Naval Postgraduate School
Dudley Knox Library

ENHANCING KNOWLEDGE FLOW WITH EXTENDED REALITY

Title	ENHANCING KNOWLEDGE FLOW WITH EXTENDED REALITY
Item Type	Thesis
Authors	Brewster, Jon M.
URI	https://hdl.handle.net/10945/75068
Publisher	Monterey, CA; Naval Postgraduate School
Date Issued	2025-12
Rights	This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.
Download date	2026-09-29 14:05:08
Link to Item	https://hdl.handle.net/10945/75068

Downloaded from NPS Archive: Calhoun

Calhoun
Institutional Archive of the Naval Postgraduate School

Calhoun is the Naval Postgraduate School's repository for research materials and institutional publications created by the NPS community. Materials in Calhoun are openly accessible and will be preserved for future generations.

Dudley Knox Library, Naval Postgraduate School
411 Dyer Rd. Bldg. 339, Monterey, CA 93943
https://library.nps.edu/



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

DISSERTATION

ENHANCING KNOWLEDGE FLOW WITH EXTENDED REALITY

by

Jon M. Brewster

December 2025

Dissertation Supervisor: Mollie R. McGuire

Distribution Statement A. Approved for public release: Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC, 20503.

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** December 2025	**3. REPORT TYPE AND DATES COVERED** Dissertation

4. TITLE AND SUBTITLE ENHANCING KNOWLEDGE FLOW WITH EXTENDED REALITY	**5. FUNDING NUMBERS**
6. AUTHOR(S) Jon M. Brewster	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING/ MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT Distribution Statement A. Approved for public release: Distribution is unlimited.	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Tacit knowledge—embodied, situational, and difficult to articulate—plays a critical role in U.S. Navy bridge operations, yet it remains challenging to transfer through resource-limited mockup training. This dissertation evaluated whether extended reality (XR) could support tacit knowledge flow by comparing XR-based team training to the Navy's Navigation, Seamanship, and Shiphandling Trainer (NSST). Sponsored by the Office of Naval Research Tech Solutions and executed by Naval Information Warfare Center, Pacific, this study compared more than 17,000 data points taken from 28 bridge teams (158 participants) who completed matched scenarios in both XR and the NSST. A within-subjects design supported over 300 statistical tests, including Wilcoxon and Mann-Whitney analyses, Hotelling's T^2, Cohen's d effect sizes, order effects, subgroup analyses, principal component analysis, sentiment analyses, and distributional fitting. Across the NASA Task Load Index, System Usability Scale, and Bridge-Watch Self-Assessment results supported three of four hypotheses, showing XR produced workload, usability, and readiness outcomes comparable to the NSST. Research findings demonstrated XR is a viable, portable, always-available training solution usable from any location at any time and also advances Knowledge Flow Theory toward future investigations into "Cognition in the Virtual," a new conceptual framework for tacit knowledge flow via synthetic experience across mixed reality environments.

14. SUBJECT TERMS extended reality; XR; Navigation, Seamanship, and Shiphandling Trainer; NSST; Knowledge Flow Theory; tacit knowledge; synthetic experience; mixed reality; MR; National Aeronautics and Space Administration; NASA; Task Load Index; TLX	**15. NUMBER OF PAGES** 331
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Distribution Statement A. Approved for public release: Distribution is unlimited.

ENHANCING KNOWLEDGE FLOW WITH EXTENDED REALITY

Jon M. Brewster
Commander (Ret.), U.S. Navy Reserve
Civilian, Department of the Navy
BA, University of Colorado, Boulder, 1987

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN INFORMATION SCIENCES

from the

NAVAL POSTGRADUATE SCHOOL
December 2025

Approved by:

Mollie R. McGuire
Department of
Information Sciences
Dissertation Supervisor

Alex Bordetsky
Department of
Information Sciences
Dissertation Chair

Don Brutzman
Department of
Information Sciences

Johnathan C. Mun
Department of
Information Sciences

Carl L. Oros
Department of
Information Sciences

Jamie R. Lukos
Naval Information
Warfare Center, Pacific

Approved by:

