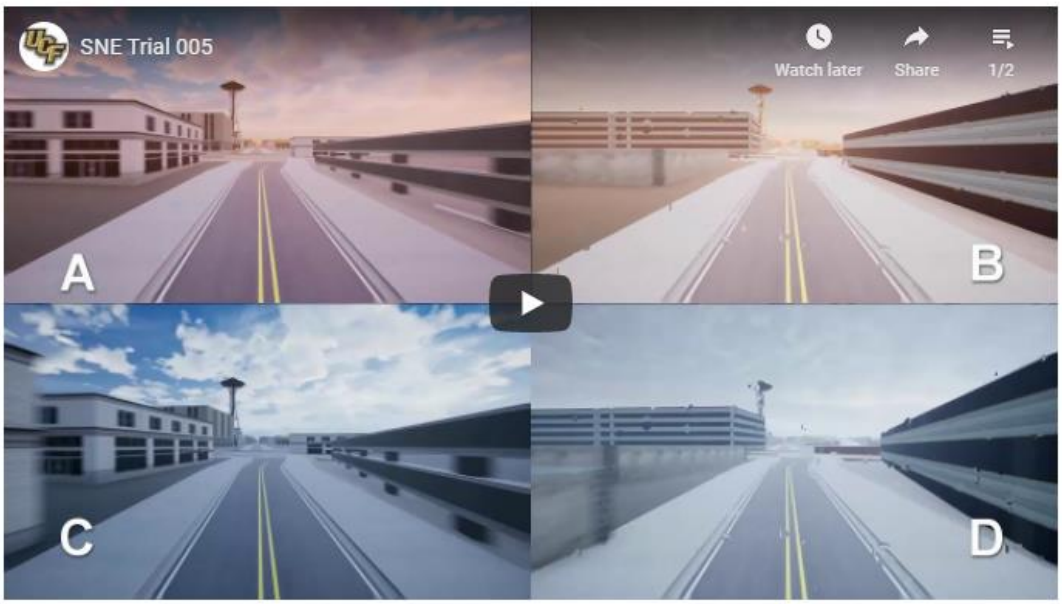
Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*



* Q5. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	◯	◯	◯	◯
Option B	◯	◯	◯	◯
Option C	◯	◯	◯	◯
Option D	◯	◯	◯	◯

What was the rationale for your ranking? (optional)

Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*

* Q6. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

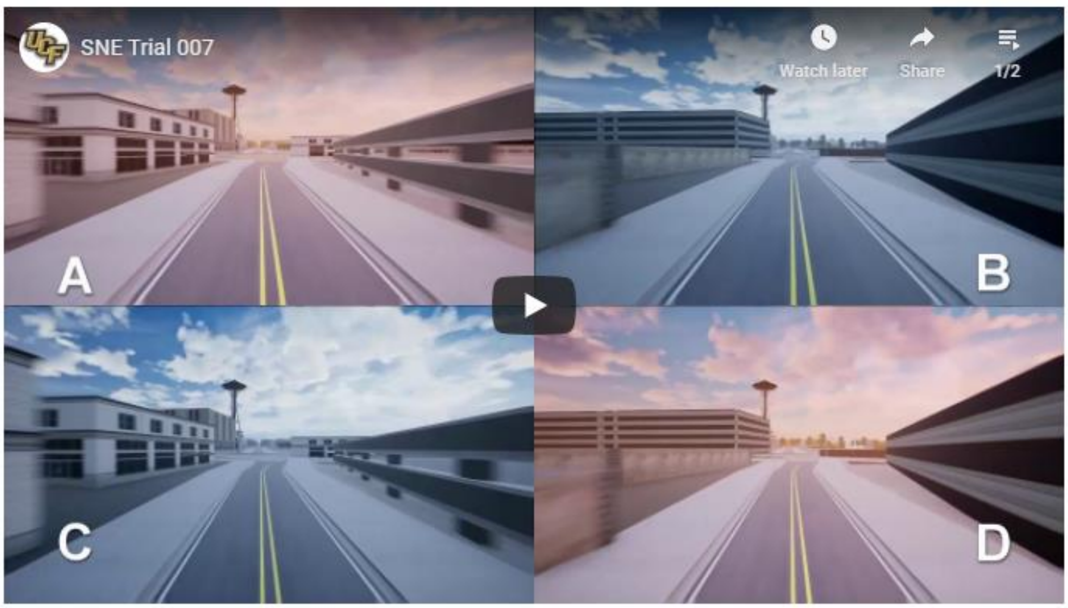
Prev Next

Exit

UNIVERSITY OF
CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*



* Q7. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

<u>Scenario Reference</u>: A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).

* Q8. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*

* Q9. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

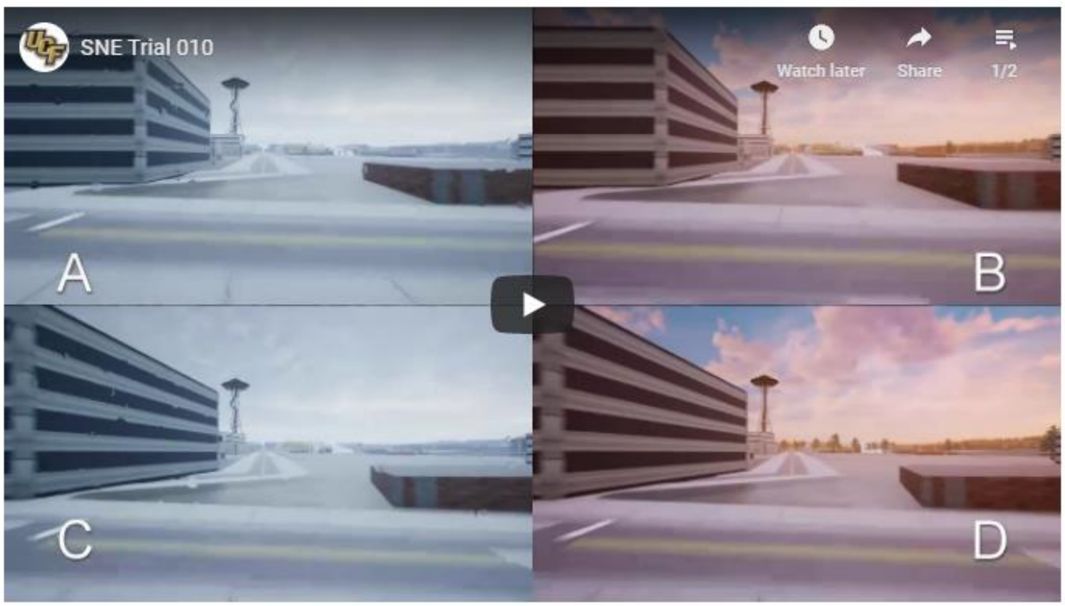
Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

<u>Scenario Reference</u>: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*



* Q10. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

Prev | Next

UNIVERSITY OF CENTRAL FLORIDA

Exit

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*

* Q11. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

Prev Next

UNIVERSITY OF CENTRAL FLORIDA

Exit

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*

* Q12. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	◯	◯	◯	◯
Option B	◯	◯	◯	◯
Option C	◯	◯	◯	◯
Option D	◯	◯	◯	◯

What was the rationale for your ranking? (optional)

Prev Next

Exit

UNIVERSITY OF
CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*

* Q13. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

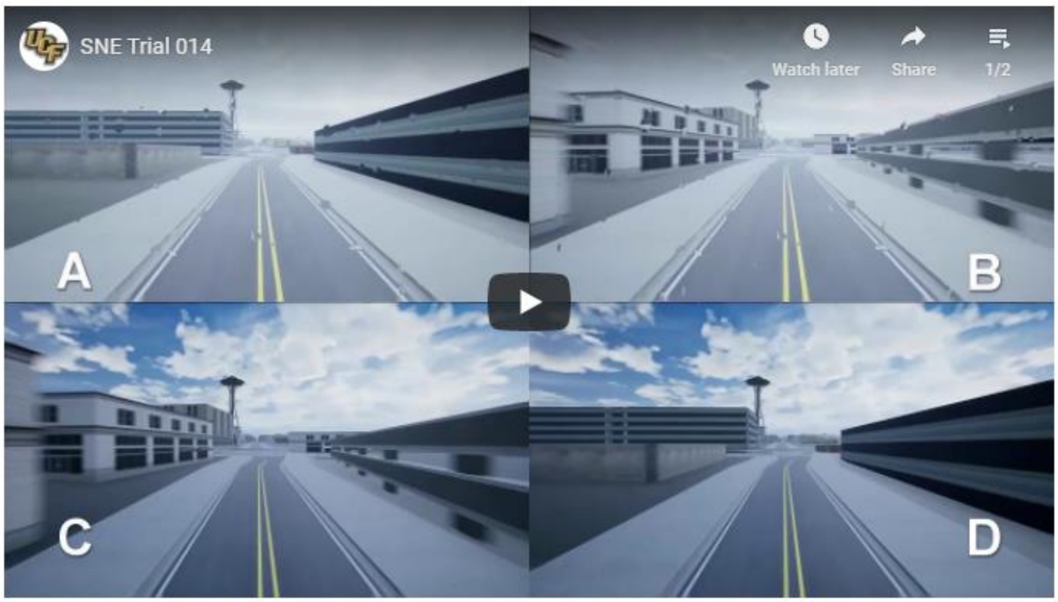
Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*