Alex Bordetsky
Chair, Department of Information Sciences

James B. Michael
Vice Provost of Academic Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Tacit knowledge—embodied, situational, and difficult to articulate—plays a critical role in U.S. Navy bridge operations, yet it remains challenging to transfer through resource-limited mockup training. This dissertation evaluated whether extended reality (XR) could support tacit knowledge flow by comparing XR-based team training to the Navy's Navigation, Seamanship, and Shiphandling Trainer (NSST). Sponsored by the Office of Naval Research Tech Solutions and executed by Naval Information Warfare Center, Pacific, this study compared more than 17,000 data points taken from 28 bridge teams (158 participants) who completed matched scenarios in both XR and the NSST. A within-subjects design supported over 300 statistical tests, including Wilcoxon and Mann-Whitney analyses, Hotelling's T^2, Cohen's d effect sizes, order effects, subgroup analyses, principal component analysis, sentiment analyses, and distributional fitting. Across the NASA Task Load Index, System Usability Scale, and Bridge-Watch Self-Assessment results supported three of four hypotheses, showing XR produced workload, usability, and readiness outcomes comparable to the NSST. Research findings demonstrated XR is a viable, portable, always-available training solution usable from any location at any time and also advances Knowledge Flow Theory toward future investigations into "Cognition in the Virtual," a new conceptual framework for tacit knowledge flow via synthetic experience across mixed reality environments.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I. INTRODUCTION 1
A. BACKGROUND 2
B. RESEARCH FOCUS 8
C. PROBLEM STATEMENT AND PURPOSE 12
D. RESEARCH QUESTION AND SCOPE 12
E. LIMITATIONS AND SAFETY CONSIDERATIONS 13
F. ORGANIZATION OF THE DISSERTATION 14

II. LITERATURE REVIEW 17
A. KNOWLEDGE FLOW 17
1. Philosophical Distinctions in Knowledge 17
2. Knowledge Management 19
3. Knowledge Distinctions in Knowledge Management 20
4. Dynamic Knowledge Flow Model 20
5. Knowledge Flow and Information Technology 22
B. THE PROBLEM OF TACIT KNOWLEDGE FLOW 23
1. Tacit Knowledge and Feedback 23
2. Tacit Knowledge and Knowledge Flow Friction 24
3. Procedural Knowledge and Embodied Cognition 24
4. Embodiment and Tacit Knowledge 25
C. TACIT KNOWLEDGE FLOW AND SYNTHETIC EXPERIENCE 27
1. Mockup-Based Synthetic Experience 27
2. Extended Reality–Based Synthetic Experience 29
3. Research Rationale 31
D. RESEARCH GAPS AND HYPOTHESES 32

III. METHOD 35
A. RESEARCH DESIGN 35
1. Overarching Research Design 36
2. Participants and Sampling Design 39
3. Tasks Requirements Matching in the Experimental Virtual Environment 41
4. Participant Access and Experiment Alignment 42
B. DATA COLLECTION 43
1. Sample Size 43
2. Sampled Participants 43

3. Setting and Apparatus 44
4. Surveys 50
5. Observations 53
6. Procedures 54
7. Institutional Review Board Considerations 60
C. DATA ANALYSIS 61
1. Data Preparation 61
2. Statistical Analysis Design 62
3. Statistical Tests Employed 63

IV. EXPERIMENTAL RESULTS AND ANALYSES 67
A. NATIONAL AERONAUTICS AND SPACE ADMINISTRATION TASK LOAD INDEX SURVEY RESULTS 67
1. Findings Summary 68
2. Descriptive Statistics, Distributional Fit, and Normality 69
3. Data Reliability and Agreement 71
4. Statistical Comparisons: Navigation, Seamanship, Shiphandling Trainer versus Extended Reality 73
5. Order Effects 74
6. Principal Component Analysis 75
7. Power Analysis 78
8. Summary of Key Findings 79
B. SYSTEM USABILITY SCALE SURVEY RESULTS 80
1. Findings Summary 81
2. Descriptive Statistics, Distributional Fit, and Normality 81
3. Data Reliability and Agreement 85
4. Statistical Comparisons: Navigation, Seamanship, Shiphandling Trainer versus Extended Reality 86
5. Order Effects 87
6. Principal Component Analysis 89
7. Power Analysis 93
8. Summary of Key Findings 96
C. MARITIME KNOWLEDGE AND PERFORMANCE RESULTS 97
1. Bridge Watchstander Self-Assessment Survey 98
2. Officer of the Deck Self-Assessment Survey 127
3. Instructor Assessment Survey 152
D. EXIT INTERVIEW SURVEY ANALYSES 162
1. Large Language Model Computed Text Analysis 162

2. Large Language Model Computed Themes 163
3. Text Analysis Filtered by Watchstander 165
4. Word Cloud Graphic of Sentiment Analysis..... 166
5. Convergence of Text Analysis and Statistical Findings..... 167
6. Summary of Text Analysis Findings 168
7. Summary of Large Language Model Text Analysis Limitations..... 168