* Q14. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

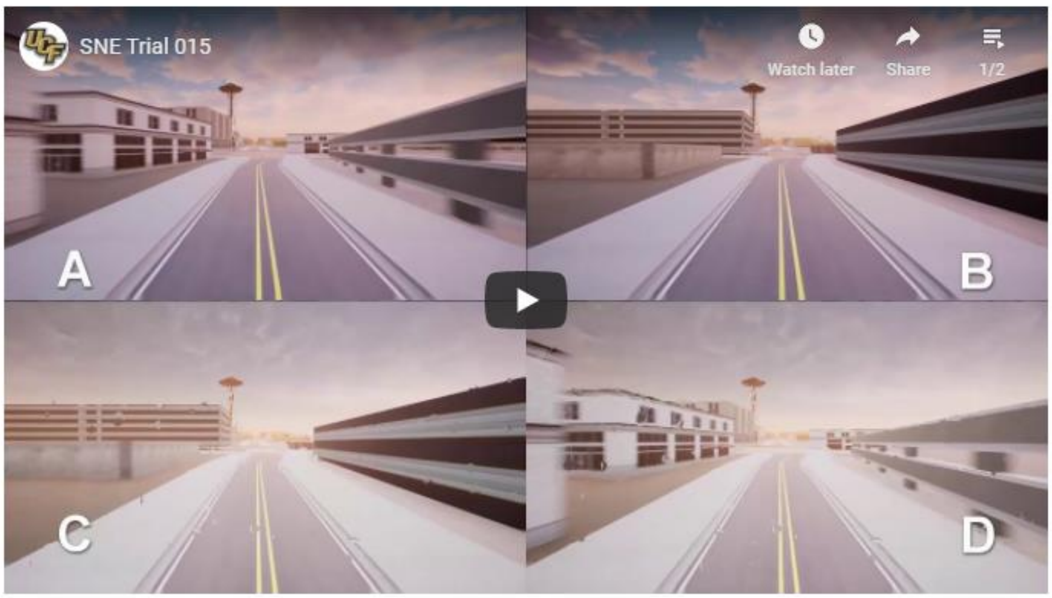
Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*



* Q15. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*

* Q16. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

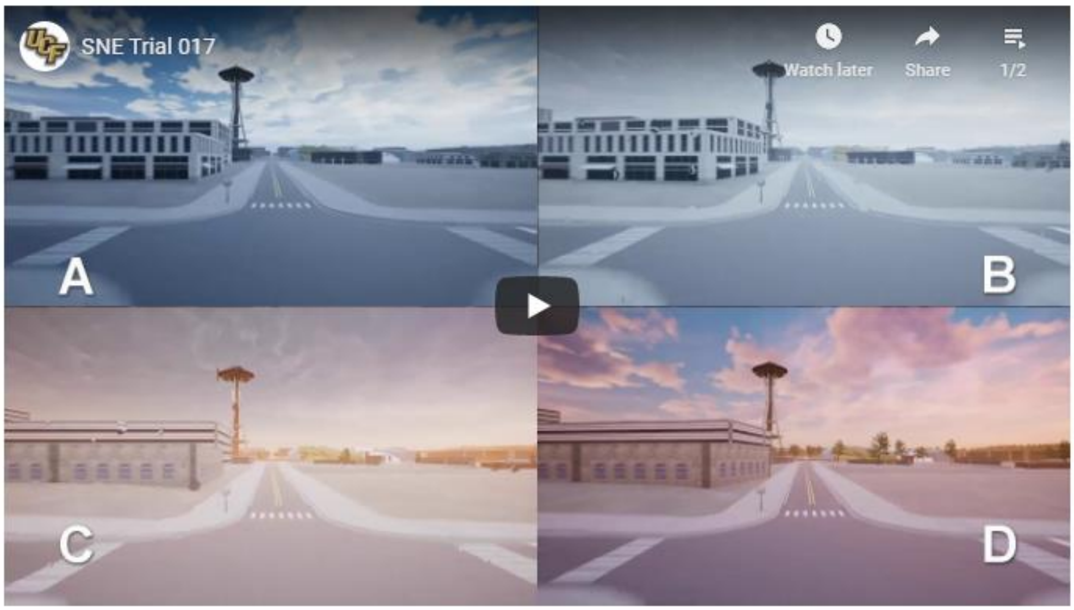
Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*



* Q17. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

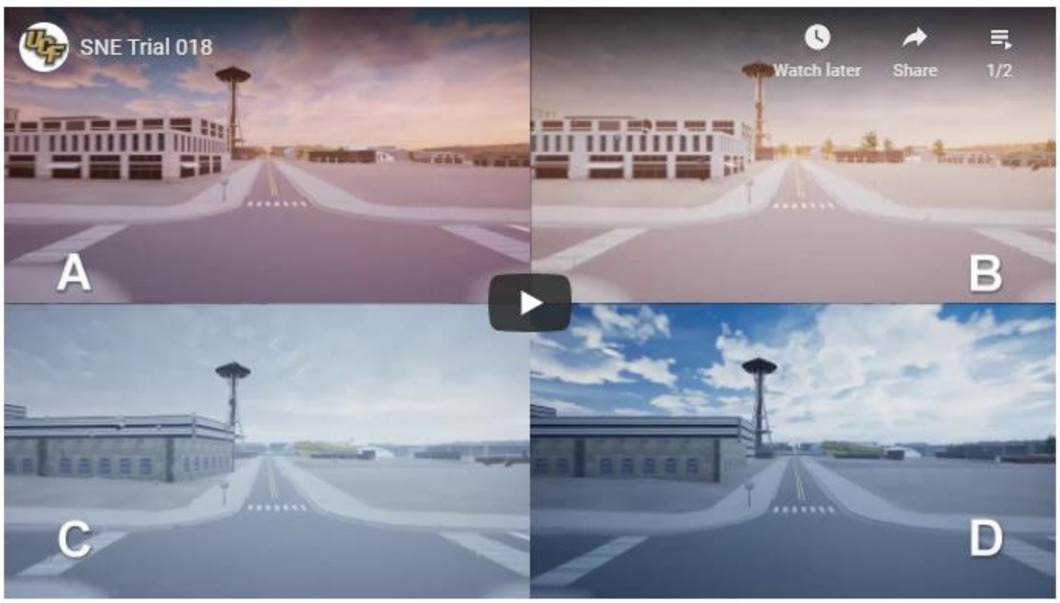
Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

<u>Scenario Reference</u>: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*



* Q18. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

Prev Next

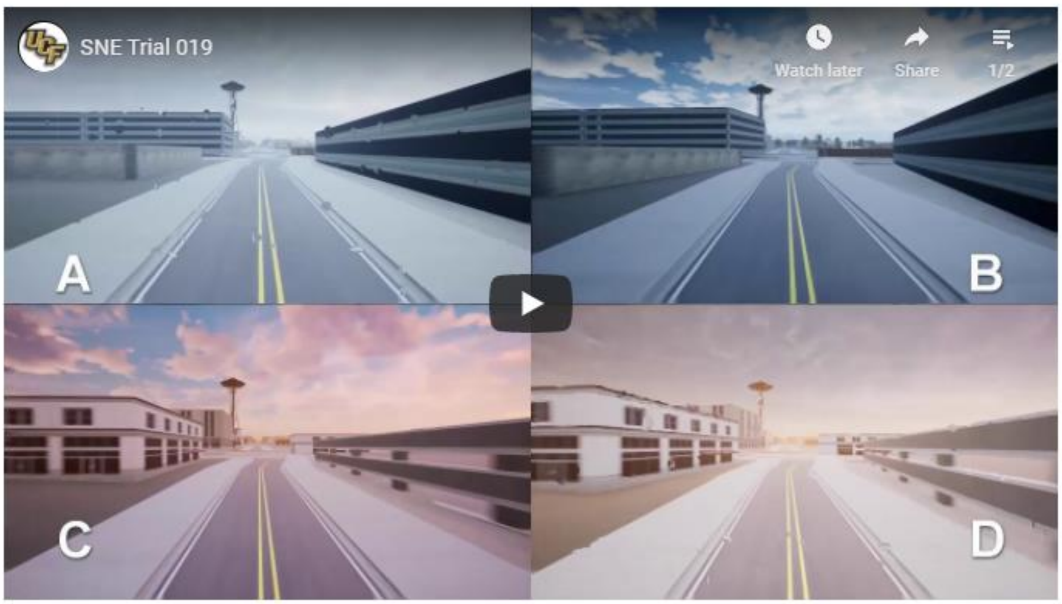
Exit

UNIVERSITY OF CENTRAL FLORIDA

UCF

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*



* Q19. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

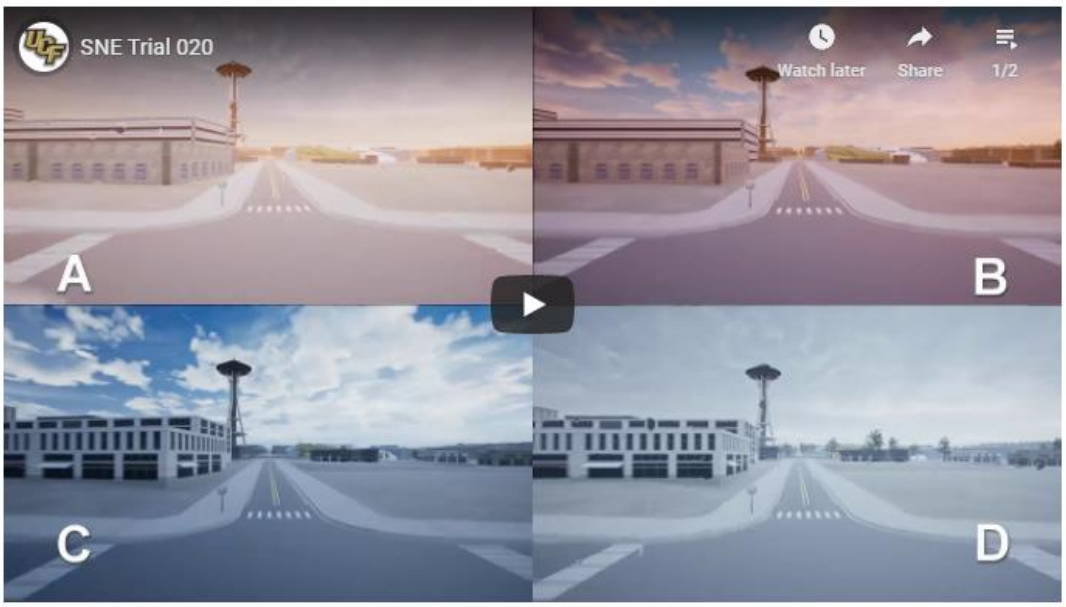
Prev Next

Exit

UNIVERSITY OF CENTRAL FLORIDA

User Expert Survey: Synthetic Terrain Visual Quality

Scenario Reference: *A multi-domain exercise consisting of ground (wheeled) vehicles following a convoy route supported by an aerial platform overhead providing route reconnaissance (ex. UAS or rotary-wing).*