V. SUMMARY OF RESEARCH FINDINGS..... 171

VI. CONCLUSIONS, CONTRIBUTIONS, AND RECOMMENDATIONS..... 175
A. RESEARCH CONCLUSIONS..... 175
1. National Aeronautics and Space Administration Task Load Index: Cognitive Workload..... 175
2. System Usability Scale: Usability and Interface Expectations..... 176
3. Bridge Watch Team Self-Assessment: Operational Competence and Tacit Processes..... 176
4. Officer of the Deck Self-Assessment: Expert Operational Judgment 176
5. Instructor Assessment: Independent Expert Evaluation 176
6. Integration of Sentiment and Text Analyses 177
7. Hypotheses Evaluation and Theoretical Synthesis 177
8. Closing Thoughts 178
B. RESEARCH CONTRIBUTIONS 178
1. Theory 179
2. Method 194
3. Practice..... 196
C. RECOMMENDATIONS FOR FUTURE RESEARCH..... 203
1. Navy Research and Extended Reality 203
2. Integrated Extended Reality–Artificial Intelligence–Navy Live Virtual Constructive Research Directions 205
3. Empirical Tacit Knowledge and Cognitive Science Research..... 205
4. Advanced Theoretical and Cross-Domain Research 206
5. Summary of Future Research Recommendations 207

APPENDIX A. EXTENDED REALITY BRIDGE VIRTUAL ENVIRONMENT CONFIGURATION DESIGN 209

APPENDIX B. EXTENDED REALITY EXTERNAL VIRTUAL ENVIRONMENT CONFIGURATION AND CONTROL 223

APPENDIX C. COMMANDER NAVAL SURFACE FORCES SEPTEMBER 23, 2024 LETTER OF SUPPORT 225

APPENDIX D. PRE-EVENT COORDINATION EMAIL 227

APPENDIX E. PRIVACY STATEMENT 229

APPENDIX F. INFORMED CONSENT DOCUMENT 231

APPENDIX G. PRE-SCREENING QUESTIONNAIRE 237

APPENDIX H. SIMULATOR SICKNESS QUESTIONNAIRE 239

APPENDIX I. VIRTUAL BRIDGE AND NAUTICAL TRAINER APPARATUS 241

APPENDIX J. NAVAL INFORMATION WARFARE CENTER PACIFIC VIRTUAL BRIDGE AND NAUTICAL TRAINER INSTITUTIONAL REVIEW BOARD PROTOCOL NIWCP.2024.0003-IR-EP7-A NOVEMBER 19, 2024 243

APPENDIX K. NATIONAL AERONAUTICS AND SPACE ADMINISTRATION TASK LOAD INDEX SURVEY 267

APPENDIX L. SYSTEM USABILITY SCALE 269

APPENDIX M. BRIDGE WATCHSTANDER SELF-ASSESSMENT SURVEY . 271

APPENDIX N. OFFICER OF THE DECK SELF-ASSESSMENT SURVEY 275

APPENDIX O. INSTRUCTOR ASSESSMENT SURVEY: BRIDGE WATCH TEAM 279

APPENDIX P. EXIT INTERVIEW QUESTIONS 283

APPENDIX Q. EXIT INTERVIEW WORD CLOUDS 285

LIST OF REFERENCES 291

INITIAL DISTRIBUTION LIST 301

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

Figure 1. Reality–virtuality continuum. Source: Milgram et al. (1995). 3

Figure 2. San Diego Bay Nautical Chart 18773 and extended reality head mounted display resting on top of the chart. 4

Figure 3. San Diego Bay Chart 18773 displayed in augmented reality via an extended reality head mounted display. 5

Figure 4. San Diego Bay approach seen in virtual reality via an extended reality head mounted display. 5

Figure 5. Synthetic experience infrastructure: Navigation, Seamanship, Shiphandling Trainer virtual environment (left) versus Virtual Bridge and Nautical Trainer virtual environment (right). 7

Figure 6. Crew members in Navigation, Seamanship, Shiphandling Trainer (left) and in extended reality Virtual Bridge and Nautical Trainer (right). 9

Figure 7. A watchstander avatar (head and hands) "inside" the bridge virtual environment standing at the pelorus. 10

Figure 8. Watchstander's first-person perspective controlling a pelorus "inside" the bridge virtual environment. 10

Figure 9. Watchstander's first-person perspective looking "out of the windows" using virtual environment binoculars. 11

Figure 10. Knowledge flow visualization model. Source: Nissen (2002). 21

Figure 11. Flow vectors in Nissen's knowledge flow theory model. Source: Nissen (2006). 22