* Q20. Rank the visual quality of each terrain depicted above for the provided training scenario

	First	Second	Third	Fourth
Option A	○	○	○	○
Option B	○	○	○	○
Option C	○	○	○	○
Option D	○	○	○	○

What was the rationale for your ranking? (optional)

Prev　Done

APPENDIC D: UCF PARTICIPANT CONSENT FORM

UNIVERSITY OF CENTRAL FLORIDA

Title of research study: ***Conjoint Analysis: Visual Quality of Synthetic Natural Environments***

Investigator: ***Thomas W. Kehr II***

Key Information: The following is a short summary of this study to help you decide whether or not to be a part of this study. More detailed information is listed later on in this form.

Why am I being invited to take part in a research study?

We invite you to take part in a research study because you have been identified as a user or expert within the military modeling, simulation, and training community.

Why is this research being done?

The purpose of this study is to identify the primary factors that affect Synthetic Natural Environment (SNE) visual aesthetic quality for virtual simulations. The results of this study will better inform the SNE development process upstream and reduce the fiscal and temporal resources required for testing and evaluation downstream.

How long will the research last and what will I need to do?

Your participation in this survey is expected to last 45-60 minutes.

You will be asked to complete an online electronic survey through the SurveyMonkey web-based survey application. This survey will require you to rank a serios of visual scenes of synthetic terrain representations.

More detailed information about the study procedures can be found under *"What happens if I say yes, I want to be in this research?"*

Is there any way being in this study could be bad for me?

There are no reasonably foreseeable risks or discomforts involved in taking part in this study.

More detailed information about the risks of this study can be found under *"Is there any way being in this study could be bad for me? (Detailed Risks)*

Will being in this study help me any way?

We cannot promise any benefits to you or others from your taking part in this research. However, possible benefits include an increased knowledge base of synthetic natural environment generation processes and quality factors as well as virtual collective simulation systems.

What happens if I do not want to be in this research?

Participation in research is completely voluntary. You can decide to participate or not to participate.

Detailed Information: The following is more detailed information about this study in addition to the information listed above.

What should I know about a research study?

- Someone will explain this research study to you.
- Whether or not you take part is up to you.
- You can choose not to take part.
- You can agree to take part and later change your mind.
- Your decision will not be held against you.
- You can ask all the questions you want before you decide.

Who can I talk to?

If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team:

Primary Investigator (student): Thomas W Kehr II, Thomas.kehr@knights.ucf.edu

Faculty Advisor: Michael D. Proctor, PhD, Michael.Proctor@ucf.edu

This research has been reviewed and approved by an Institutional Review Board ("IRB"). You may talk to them at 407-823-2901or irb@ucf.edu if:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You have questions about your rights as a research subject.
- You want to get information or provide input about this research.

How many people will be studied?

We expect 45-150 people will be in this research study.

What happens if I say yes, I want to be in this research?

The survey will take 30-60 minutes to complete.

You will first be asked to read an experimental consent form and acknowledge your participation.

Next, you will be asked to read an overview and instruction for completing the survey.

You will then be asked to complete a short demographic questionnaire to determine your proficiency and experience with SNE and virtual collective simulation.

Finally, you will be asked to rank a series of Synthetic Natural Environment (SNE) representations through the electronic survey. These SNEs will be presented as a series videos through YouTube. You will be asked to rank order the SNE views across 20 survey pages

What happens if I say yes, but I change my mind later?

You can leave the research at any time it will not be held against you. Refusal to take part in or withdrawal from this study will involve no penalty or loss of benefits you would receive by staying in it. Military personnel cannot be punished under the Uniform Code of Military Justice for choosing not to take part in or withdrawing from this study and cannot receive administrative sanctions for choosing not to participate.

Is there any way being in this study could be bad for me? (Detailed Risks)

There are no reasonably foreseeable risks or discomforts involved in taking part in this study.

There is a small risk that people who take part will develop what is ordinarily referred to as simulator sickness. Side effects of VE (virtual environment) use may include stomach discomfort, headaches, sleepiness, dizziness and decreased balance. However, these risks are no greater than the sickness risks participants may be exposed to if they were to visit an amusement park such as Disney World or Universal Studios parks and ride attractions such as roller coasters. You can take a break at any time during the survey to lessen the chance that you will feel sick. If you experience any of the symptoms mentioned, please pause the survey until the symptoms disappear.

What happens to the information collected for the research?

Personally Identifiable Information (PII) will not be collected through this survey. We will limit your data collected in this study to people who have a need to review this information. We cannot promise complete secrecy. Organizations that may inspect and copy your information include the IRB and other representatives of UCF. Data from the research survey will be stored on the Survey Monkey web portal. Access to the SurveyMonkey web portal is password protected and the data is hosted on secure servers. Once the survey and data analysis is complete, SurveyMonkey will be contacted to wipe the data from its servers, although the data can potentially still be retrievable for up to 12 months due to applicable law. SurveyMonkey maintains a high level of security from a variety of means. From a physical perspective, they store servers in locked cages, monitor digital surveillance systems, and utilize facility intrusion detections systems. Electronically they use Secure Socket Layer (SSL) password encryption, weekly network security audits, firewall restrictions, and daily hacker safe scans. Additionally, all data is backed up hourly, with daily backup and offsite storage.

What else do I need to know?

The primary investigator understands that this research may be of interest to the participants of this study. Results of this research can be obtained from Thomas Kehr, PhD Candidate, School of Modeling, Simulation, and Training, thomas.kehr@knights.ucf.edu , following the publication of the related dissertation.

APPENDIX E: UCF IRB APPROVAL LETTER

UCF

UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board
FWA00000351
IRB00001138Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

APPROVAL

October 7, 2019

Dear Thomas Kehr:

On 10/7/2019, the IRB reviewed the following submission:

Type of Review:	Initial Study
Title:	Exploring Delphi Method Generated Visual Aesthetic Quality (VAQ) Factor Forecasts and Preferences Through Conjoint Analysis of SNE End User Assessments
Investigator:	Thomas Kehr
IRB ID:	STUDY00000949
Funding:	None
Documents Reviewed:	• Adult Consent - Kehr 2019 • IRB Protocol - Kehr 2019 • Recruitment Letter - Kehr 2019 • Survey Content - Kehr 2019

The IRB approved the protocol on 10/7/2019.

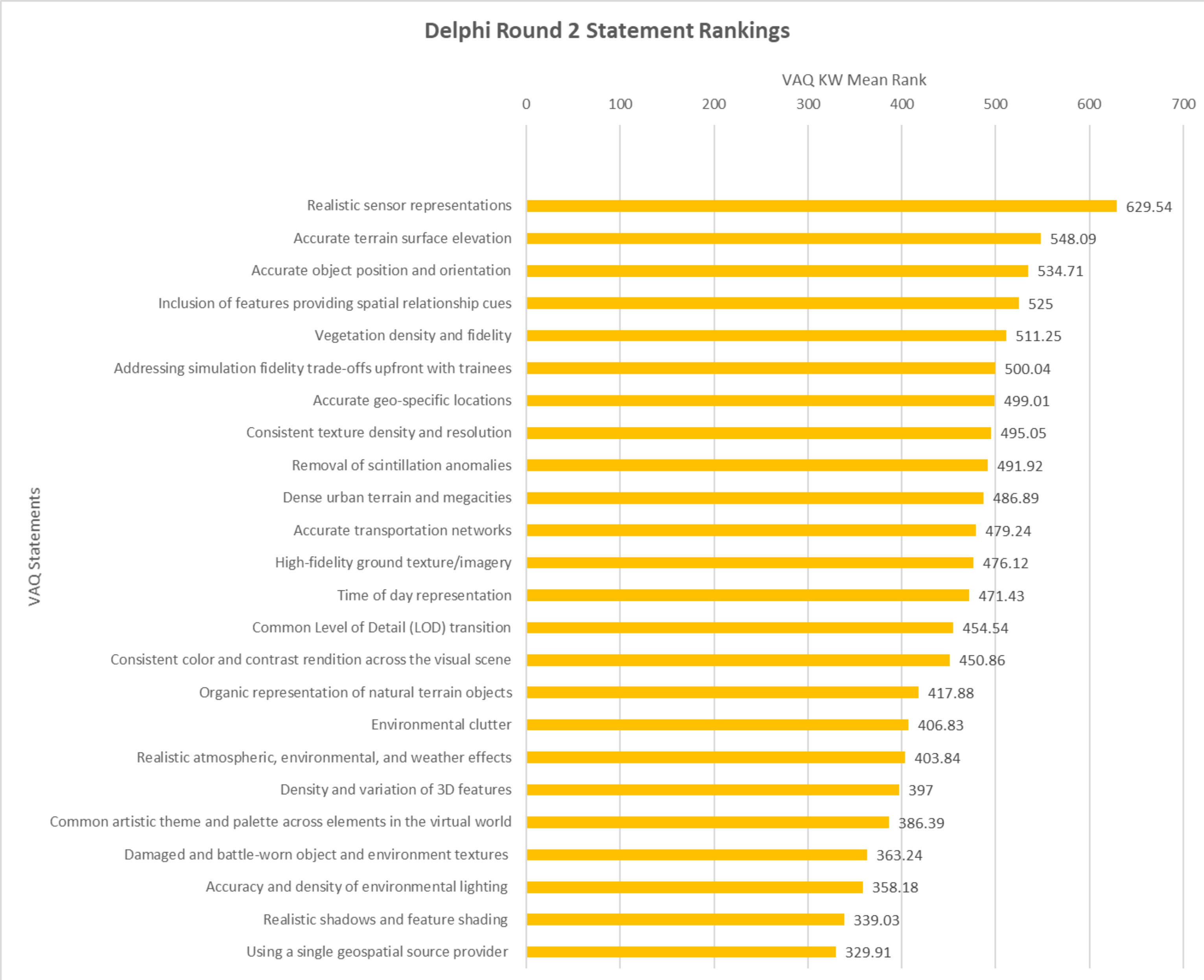
In conducting this protocol, you are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system.

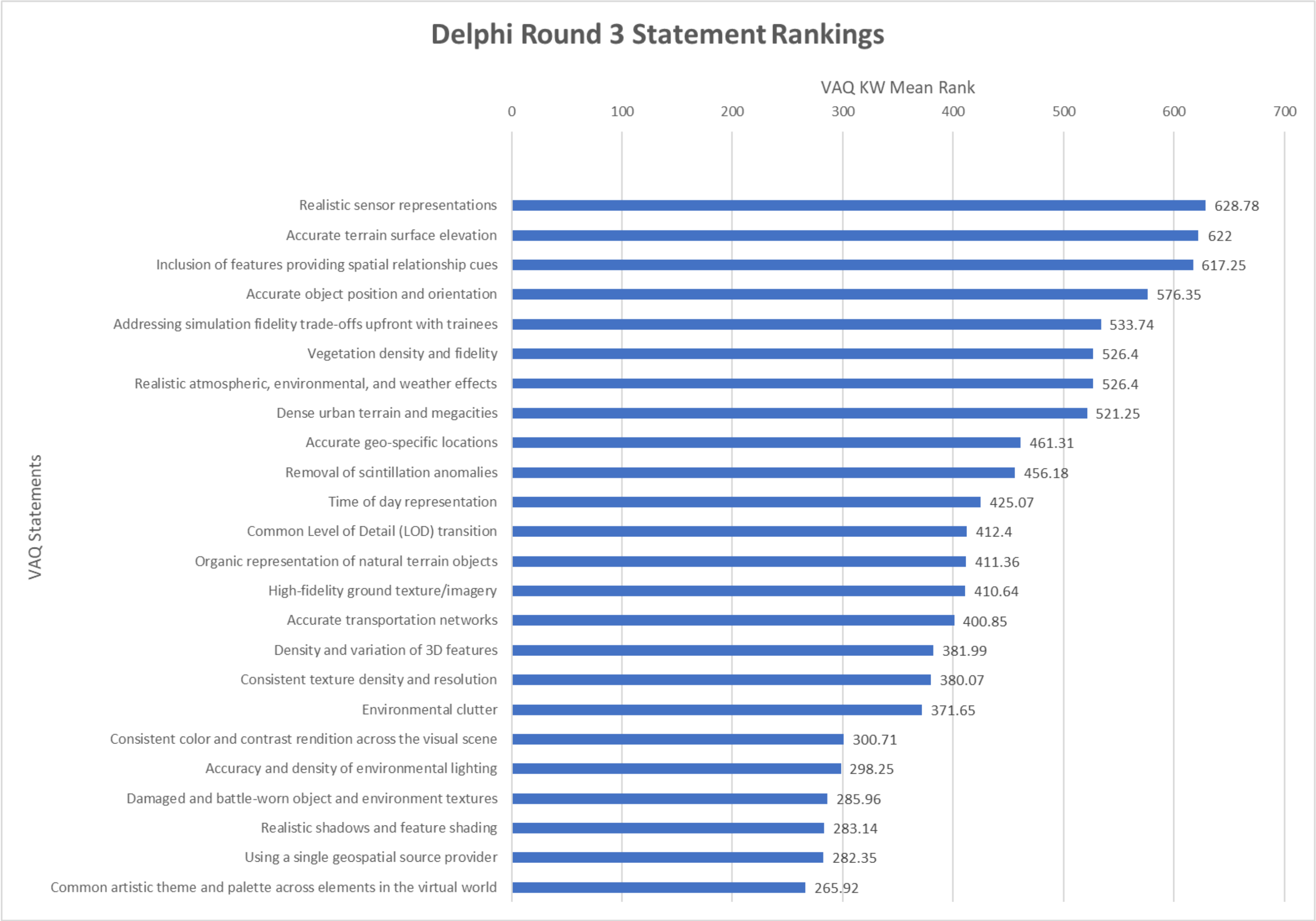
If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

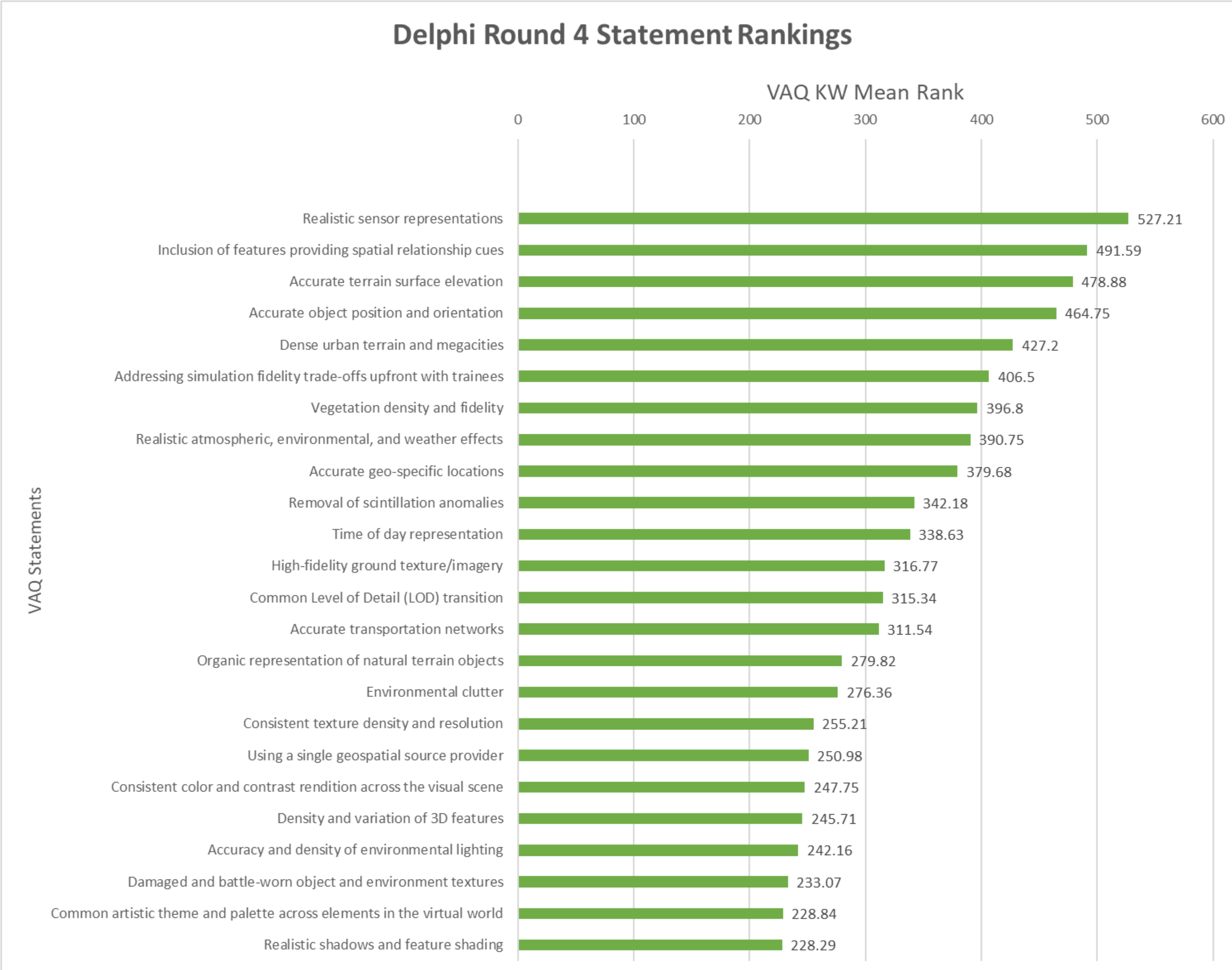
Sincerely,

Renea Carver
Designated Reviewer

APPENDIX F: VAQ STATEMENT IMPORTANCE RANKING CHARTS FROM DELPHI STUDY


Delphi Round 2 Statement Rankings
VAQ KW Mean Rank
0
100
200
300
400
500
600
700
VAQ Statements
Realistic sensor representations
629.54
Accurate terrain surface elevation
548.09
Accurate object position and orientation
534.71
Inclusion of features providing spatial relationship cues
525
Vegetation density and fidelity
511.25
Addressing simulation fidelity trade-offs upfront with trainees
500.04
Accurate geo-specific locations
499.01
Consistent texture density and resolution
495.05
Removal of scintillation anomalies
491.92
Dense urban terrain and megacities
486.89
Accurate transportation networks
479.24
High-fidelity ground texture/imagery
476.12
Time of day representation
471.43
Common Level of Detail (LOD) transition
454.54
Consistent color and contrast rendition across the visual scene
450.86
Organic representation of natural terrain objects
417.88
Environmental clutter
406.83
Realistic atmospheric, environmental, and weather effects
403.84
Density and variation of 3D features
397
Common artistic theme and palette across elements in the virtual world
386.39
Damaged and battle-worn object and environment textures
363.24
Accuracy and density of environmental lighting
358.18
Realistic shadows and feature shading
339.03
Using a single geospatial source provider
329.91