Figure 12. Embodiment and structural coupling. Source: Mingers (2001). 26

Figure 13. Photograph of an actual U.S. Navy bridge (left) and the representative mockup virtual environment (right). 29

Figure 14. Fixed, physical mockup virtual environment (left) and mobile, information-based virtual environment (right). 30

Figure 15. Mingers' structural coupling to an extended reality–based synthetic experience. Adapted from Mingers (2001). 31

Figure 16. Virtual Bridge and Nautical Trainer extended reality testing session in Mariner Skills Training Center, Pacific, Classroom 215. 47

Figure 17. Virtual Bridge and Nautical Trainer extended reality pack-up kit in Mariner Skills Training Center, Pacific, Classroom 215. 47

Figure 18. Photograph of Navigation, Seamanship, Shiphandling Trainer 5B from outside and below the bridge mockup windows. 49

Figure 19. Photograph of Navigation, Seamanship, Shiphandling Trainer 5B portside video projectors taken from outside and above the mockup bridge. 49

Figure 20. Officer of the deck wearing Pupil Labs' eye tracking glasses. 53

Figure 21. Exit interview word cloud from officer of the deck responses. 167

Figure 22. Dynamic knowledge flow model. Source: Nissen (2016). 181

Figure 23. Dissertation theoretic extension of Nissen's Knowledge Flow Model to incorporate synthetic experience. 182

Figure 24. Dissertation theoretic construct informational structural coupling to the real world versus the virtual world. 185

Figure 25. Dissertation notional operational view of extended reality bridge training in a classroom or other locally connected space. 198

Figure 26. Dissertation notional operational view of extended reality bridge training dispersed to any location with a WiFi access point. 199

Figure 27. Dissertation notional operational view of extended reality bridge training environments integrated with live, virtual, constructive networks. 202

LIST OF TABLES

Table 1. Experimental design and independent and dependent variables. 35

Table 2. Threats to internal validity and controls. 38

Table 3. Threats to external validity and controls. 39

Table 4. Summary of survey response statistical analyses. 63

Table 5. National Aeronautics and Space Administration Task Load Index subscale question key. 67

Table 6. National Aeronautics and Space Administration Task Load Index survey, descriptive statistics of Navigation, Seamanship, Shiphandling Trainer control and extended reality experimental conditions for *All Participants* (N = 129). 69

Table 7. National Aeronautics and Space Administration Task Load Index survey, distributional fit analysis of Navigation, Seamanship, Shiphandling Trainer and extended reality control and experimental conditions for *All Participants* (N = 129). 70

Table 8. National Aeronautics and Space Administration Task Load Index survey, normality checks, Shapiro–Wilk and D'Agostino–Pearson of Navigation, Seamanship, Shiphandling Trainer and extended reality control and experimental conditions for *All Participants* (N = 129). 71

Table 9. National Aeronautics and Space Administration Task Load Index survey, Kendall's W (with ties) of Navigation, Seamanship, Shiphandling Trainer and extended reality control and experimental conditions for *All Participants* and *Subgroups*. 72

Table 10. National Aeronautics and Space Administration Task Load Index survey, Guttman's Lambda of Navigation, Seamanship, Shiphandling Trainer and extended reality control and experimental conditions for *All Participants* and *Subgroups*. 73

Table 11. National Aeronautics and Space Administration Task Load Index summary of comparisons of Navigation, Seamanship, Shiphandling Trainer versus extended reality for *All Participants* and *Subgroups*. 74

Table 12. National Aeronautics and Space Administration Task Load Index of order effects analysis after exposure to the extended reality condition

of Navigation, Seamanship, Shiphandling Trainer for *All Participants* (N = 129). 75

Table 13. National Aeronautics and Space Administration Task Load Index principal component analysis of Navigation, Seamanship, Shiphandling Trainer condition for *All Participants*. 76

Table 14. National Aeronautics and Space Administration Task Load Index principal component analysis of the extended reality condition for *All Participants*. 76

Table 15. National Aeronautics and Space Administration Task Load Index rotated principal component analysis of the Navigation, Seamanship, Shiphandling Trainer condition responses for *All Participants*..... 77

Table 16. National Aeronautics and Space Administration Task Load Index rotated principal component analysis of the extended reality condition responses for *All Participants*..... 77

Table 17. National Aeronautics and Space Administration Task Load Index power analysis of compared control and experimental condition findings of mental demand and performance for *All Participants* and *OOD Only*. 78