Delphi Round 3 Statement Rankings
VAQ KW Mean Rank
0
100
200
300
400
500
600
700
VAQ Statements
Realistic sensor representations
628.78
Accurate terrain surface elevation
622
Inclusion of features providing spatial relationship cues
617.25
Accurate object position and orientation
576.35
Addressing simulation fidelity trade-offs upfront with trainees
533.74
Vegetation density and fidelity
526.4
Realistic atmospheric, environmental, and weather effects
526.4
Dense urban terrain and megacities
521.25
Accurate geo-specific locations
461.31
Removal of scintillation anomalies
456.18
Time of day representation
425.07
Common Level of Detail (LOD) transition
412.4
Organic representation of natural terrain objects
411.36
High-fidelity ground texture/imagery
410.64
Accurate transportation networks
400.85
Density and variation of 3D features
381.99
Consistent texture density and resolution
380.07
Environmental clutter
371.65
Consistent color and contrast rendition across the visual scene
300.71
Accuracy and density of environmental lighting
298.25
Damaged and battle-worn object and environment textures
285.96
Realistic shadows and feature shading
283.14
Using a single geospatial source provider
282.35
Common artistic theme and palette across elements in the virtual world
265.92



Delphi Round 4 Statement Rankings
VAQ KW Mean Rank
0
100
200
300
400
500
600
VAQ Statements
Realistic sensor representations
527.21
Inclusion of features providing spatial relationship cues
491.59
Accurate terrain surface elevation
478.88
Accurate object position and orientation
464.75
Dense urban terrain and megacities
427.2
Addressing simulation fidelity trade-offs upfront with trainees
406.5
Vegetation density and fidelity
396.8
Realistic atmospheric, environmental, and weather effects
390.75
Accurate geo-specific locations
379.68
Removal of scintillation anomalies
342.18
Time of day representation
338.63
High-fidelity ground texture/imagery
316.77
Common Level of Detail (LOD) transition
315.34
Accurate transportation networks
311.54
Organic representation of natural terrain objects
279.82
Environmental clutter
276.36
Consistent texture density and resolution
255.21
Using a single geospatial source provider
250.98
Consistent color and contrast rendition across the visual scene
247.75
Density and variation of 3D features
245.71
Accuracy and density of environmental lighting
242.16
Damaged and battle-worn object and environment textures
233.07
Common artistic theme and palette across elements in the virtual world
228.84
Realistic shadows and feature shading
228.29

REFERENCES

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*, 888–918.

Akao, Y. (1997). QFD: past, present, and future. *International Symposium on QFD*, *97*(2), 1–12. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:QFD:+Past,+Present,+and+Future#0

Alderton, M. (2019). Fighting on New Frontiers. Retrieved October 5, 2019, from https://trajectorymagazine.com/fighting-on-new-frontiers/

Allen, R. (2011). The unreal enemy of America's army. *Games and Culture*, *6*(1), 38–60. https://doi.org/10.1177/1555412010377321

Altschuld, J. W., & Thomas, P. M. (1991). Considerations in the Application of a Modified Scree Test for Delphi Survey Data. *Evaluation Review*, *15*(2), 179–188. https://doi.org/10.1177/0193841X9101500201

American Supplier Institute. (1989). *Quality Function Deployment Implementation Manual*. Dearborn, MI: American Supplier Institute.

Andrei, S. S. (2012). Efficient terrain & ocean rendering for a real size planet. *ACM SIGGRAPH 2012 Posters, SIGGRAPH'12*, 4503. https://doi.org/10.1145/2342896.2343017

Baca, L. D., & Proctor, M. D. (2017). Friendly casualty outcomes: choice of counter Improvised Explosive Device tactics matters on the asymmetric battlefield. *Journal of Defense Modeling and Simulation*, *14*(3), 303–314. https://doi.org/10.1177/1548512916677837

Balci, O. (1998). Verification, validation, and testing. In J. Banks (Ed.), *The Handbook of Simulation* (pp. 335–393). New York, NY: Wiley.

Barrios, M., Villarroya, A., Borrego, A., & Olle, C. (2011). Response rates and data quality in web and mail surveys administered to PhD holders. *Social Science Computer Review*, *29*(2), 208–220. https://doi.org/10.1177/0894439310368031

Beeland, W. (2002). Student engagement, visual learning and technology: Can interactive whiteboards help? *Annual Confernce of the Association of Information Technology for Teaching Education*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.135.3542&rep=rep1&type=pdf

Bloch, P. H. (1995). Seeking the Ideal Form: Product Design and Consumer Response. *Journal of Marketing*, *59*(3), 16. https://doi.org/10.2307/1252116

Bohemia Interactive Simulations. (2015). BISim at I/ITSEC 2015: See the Next Generation VBS Framework. Retrieved January 1, 2016, from https://www.youtube.com/watch?v=fisi0gsKFQQ

Breillatt, A. (2009). You Can't Innovate Like Apple.

Brown, B. (1968). *Delphi Process: A methodology Used for the Elicitation of Opinions of*

Experts. Santa Monica, CA. Retrieved from https://www.rand.org/pubs/papers/P3925.html

Calvin, J., Dickens, A., Gaines, B., Metzger, P., Miller, D., & Owen, D. (1993). The SIMNET virtual world architecture. In *Proceedings of IEEE Virtual Reality Annual International Symposium* (pp. 450–455). IEEE.

Chief of Staff of the Army Strategic Studies Group. (2014). Megacities and the United States Army, (June). Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwijl9DksaXMAhUIWCYKHZxYBBgQFggcMAA&url=http%3A%2F%2Fwww.army.mil%2Fe2%2Fc%2Fdownloads%2F351235.pdf&usg=AFQjCNEmW6G9O33Cbg9FryOhySNFe9yIow&sig2=poJiUSSoLEk-vv_2eW48aA

Chung, C. a. (2004). *Simulation Modeling Handbook A Practical Approach*. *CRC Press*.

Ciotti, G. (2013). Why Steve Jobs Didn't Listen to His Customers.

Cohen, J. (1992). Quantitative Methods in Psychology. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cox, A., Cairns, P., Shah, P., & Carroll, M. (2012). Not Doing But Thinking: The Role of Challenge in the Gaming Experience. *Proceedings of the CHI 2012*, 79–88. https://doi.org/10.1145/2207676.2207689

Crow, K. (2016). Customer-Focused Development with QFD. Retrieved July 7, 2016, from http://www.npd-solutions.com/qfd.html

Curry, J. (2001). Understanding Conjoint Analysis in 15 Minutes. *Quirk's Marketing Research Review*, *98382*(360), 1–6. Retrieved from http://www.quirks.com/articles/a2001/20010602.aspx?searchID=17327434andsort=9

Cyphert, F., & Gant, W. (1971). The Delphi technique: A case study. *Phi Delta Kappan*, *52*(5), 272-273.

Czupinski, G. W., & Kerska, D. H. (1992). Chrysler LH Powertrain: An Early Automotive QFD Example. In *The 4th Symposium on QFD*. QFD Institute. Retrieved from http://www.qfdi.org/newsletters/chrysler_LHpowertrain_qfd_example.html

Dalkey, N. C., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, *9*(3), 458–467. https://doi.org/10.1287/mnsc.9.3.458

DARPA. (2016). Budget. Retrieved November 9, 2016, from http://www.darpa.mil/about-us/budget

Davis, F. D. (1985). *A Technology Acceptance Model for Empirically Testing New End-User Information Systems*. Massachusetts Institute of Technology.

Davis, F. D. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioural impacts. *International Journal of Man–Machine Studies*, *38*(3), 475–487.

DecisionPro Inc. (2014). *Marketing Engineering for Excel: Conjoint Tutorial*. State College, PA.

Dede, C. (2009). Immersive Interfaces for Engagement and Learning. *Science*, *323*(5910), 66–

69. https://doi.org/10.1126/science.1167311

Defense Acquisition University. (2014). Systems Engineering Process. Retrieved May 6, 2015, from https://acc.dau.mil/CommunityBrowser.aspx?id=638297

Defense Innovation Unit Experimental. (2016). ACCELERATING INNOVATION TO THE WARFIGHTER. Retrieved April 9, 2016, from https://www.diux.mil/

Department of Defense. (2015). *Risk , Issue , and Opportunity Management Guide for Defense Acquisition Programs*. Retrieved from http://bbp.dau.mil/docs/RIO-Guide-Jun2015.pdf

Department of the Army. (2017). *Request for White Paper: Synthetic Training Environment Reconfigurable Virtual Collective Trainers and Global Terrain/One World Terrain*. Orlando, FL. Retrieved from https://govtribe.com/project/request-for-white-paper-synthetic-training-environment-reconfigurable-virtual-collective-trainers-and-global-terrainone-world-terrain

Diener, C., Narang, R., Shant, M., Chander, H., & Goyal, M. (2013). Making Conjoint Mobile: Adapting Conjoint to the Mobile Phenomenon. *Proceedings of the Sawtooth Software Conference*, 55–68.