Table 18. System Usability Scale question key. 80

Table 19. System Usability Scale descriptive statistics of Navigation, Seamanship, Shiphandling Trainer and extended reality condition responses for *All Participants* (N = 130). 82

Table 20. System Usability Scale distributional fit analysis of Navigation, Seamanship, Shiphandling Trainer and extended reality condition responses for *All Participants* (N = 130). 83

Table 21. System Usability Scale Shapiro–Wilk and D'Agostino–Pearson normality checks of Navigation, Seamanship, Shiphandling Trainer and extended reality condition responses for *All Participants* (N = 130). 84

Table 22. System Usability Scale Guttman's Lamda and split-half internal consistency tests of Navigation, Seamanship, Shiphandling Trainer and extended reality condition responses for *All Participants* and *Subgroups*. 86

Table 23. System Usability Scale Hotelling's T^2, Wilcoxon signed-rank, Cohen's *d*, and Bonferroni comparison tests of Navigation,

Seamanship, Shiphandling Trainer versus extended reality condition responses for *All Participants* and *Subgroups*. 87

Table 24. System Usability Scale order effects analysis of Navigation, Seamanship, Shiphandling Trainer response after exposure to extended reality for *All Participants* (N = 130). 88

Table 25. System Usability Scale principal component analysis of Navigation, Seamanship, Shiphandling Trainer condition responses for *All Participants*. 89

Table 26. System Usability Scale principal component analysis of extended reality condition responses for *All Participants*. 90

Table 27. System Usability Scale rotated principal component analysis of Navigation, Seamanship, Shiphandling Trainer condition responses for *All Participants*. 90

Table 28. System Usability Scale rotated principal component analysis of extended reality condition responses for *All Participants*. 91

Table 29. System Usability Scale principal component factor loadings of extended reality condition responses for *OOD Only*. 92

Table 30. System Usability Scale principal component analysis of extended reality conditions to statement summary for *OOD Only*. 92

Table 31. System Usability Scale statement to factor principal component analysis summary of Navigation, Seamanship, Shiphandling Trainer and extended reality conditions for *OOD Only*. 93

Table 32. System Usability Scale power analysis of "ease and integration" and "confidence and attitude" components of Navigation, Seamanship, Shiphandling Trainer and extended reality conditions for *All Participants*. 94

Table 33. System Usability Scale power analysis "ease and integration" and "confidence and attitude" components of Navigation, Seamanship, Shiphandling Trainer and extended reality conditions for *OOD Only*. 95

Table 34. Bridge watchstander self-assessment question key. 98

Table 35. Bridge watchstander self-assessment descriptive statistics for *All Participants*. 101

Table 36. Bridge watchstander self-assessment distributional fit results by question and condition for *All Participants*. 104

Table 37. Bridge watchstander self-assessment Shapiro–Wilk normality findings of Navigation, Seamanship, Shiphandling Trainer and extended reality conditions for *All Participants*. 107

Table 38. Bridge watchstander self-assessment survey Kendall's W (with ties) analysis of Navigation, Seamanship, Shiphandling Trainer and extended reality responses for *All Participants* and *Subgroups*. 109

Table 39. Bridge watchstander self-assessment survey, Guttman's Lamda and split-half analysis of Navigation, Seamanship, Shiphandling Trainer and extended reality responses for *All Participants* and *Subgroups*..... 109

Table 40. Bridge watchstander self-assessment survey, sample and questions retained for Hotelling's T^2 analysis by watchstander for *All Participants.* 111

Table 41. Bridge watchstander self-assessment survey for Hotelling's T^2, Wilcoxon signed-rank, Cohen's *d*, and Bonferroni comparison of Navigation, Seamanship, Shiphandling Trainer versus extended reality condition responses for *All Participants* and *Subgroups*..... 112

Table 42. Bridge watchstander self-assessment survey, Mann–Whitney and Cohen's *d* order effects analysis of Navigation, Seamanship, Shiphandling Trainer responses after exposure to extended reality for *All Participants* (N = 36)..... 114

Table 43. Bridge watchstander self-assessment survey, principal component analysis of Navigation, Seamanship, Shiphandling Trainer condition responses for *All Participants.* 116

Table 44. Bridge watchstander self-assessment survey, principal component factor loading of Navigation, Seamanship, Shiphandling Trainer condition responses for *All Participants*..... 117

Table 45. Bridge watchstander self-assessment survey, principal component analysis rotated factors and thematic descriptions of Navigation, Seamanship, Shiphandling Trainer condition responses for *All Participants*..... 118

Table 46. Bridge watchstander self-assessment survey principal component analysis of extended reality condition responses for *All Participants*.... 119

Table 47. Bridge watchstander self-assessment survey, factor loading principal component analysis of extended reality condition responses for *All Participants*..... 120

Table 48. Bridge watchstander self-assessment survey, principal component analysis of rotated factors and thematic descriptions of extended reality condition responses for *All Participants*..... 121

Table 49. Bridge watchstander self-assessment survey, principal components comparison of Navigation, Seamanship, Shiphandling Trainer versus extended reality responses for *All Participants*. 122

Table 50. Bridge watchstander self-assessment survey, power analysis across principal components Navigation, Seamanship, Shiphandling Trainer versus extended reality condition responses for *All Participants*..... 124

Table 51. Officer of the deck self-assessment survey question key. 127

Table 52. Officer of the deck self-assessment survey, descriptive statistics of Navigation, Seamanship, Shiphandling Trainer and extended reality condition responses for *OOD Only*..... 130

Table 53. Officer of the deck self-assessment survey, distributional fit results by question and condition, revised Akaike Information Criterion rule of Navigation, Seamanship, Shiphandling Trainer and extended reality responses for *OOD Only*..... 132

Table 54. Officer of the deck self-assessment survey Shapiro–Wilk and D'Agostino–Pearson normality checks of Navigation, Seamanship, Shiphandling Trainer and extended reality responses for *OOD Only*. ... 134

Table 55. Officer of the deck self-assessment survey, Kendall's W (with ties) analysis, optimized $N \geq 20$, of Navigation, Seamanship, Shiphandling Trainer and extended reality conditions for *OOD Only*. .. 136

Table 56. Officer of the deck self-assessment survey, Guttman's Lamda and split-half analysis of Navigation, Seamanship, Shiphandling Trainer and extended reality responses for *OOD Only*. 137

Table 57. Officer of the deck self-assessment survey, sample and questions retained for Hotelling's T^2 analysis for *OOD Only*. 139

Table 58. Officer of the deck self-assessment survey, Hotelling's T^2, Wilcoxon signed-rank, Cohen's *d*, and Bonferroni comparison tests of Navigation, Seamanship, Shiphandling Trainer versus extended reality responses for *OOD Only*..... 140

Table 59. Officer of the deck self-assessment survey, Mann–Whitney order effects tests of Navigation, Seamanship, Shiphandling Trainer responses compared to responses after exposure to extended reality for *OOD Only*. 141

Table 60. Officer of the deck self-assessment survey, principal component analysis of Navigation, Seamanship, Shiphandling Trainer condition responses for *OOD Only*. 143

Table 61. Officer of the deck self-assessment survey, principal component factor loadings analysis of Navigation, Seamanship, Shiphandling Trainer condition responses for *OOD Only*. 144

Table 62. Officer of the deck self-assessment survey, principal component analysis of rotated factors and thematic descriptions of Navigation, Seamanship, Shiphandling Trainer condition responses for *OOD Only*. 145

Table 63. Officer of the deck self-assessment survey, principal component analysis for extended reality condition responses for *OOD Only*. 146

Table 64. Officer of the deck self-assessment survey, factor loadings principial component analysis for extended reality condition responses for *OOD Only*. 146

Table 65. Officer of the deck self-assessment survey, principal component analysis of rotated factors and thematic descriptions of extended reality condition responses for *OOD Only*. 147

Table 66. Officer of the deck self-assessment survey, Navigation, Seamanship, Shiphandling Trainer and extended reality principal component analysis comparison for *OOD Only*. 148

Table 67. Officer of the deck self-assessment survey statistical power analysis for *OOD Only*. 150

Table 68. Instructor assessment question key. 152

Table 69. Instructor assessment, descriptive statistics of the full dataset with normalized items for *Instructors Only*. 155

Table 70. Instructor Assessment, summary of Wilcoxon signed-rank Test and Bonferroni results for *Instructors Only*. 158

Table 71. Bridge watchstander sentiment summary for *All Subgroups*. 165

Table 72. Role-based themes with representative quotes for *All Subgroups*. 166

Table 73. Text analysis findings for extended reality, “strengths” versus “areas for improvement” for *All Participants*. 167