Dobrian, F., Sekar, V., Awan, A., Stoica, I., Joseph, D., Ganjam, A., … Zhang, H. (2011). Understanding the impact of video quality on user engagement. In *Proceedings of the ACM SIGCOMM 2011 conference on SIGCOMM - SIGCOMM '11* (p. 362). https://doi.org/10.1145/2018436.2018478

DoD Modeling & Simulation Coordination Office. (2011). VV&A Recommended Practice Guide (RPG). Retrieved January 1, 2016, from http://www.msco.mil/vva_rpg.html

DoD Modeling & Simulation Coordination Office. (2013). M&S Glossary. Retrieved from http://msco.mil/MSGlossary.html

Durall, D. (2018). STE and the Digital Revolution. *Infantry*, (Vol. 107 Issue 3), p37-38.

Elliot, J. (2012). *Leading Apple With Steve Jobs: Management Lessons From a Controversial Genius* (1st Ed.). Wiley.

Epic Games, I. (2020). UE4 Features. Retrieved from https://www.unrealengine.com/en-US/features

Evangelista, P. F., Darken, C. J., & Jungkunz, P. (2013). Modeling and integration of situational awareness and soldier target search. *Journal of Defense Modeling and Simulation*, *10*(1), 3–21. https://doi.org/10.1177/1548512911415726

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, *26*(2), 132–139. https://doi.org/10.1016/j.chb.2009.10.015

Ferwerda, J. A. (2003). Three varieties of realism in computer graphics. In *Proceedings SPIE Human Vision and Electronic …* (Vol. SPIE 5007, pp. 290–297). https://doi.org/10.1117/12.473899

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to*

Theory and Research. Reading, MA: Addison-Wesley.

Fleiss, J. L. (1971). Measuring Nominal Scale Agreement Among Many Raters. *Psychological Bulletin*, *76*(5), 378–382.

Fujimoto, R. M., & Weatherly, R. M. (1996). Time Management in the DoD High Level Architecture. In *Proceedings of the Tenth Workshop Parallel & Distributed Simulation* (pp. 60–67). IEEE.

Gill, F. J., Leslie, G. D., Grech, C., & Latour, J. M. (2013). Using a web-based survey tool to undertake a Delphi study: Application for nurse education research. *Nurse Education Today*, *33*(11), 1322–1328. https://doi.org/10.1016/j.nedt.2013.02.016

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research*, *42*(3), 237–288. https://doi.org/10.3102/00346543042003237

Goldiez, B. F. (1996). Techniques For Interoperability Between Terrain Data Bases. *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*. Retrieved from http://www.iitsecdocs.com/download/1996/GU50650216W34741

Gordon, T. J. (2009). The Delphi Method. *Futures Research Methodology v3.0 [CD-ROM]*. Retrieved from https://www.researchgate.net/profile/Harshvardhan_Singh7/post/How_to_develop_a_proper_Delphi_questionnaire/attachment/59d62ec779197b807798cf63/AS%3A355517395423233%401461773360316/download/2+Delphi.pdf

Graeco-Latin Square Design. (2008). In *The Concise Encyclopedia of Statistics*. New York, NY: Springer.

Graniela, B., & Proctor, M. D. (2012). A network-centric terrain database regeneration architecture. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*. Retrieved from http://dms.sagepub.com/content/early/2012/05/31/1548512912444178.full.pdf

Graniela, Benito. (2011). *HARMONY: An Architecture for Network Centric Heterogenous Terrain Database Re-Generation*. University of Central Florida.

Green, B., Jones, M., Hughes, D., & Williams, A. (1999). Applying the Delphi technique in a study of GPs' information requirements. *Health and Social Care in the Community*, *7*(3), 198–205. https://doi.org/10.1046/j.1365-2524.1999.00176.x

Green, P. (1982). The Content of a College-Level Outdoor Leadership Course. In *Conference of the Northwest District Association for the American Alliance for Health, Physical Education, Recreation and Dance* (p. 26).

Green, P. E. (1974). On the Design of Choice Experiments Involving Multifactor Alternatives. *Journal of Consumer Research*, *1*(2), 61–68. Retrieved from www.jstor.org/stable/2489108

Green, P. E., & Rao, V. R. (1971). Conjoint Measurement for Quantifying Judgmental Data. *Journal of Marketing Research*, *8*(3), 355. https://doi.org/10.2307/3149575

Green, P. E., & Wind, Y. (1975). New way to measure consumers' judgments. *Harvard Business*

Review, *53*(4), 107–117. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=3867280&site=ehost-live

Hackett, M., & Proctor, M. (2016). Three-Dimensional Display Technologies for Anatomical Education: A Literature Review. *Journal of Science Education and Technology*, *25*(4), 641–654. https://doi.org/10.1007/s10956-016-9619-3

Hartmann, J., Sutcliffe, A., & Angeli, A. De. (2008). Towards a theory of user judgment of aesthetics and user interface quality. *ACM Transactions on Computer-Human Interaction*, *15*(4), 1–30. https://doi.org/10.1145/1460355.1460357

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*(4), 1008–1015. https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x

Herz, J. C., & Macedonia, M. R. (2002). Computer Games and the Military: Two Views. *Defence Horizons*, (11), 1–8. Retrieved from https://www.files.ethz.ch/isn/135079/DH11.pdf

Hill, K. Q., & Fowles, J. (1975). The methodological worth of the Delphi forecasting technique. *Technological Forecasting and Social Change*, *7*(2), 179–192. https://doi.org/10.1016/0040-1625(75)90057-8

Hodgins, J., Jörg, S., O'Sullivan, C., Park, S. Il, & Mahler, M. (2010). The saliency of anomalies in animated human characters. *ACM Transactions on Applied Perception*, *7*(4), 1–14. https://doi.org/10.1145/1823738.1823740

Hsu, C.-C., & Sanford, B. A. (2007). The Delphi Technique: Making Sense Of Consensus. *Proactical Assessment, Research & Evaluation*, *12*(10). Retrieved from https://pareonline.net/pdf/v12n10.pdf

Huber, J. (1997). What We Have Learned From 20 Years of Conjoint Research: When to Use Self-Explicated, Graded Pairs, Full Profiles or Choice Experiments. *Proceedings of the Sawtooth Software Conference*, 243–257.

IEEE Computer Society. (2012). *IEEE Standard for Distributed Interactive Simulation—Application Protocols (IEEE Std 1278.1-2012)*. *Institute of Electrical and Electronics Engineers*. New York, NY. Retrieved from https://ieeexplore.ieee.org/servlet/opac?punumber=6587042

Isaacson, W. (2015). *Steve Jobs* (2nd Ed). Simon & Schuster.

Jacobs, E. L. (1999). Sampling criteria for sensor simulation. *Optical Engineering*, *38*(5), 827. https://doi.org/10.1117/1.602280

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, *38*(12), 1217–1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: a simulation study of categorization error. *American Sociological Review*, *48*(3), 398–407. https://doi.org/10.2307/2095231

Johnson, M. V. R. S., McKeon, M. F., & Szante, T. R. (1998). *Simulation Based Acquisition: A*

New Approach. Fort Belvoir, VA. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a359264.pdf

Joint Research Centre. (2006). Delphi Survey. Retrieved from http://forlearn.jrc.ec.europa.eu/guide/2_scoping/meth_delphi.htm

Juran, J., & Godfrey, A. B. (1999). *Juran's Quality Handbook* (Fifth Ed.). McGraw-Hill.

Kahney, L. (2012). *Inside Steve's Brain* (Second Ed.). Portfolio.

Kalaian, S. A., & Kasim, R. M. (2012). Terminating Sequential Delphi Survey Data Collection - Practical Assessment, Research & Evaluation. *Practical Assessment, Research & Evaluation*, *17*(5). Retrieved from https://eric.ed.gov/?id=EJ977579

Kang, Y., Kim, H., & Han, S. (2015). Visualization of the Synthetic Environment Data Representation & Interchange Specification data for verifying large-scale synthetic environment data. *Journal of Defense Modeling and Simulation*, *12*(4), 507–518. https://doi.org/10.1177/1548512914548601

Keeney, S., Hasson, F., & Mckenna, H. (2006). Consulting the oracle : ten lessons from using the Delphi technique in nursing research. *Journal of Advanced Nursing*, *53*(2), 205–212. https://doi.org/10.1111/j.1365-2648.2006.03716.x

Kehr, T., Godwin, T., & Mcintire, R. (2014). A Paradigm Shift in the Test and Evaluation of Terrain Databases, (14200), 1–13.

Kenyon, H. (2016). Virtual Training Would Let Army Sharpen Skills, Prep for New Missions. Retrieved November 10, 2016, from http://www.defensenews.com/articles/virtual-training-would-let-army-sharpen-skills-prep-for-new-missions

Kessels, R. (2016). Case Studies on Designing and Analysing Discrete Choice Experiments using JMP ®. In *Europe Discovery Summit*. Antwerp.

Khrnos Group. (2018). Depth Test. Retrieved May 25, 2020, from https://www.khronos.org/opengl/wiki/Depth_Test

King, W. C., Hill, A., & Orme, B. K. (2004). The "Importance" Question in ACA: Can it be Omitted? In *Proceedings of the Sawtooth Software Conference* (pp. 53–65).

Kooi, F. L., & Toet, A. (2005). What's crucial in night vision goggle simulation. *Enhanced and Synthetic Vision 2005*, *5802*(May 2005), 37–46. https://doi.org/10.1117/12.601432

Ladner, R., & Shaw, K. (2001). An Overview of 3D Synthetic Environment Construction. In *3D Synthetic Environment Reconstruction*. https://doi.org/10.1007/978-1-4419-8756-3

Lalonde, B. (2008). Converging towards synthetic environment interoperability. *Aeronautical Journal*, *112*(1129), 171–177. https://doi.org/10.1017/S0001924000002104

Lashinsky, A. (2012). *Inside Apple: How America's Most Admired--and Secretive--Company Really Works*. (First Ed., Ed.). Business Plus.