Table 74. Summary of contributions to method. 196

LIST OF ACRONYMS AND ABBREVIATIONS

2D	two dimensional
3D	three dimensional
AIC	Akaike Information Criterion
AOI	area of interest
AR	augmented reality
AV	augmented virtuality
BCI	brain computer interface
BWS	bridge watchstander
CCS	central control station
CIC	combatant information center
CNSF	Commander Naval Surface Forces
CO	commanding officer
CONN	conning officer
CP	centerline pelorus
DDG	guided missile destroyer
EOOW	engineering officer of the watch
GPT	generative pre-trained transformer
HELM	helmsman
HMD	head mounted device
HR fMRI	high resolution functional magnetic resonance imaging
HVAC	heating ventilation and air conditioning
ICR	internal consistency reliability
IRR	inter-rater reliability
IT	information technology
JOOD	junior officer of the deck
KFT	knowledge flow theory
KM	knowledge management
LEE HELM	lee helmsman
LLM	large language model
LOB	line of bearing

LOE	letter of endorsement
LoS	letter of support
LVC	live virtual constructive
MR	mixed reality
MSTCPAC	Mariner Skills Training Center, Pacific
NASA TLX	National Aeronautics and Space Administration Task Load Index
NAV	navigator
NIWCPAC	Navy Information Warfare Center Pacific
NSST	Navigation, Seamanship, Shiphandling Trainer
OJT	on the job training
OOD	officer of the deck
PCA	principal component analysis
PoV	point of view
PUK	pack-up kit
RADAR	radio detection and ranging
ROI	return on investment
RV	reality–virtuality
SA	situational awareness
SCC PC	Scenario Creation and Control Personal Computer
SECI	socialization externalization combination and integration
SUS	System Usability Scale
SWO	surface warfare officer
TA	text analysis
USN	U.S. Navy
VE	virtual environment
VIBRaNT	Virtual Bridge and Nautical Trainer
VMS	voyage management system
VR	virtual reality
WSRT	Wilcoxon signed-rank test
XO	executive officer
XR	extended reality

ACKNOWLEDGEMENTS

It has been said “no man is an island,” and that is certainly the case here. This dissertation would not have been possible without the generosity, encouragement, and expertise of many people—far more than I could ever name. At the risk of overlooking others whose contributions were equally meaningful, I offer the following acknowledgements with deep appreciation.

I thank the Naval Postgraduate School (NPS) for providing the academic home that made this work possible. I am especially grateful to Dr. Alex Bordetsky, Chair of the Information Sciences Department, for his guidance on theoretical considerations and steady support throughout my doctoral studies. My deepest thanks go to my PhD supervisors, Dr. Don Brutzman (now retired), for his ever steadfast and supportive mentorship even in retirement and Dr. Mollie McGuire, who offered to step in as my supervisor after Don retired. Both supported this dissertation and my development as a scholar with superior mentorship, patience, and encouragement. I am also sincerely grateful to the entire dissertation committee—Dr. Johnathan Mun, Dr. Carl Oros, Dr. Mark Nissen (retired), and Dr. Jamie Lukos—for their thoughtful guidance and enduring support. Each member offered expertise from their unique lens of scholarship to build this up into a greater whole than would have otherwise been possible without their contributions.

Several faculty members shaped my progress from the very beginning. Dr. Erik Jansen (now retired) provided steady grounding in Philosophy of Science during the darker days of the COVID-19 pandemic, a time when persistence was not always easy. Dr. Thomas Housel offered many hours of insight and challenge during our Information Sciences PhD seminar, helping to refine the questions that guided this research. Also, to Dr. Mark Nissen (now retired) for his guidance of my 5th area studies and in particular for his expertise in knowledge flow, I extend a hearty thanks! In the final stages of writing, the NPS Graduate Writing Center, especially Michael Thomas, played a crucial role through his patient review and tireless commitment to concision and clarity. I am also thankful for the support of the NPS Surface Warfare Chair, Captain Jeremy Gray, for his support in

liaison with the Commander Surface Warfare Forces Pacific, in attaining their support of VIBRaNT XR research efforts at Mariner Skills Training Center Pacific (MSTCPAC).

I am grateful to my NIWC Pacific PhD NPS Information Science Cohort 3 colleagues for their camaraderie over the many years of coursework, writing, and paper chasing. I am especially thankful to Dr. Chris Cirullo, whose steady encouragement helped me maintain momentum when it mattered most.

This work benefitted immeasurably from the support of Naval Information Warfare Center Pacific (NIWC Pacific). I am deeply grateful to Ms. Amelia Kracinovich, VIBRaNT Principal Investigator (PI) and Integrated Product Team (IPT) Lead, whose leadership and human-systems integration expertise contributed to survey instrument selection (NASA-TLX, SUS, SSQ, and Self-Assessment Rubrics), IRB development, and overall VIBRaNT project design. I also thank Jacob Anderson for his technical wizardry in setting up VIBRaNT equipment and data recording hardware, project interns Noah Walsh and Luke Lastomirski, and early champions Robert Tanner, Dr. Paul Shigley and Dr. Clare Morton, whose encouragement helped launch this research. My heartfelt appreciation goes to my NIWC leadership—past Division Heads Greg Settelmayer and Jodi McGee, current Division Head Alex Vasel, and especially Dr. Laura Baker, Division Operations Officer, whose support has been steady and unwavering from the very beginning. I am also grateful to Dr. Alex Phipps, my NIWC Pacific PhD Sponsor, for his advocacy and support throughout this journey.