Lashlee, D., Bricio, J., Holcomb, R., & Richards, W. T. (2012). GIS Data for Combat Modeling. In A. Tolk (Ed.), *Engineering Principles of Combat Modeling and Distributed Simulation* (First Edit). Hoboken, New Jersey: John Wiley & Sons, Inc.

Latour, J. M., Hazelzet, J. A., Duivenvoorden, H. J., & van Goudoever, J. B. (2009). Construction of a parent satisfaction instrument: Perceptions of pediatric intensive care nurses and physicians. *Journal of Critical Care*, *24*(2), 255–266. https://doi.org/10.1016/j.jcrc.2008.06.002

Lenth's Analysis of Unreplicated Factorial Experiments. (2011). Retrieved from https://scss.tcd.ie/postgraduate/pgcertstats/current/Lecture Notes/2011-12/7003/Lenth'sAnalysis.pdf

Lenth, R. V. (1989). Quick and easy analysis of unreplicated factorials. *Technometrics*, *31*(4), 469–473. https://doi.org/10.1080/00401706.1989.10488595

Lockamy, A., & Khurana, A. (1995). Quality Function Deployment: a Case Study. *Production and Inventory Management Journal*, *36*(2), 56.

Loper, M., & Turnitsa, C. (2012). History of Combat Modeling and Distributed Simulation. In A. Tolk (Ed.), *Engineering Principles of Combat Modeling and Distributed Simulation* (pp. 331–356). Hoboken, New Jersey: Wiley.

Madu, C. N. (2006). *House of Quality in a Minute* (2nd Editio). Fairfield, CT: Chi.

Mamaghani, F. (1995). Creation and usage of synthetic environments in realtime networked interactive simulation. In *SPIE's 1995 Symposium on OE/Aerospace Sensing and Dual Use Photonics* (Vol. 2486, pp. 186–195). International Society for Optics and Photonics. Retrieved from http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1000327

Mamaghani, F. (2008). An Introduction to SEDRIS. In *Proceedings of the 2008 Interservice/Industry Training Systems and Education Conference (I/ITSEC)* (p. 95). Orlando, FL: NTSA. Retrieved from https://www.sedris.org/iitsec_2008/sld001.htm

Manktelow, J., Jackson, K., Swift, C., Edwards, S., Bishop, L., Pearcey, E., … Robinson, R. (2018). Avoiding Groupthink: Avoiding Fatal Flaws in Group Decision Making. Retrieved April 4, 2018, from https://www.mindtools.com/pages/article/newLDR_82.htm

Martin, G. A., Jewett, R. M., Hollander, C., & Hicks, C. (2007). Dead Reckoning on the GPU : A Comparative Study vs . the CPU. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)* (pp. 1–7).

McKenna, H. P. (1994). The Delphi technique: a worthwhile research approach for nursing? *Journal of Advanced Nursing*, *19*(6), 1221–1225. https://doi.org/10.1111/j.1365-2648.1994.tb01207.x

Montgomery, D. C. (2000). *Design and Analysis of Experiments*. *Book* (5th ed.). New York: John Wiley & Sons, Inc. https://doi.org/10.1198/tech.2006.s372

Montijo, G. a, Kaiser, D., International, C. T., Spiker, V. A., Sciences, A., Barbara, S., … Force, A. (2008). Training Interventions for Reducing Flight Mishaps. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2008* (pp. 1–12).

Moorthy, A. K., & Bovik, A. C. (2011). Visual quality assessment algorithms: What does the future hold? *Multimedia Tools and Applications*, *51*(2), 675–696. https://doi.org/10.1007/s11042-010-0640-x

Mourkoussis, N., Rivera, F. M., Troscianko, T., Dixon, T., Hawkes, R., & Mania, K. (2010). Quantifying fidelity for virtual environment simulations employing memory schema assumptions. *ACM Transactions on Applied Perception*, *8*(1), 1–21. https://doi.org/10.1145/1857893.1857895

Nielsen, B., & Schmidt, M. (1990). Performing Conjoint Analysis Using PROC TRANSREG. In *SEUGI '90 : Proceedings of the SAS European Users Group International Conference* (pp. 357–362). Heidelberg: SAS Institute. Retrieved from https://research.cbs.dk/en/publications/performing-conjoint-analysis-using-proc-transreg

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Orme, B. (2010). Traditional Conjoint Analysis with Excel. *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*, 67–75. Retrieved from http://www.sawtoothsoftware.com/support/technical-papers/general-conjoint-analysis/analysis-of-traditional-conjoint-using-excel-an-introductory-example-2009

Orme, B. K. (2010a). A Short History of Conjoint Analysis. *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*, 29–37.

Orme, B. K. (2010b). *Getting Started with Conjoint Analysis* (2nd Ed.). Madison, WI: Research Publishers LLC.

Orme, B. K. (2010c). Sample Size Issues for Conjoint Analysis. *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*, 57–66.

Orme, B. K. (2013). *Which Conjoint Method Should I Use? Sawtooth Software RESEARCH PAPER SERIES*. Retrieved from http://www.sawtoothsoftware.com/download/techpap/cbctech.pdf

Palter, V. N., Graafland, M., Schijven, M. P., & Grantcharov, T. P. (2012). Designing a proficiency-based, content validated virtual reality curriculum for laparoscopic colorectal surgery: A Delphi approach. *Surgery*, *151*(3), 391–397. https://doi.org/10.1016/j.surg.2011.08.005

Peck, M. (2012). Forget Call of Duty and Battlefield. Real Military Training Needs a Different Kind of Game. Retrieved August 9, 2016, from http://kotaku.com/5883105/forget-call-of-duty-and-battlefield-real-military-training-needs-a-different-kind-of-game

Peck, Michael. (2015). Army Developing Military-Grade 'Google Earth.' Retrieved January 1, 2016, from https://www.govtechworks.com/army-developing-military-grade-google-earth/#gs.Mj8vCcw

Persson, E. (2012). Creating vast game worlds: Experiences from Avalanche studios. *ACM SIGGRAPH 2012 Talks,SIGGRAPH'12*, 4503. https://doi.org/10.1145/2343045.2343089

Petty, M. D. (2010). Verification, Validation, and Accreditation. In J. A. Sokolowski & C. M. Banks (Eds.), *Modeling and Simulation Fundamentals* (pp. 325–372). Wiley.

Piorkowski, R., Mantiuk, R., & Siekawa, A. (2017). Automatic Detection of Game Engine Artifacts Using Full Reference Image Quality Metrics. *ACM Transactions on Applied*

Perception, *14*(3), 14:1--14:17. https://doi.org/10.1145/3047407

Prasithsangaree, P., Manojlovich, J., Hughes, S., & Lewis, M. (2004). UTSAF: A Multi-Agent-Based Software Bridge for Interoperability between Distributed Military and Commercial Gaming Simulation. *Simulation*, *80*(12), 647–657. https://doi.org/10.1177/0037549704050907

Proctor, M. D., & Paulo, G. (1996). Modeling in support of operational testing. *Mathematical and Computer Modelling*, *23*(1–2), 9–14. https://doi.org/10.1016/0895-7177(95)00211-1

Proctor, Michael D., & Campbell-Wynn, L. (2014). Effectiveness, Usability, and Acceptability of Haptic-Enabled Virtual Reality and Mannequin Modality Simulators for Surgical Cricothyroidotomy. *Military Medicine*, *179*(3), 260–264. https://doi.org/10.7205/milmed-d-13-00365

Proust, M. (2018). *JMP 14 Design of Experiments Guide*. Cary, NC: SAS Institute. Retrieved from https://support.sas.com/documentation/onlinedoc/jmp/14.0/DOE-Guide.pdf

Pullman, M. E. (Colorado S. U., Moore, W. L. (University of U., & Wardell, D. G. (University of U. (2002). A comparison of quality function deployment and conjoint analysis in new product design. *Journal of Product Innovation Management*. https://doi.org/10.1016/S0737-6782(02)00152-2

Purdy, R., & Goldiez, B. (1995). Interoperability Issues Associated With the Use Of Dissimilar Simulators. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*. Retrieved from http://www.iitsecdocs.com/download/1995/335466Q256571403

QFD Insitute. (n.d.). NATO Combat Aircraft Mobility Study. Retrieved May 9, 2016, from http://www.qfdi.org/newsletters/nato_combat_aircraft_mobility_study.html

Rawson, R. (2013). Hiring Strategies of Google, Facebook, Apple and other Tech Giants. Retrieved April 9, 2016, from https://blog.staff.com/hiring-strategies-of-google-facebook-apple-and-other-tech-giants/

Riddle, S. P., & Olejniczak, S. (2000). The Advanced Technology Crew Station (ATCS) Design Methodology: A Crew-Centered Approach. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. Retrieved from http://www.iitsecdocs.com/download/2000/A2R57C8EX78DW7HC

Santiago, F., Verdesca, M., Watkins, J., & de la Cruz, J. (2012). Geospatial Correlation Testing Framework and Toolset. *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, *2012*. Retrieved from http://www.iitsecdocs.com/download/2012/9N5718220V6213R5

Sawtooth Software. (2002). Conjoint Value Analysis (CVA) Version 3.0. *Sawtooth Software Technical Papers Series*.

Scheibe, M., Skutsch, M., & Schofer, J. (1975). Experiments in Delphi methodology. In H. A. Lindstone & M. Turoff (Eds.), *The Delphi method: techniques and applications* (pp. 262–287). Reading, MA: Addison-Wesley Pub. Co.

Schiavone, G. A., Nelson, R. S., & Goldiez, B. (1994). Statistical Certification of Terrain

Databases. In *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*. Retrieved from http://www.iitsecdocs.com/download/1994/8J45516668K331PT

Schnitzer, A. P. (1976). A data base generation system for digital image generation. In *Proceedings Ninth NTEC Conference* (pp. 103–114). Retrieved from http://www.iitsecdocs.com/download/1976/F16WTQ5J06155006

Scoresby, J., & Shelton, B. E. (2011). Visual perspectives within educational computer games: Effects on presence and flow within virtual immersive learning environments. *Instructional Science*, *39*(3), 227–254. https://doi.org/10.1007/s11251-010-9126-5

SE Core. (2013). I/ITSEC'13 State of SE Core Brief. U.S. Army PEO STRI. Retrieved from Non-Public Domain Reference

SEDRIS. (2007). SEDRIS Glossary. Retrieved January 1, 2016, from http://www.sedris.org/glossary.htm

Sellers, G., Van Waveren, J. M. P., Cozzi, P., Ring, K., Persson, E., & De Vahl, J. (2013). Rendering massive virtual worlds. *ACM SIGGRAPH 2013 Courses, SIGGRAPH 2013*. https://doi.org/10.1145/2504435.2504458

Shah, H. A., & Kalaian, S. A. (2009). Which Is the Best Parametric Statistical Method For Analyzing Delphi Data? *Journal of Modern Applied Statistical Methods*, *8*(1), 226–232. https://doi.org/10.22237/jmasm/1241137140

Shen, Z., & Zhou, S. (2006). Behavior Representation and Simulation for Military Operations on Urbanized Terrain. *Simulation*, *82*(9), 593–607. https://doi.org/10.1177/0037549706072093

Shepard, S. B. (1998, May). Apple - Steve Jobs. *BusinessWeek*.

Shufelt, J. W. J. (2006). A vision for future virtual training. *Nato: Rto-Mp-Hfm-136*, (October 2002), 4–15. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA473302

Slater, M. (1999). Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoper. Virtual Environ.*, *8*(5), 560–565. https://doi.org/10.1086/250095

Smelik, R., van Wermeskerken, F., Krijnen, R., & Kuijper, F. (2019). Dynamic synthetic environments: a survey. *Journal of Defense Modeling and Simulation*, *16*(3), 255–271. https://doi.org/10.1177/1548512918811954

Smith, R. (2006). Technology Disruption in the Simulation Industry. *The Journal of Defense Modeling and Simulation: Applications*, *Methodology*, *Technology*, *3*(1), 3–10. https://doi.org/10.1177/875647930600300102

Smith, R. D. (2009). *Military Simulation and Serious Games* (First Edit). Modelbenders Press.

Stalder, J. (2016). Army launches Rapid Capabilities Office. Retrieved April 9, 2016, from https://www.army.mil/article/174290/army_launches_rapid_capabilities_office

Standards Committee on Interactive Simulation. (1995). IEEE Standard for Distributed Interactive Simulation — Application Protocols (IEEE Std 1278.1-1995). IEEE Computer

Society. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=499701&isnumber=10849

STE, OTW, OTA, and TReX. (2019). Retrieved from https://militarysimulation.training/articles/ste-otw-ota-and-trex/

Stevens, J., Kincaid, P., & Sottilare, R. (2015). Visual modality research in virtual and mixed reality simulation. *Journal of Defense Modeling and Simulation*, *12*(4), 519–537. https://doi.org/10.1177/1548512915569742

Sullivan, G. M., & Artino, A. R. (2013). Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education*, *5*(4), 541–542. https://doi.org/10.4300/jgme-5-4-18

Sumison, T. (1998). The Delphi Technique: An Adaptive Research Tool. *British Journal of Occupational Therapy*, *61*(4), 153–156.

SurveyMonkey. (2016a). Ranking Question. Retrieved July 24, 2016, from http://help.surveymonkey.com/articles/en_US/kb/How-do-I-create-a-Ranking-type-question

SurveyMonkey. (2016b). Security Statement. Retrieved July 16, 2016, from https://www.surveymonkey.com/mp/policy/security/

Tang, J., & Grenville, A. (2013). Can Conjoint be Fun?: Improving Respondent Engagement in CBC Experiments. *Proceedings of the Sawtooth Software Conference*, 39–52.

The Pennsylvania State University. (2018). The Latin Square Design. Retrieved from https://newonlinecourses.science.psu.edu/stat503/node/21/

Toet, A. (1998). Visual conspicuity determines human target acquisition performance. *Optical Engineering*, *37*(7), 1969. https://doi.org/10.1117/1.601903

Tolk, A. (2012a). Modeling the Environment. In A. Tolk (Ed.), *Engineering Principles of Combat Modeling and Distributed Simulation* (First Edit). Hoboken, New Jersey: John Wiley & Sons, Inc.

Tolk, A. (2012b). Terms and Application Domains. In A. Tolk (Ed.), *Engineering Principles of Combat Modeling and Distributed Simulation* (First Edit). Hoboken, New Jersey: John Wiley & Sons, Inc.

Tolk, A. (2012c). Verification and Validation. In A. Tolk (Ed.), *Engineering Principles of Combat Modeling and Distributed Simulation* (First Edit, pp. 263–294). Hoboken, New Jersey: John Wiley & Sons, Inc.

U.S. Army. (2016). *PE 0604780A: Combined Arms Tactical Trainer (CATT)*. Retrieved from http://www.i2insights.com/library/defense_budget-documents/fy2017-defense_budget/2040A/2040A-0604780A-R.pdf

U.S. Army PEO STRI. (2015). SE Core Standard Terrain Database Generation Capability (STDGC) Diagram. Orlando, FL: US Army. Retrieved from Non-Public Domain Reference

U.S. Army PEO STRI. (2016a). Assistant Project Manager for Synthetic Environment Core (SE) CORE. Retrieved November 10, 2016, from http://www.peostri.army.mil/PM-

ITE/SECore.jsp

U.S. Army PEO STRI. (2016b). SE Core Internal Program Management Documentation. Orlando, FL: APM SE Core.

U.S. Army PEO STRI. (2017a). Synthetic Environment Core (SE Core): 3D Geospatial M&S Summit Follow On. Orlando, FL. Retrieved from https://portal.opengeospatial.org/files/?artifact_id=74238

U.S. Army PEO STRI. (2017b). Synthetic Environment Core (SE Core) 3D Geospatial M&S Summit Follow On. Retrieved from file:///C:/Users/thomas.kehr/Downloads/PEO-STRI_Presentation_Releasable.pdf

U.S. Army STE Cross-Functional Team (CFT). (n.d.). Draft Common Synthetic Environment (CSE) Statement of Need. NSTXL.org. Retrieved from https://trainingaccelerator.org/wp-content/uploads/2019/03/Draft-CSE-SoN_28Feb.pdf

U.S. Army Training and Doctrine Command. (2010). *The United States Army Operating Concept 2016-2028*. Retrieved from http://www.tradoc.army.mil/tpubs/pams/tp525-3-1.pdf

Ulschak, F. L. (1983). Human resource development : the theory and practice of need assessment (pp. 111–131). Reston, VA: Reston Publishing company, Inc.

Under Secretary of Defense. (2015). *Implementation Directive for Better Buying Power 3.0*. Retrieved from http://www.acq.osd.mil/fo/docs/betterBuyingPower3.0(9Apr15).pdf

Under Secretary of Defense for Acquisition and Technology. (1995). *Modeling and Simulation (M&S) Master Plan*. Washington, D.C.

United States Geological Survey. (2020a). 3D Elevation Program. Retrieved from https://www.usgs.gov/core-science-systems/ngp/3dep

United States Geological Survey. (2020b). National Geospatial Program. Retrieved from https://www.usgs.gov/core-science-systems/national-geospatial-program/national-map

Veziridis, S., Karampelas, P., & Lekea, I. (2017). Learn by Playing. *IEEE Global Engineering Education Conference (EDUCON)*, (April), 920–925.

Viera, A. J., & Garrett, J. M. (2005). Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*, *37*(5), 360–363. https://doi.org/Vol. 37, No. 5

Vilikus, O. (n.d.). Optimization of Sample Size and Number of Tasks Per Respondent in Conjoint Studies Using.

Vries, J. de. (2020). PBR Theory. Retrieved from https://learnopengl.com/PBR/Theory

Wainer, G. A., & Al-Zoubi, K. (2010). An Introduction to Distributed Simmulation. In J. Sokolowski & C. Banks (Eds.), *Modeling and Simulation Fundamentals* (pp. 373–02). Suffolk, VA: John Wiley & Sons, Inc.

Warfare History Network. (2014). Video Game Training: How Do Games Compare to Military Simulations? Retrieved from http://warfarehistorynetwork.com/daily/military-games/video-game-training-how-do-games-compare-to-military-simulations/

Whitton, M., & Loftin, R. B. (2009). Virtual Environment Component Technologies. In *The PSI*

Handbook of Virtual Environments for Training and Education. Greenwood Publishing Group.

Winkler, S. (2001). Visual fidelity and perceived quality: toward comprehensive metrics. In *Proceedings of SPIE - The International Society for Optical Engineering* (Vol. 4299, pp. 114–125). https://doi.org/10.1117/12.429540

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoper. Virtual Environ.*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Wolniak, R. (2018). The use of QFD method advantages and limitation. *Production Engineering Archives*, *18*(18), 14–17. https://doi.org/10.30657/pea.2018.18.02

Yamamoto, C., Hara, F., Kishi, K., & Satoh, K. (2005). Using Quality Function Deployment To Evaluate Government Services From the Customer ’ S Perspective. *Journal of the Eastern Asia Society for Transportation Studies, 6*, 4160–4175.

Zevin, B., Levy, J. S., Satava, R. M., & Grantcharov, T. P. (2012). A consensus-based framework for design, validation, and implementation of simulation-based training curricula in surgery. *Journal of the American College of Surgeons*, *215*(4), 580–586. https://doi.org/10.1016/j.jamcollsurg.2012.05.035

Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology/Psychologie Canadienne*, *34*(4), 390–400. https://doi.org/10.1037/h0078865