I am further indebted to Kara Sorensen, NIWC Pacific IRB Chair, and Kathleen Brocker for their essential guidance in preparing the VIBRaNT IRB Protocol (NIWCP.2024.0003-IR-EP7-A). I also thank the Office of Naval Research Technical Solutions Team, including Dr. Jessica Jones, Mr. Jason Payne, and Mr. John Coffey, for having the vision and willingness to take the risk in funding NIWC Pacific's VIBRaNT research project. Their response to fleet needs sponsorship of research into XR made this study possible. This research would also not have been possible without the support of the Commander Naval Surface Forces Pacific Science Advisor, Mr. Keith Hines, in gaining permission to conduct the research at MSTCPAC within the allowed training windows afforded ships stationed in the Pacific Fleet Area of Responsibility. I also extent a warm

thanks to the U.S. Navy Officers and Sailors who volunteered their time professionalism to participate in the study with patient, open-minded professionalism so indicative of the best traditions of naval service. I am equally grateful for the support of Surface Warfare Schools Command, including Ms. Elleanor Richards, and Mr. Eric Stailey, Executive Director of Mariner Skills Training Center Pacific, whose coordination enabled access to essential facilities and personnel. I also acknowledge Andy Parkin and the XR development team at Kilo Solutions, UK, whose technical expertise was indispensable to Project VIBRaNT and enabled the findings contained this study.

Finally—and most importantly—I thank my family. My wife, Dr. Michele Brewster, a published author herself, has been my greatest source of strength, grounding, and perspective. Her patience and encouragement carried me through every late night, and long weekend spent at the keyboard. My daughters, Audrey and Ellie, showed remarkable understanding and resilience throughout this journey; their humor, love, and light sustained me more than they know. This dissertation is as much theirs as it is mine.

THIS PAGE INTENTIONALLY LEFT BLANK

DISCLOSURES

To support my writing process for this research paper, I used OpenAI's ChatGPT 4 and GPT5. I received permission from Professors Dr. Alex Bordetsky and Dr. Mollie McGuire, who required a specific description of how I would use AI and emphasized the importance of maintaining academic integrity.

I used the tool at several stages of the writing process to help me generate ideas, build tables, cross check analyses, and generate summary text from statistical summary tables and statistical findings. GPT5 was also used to conduct guided sentiment analyses of captured exit interview data and generate pattern summaries and word clouds. First, I identified a broad topic area of interest and asked it to confirm a list of possible research gaps for my topic. I then explored some of these gaps conversationally with the tool. Second, I asked it to make tables from statistical results generated by the tool after entry into GPT5. Tables generated were then confirmed via comparison to raw data results. Third, I used it to generate narrative summaries of analytic findings from the statistics analysis software.

To avoid the risk that this work might rely inappropriately on an AI tool, I also reviewed multiple journals, library references, on-line resources, e.g., Google Scholar, and other materials through consultation with my professor on my research and theoretically related ideas, gaps, and findings. All of the work in this dissertation was reviewed with my dissertation committee, including the use of GPT5 as described above, before committing to its suggestions. I consulted with my GWC writing coach throughout the drafting of this dissertation, to include incorporation of GPT5 generated tables and other ancillary content.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

What is information? This simple question, deeply intertwined with the questions concerning the nature of knowledge, has challenged philosophers since the time of Aristotle. In its classical conception, information was posited as an agent to impress or inform as

> a piece of wax takes on the impress (information) of a signet ring without the iron or gold; we say that what produces the impression is a signet of bronze or gold, but its particular metallic constitution makes no difference: in a similar way the sense (human perception) is affected by what is colored or flavored or sounding, but it is indifferent what in each case what the substance is; what alone matters is the quality it has, i.e., in what ratio its constituents are combined. (Aristotle, 1989)

This analogy, of information as external artifact making impressions on "the sense," highlights the importance of the coupling of *dead* information to the *living* human mind. While it is not uncommon to find philosophers at odds over that which is knowledge in relation to that which is information, what remains consistent is the coupling of the two concepts one to another, regardless of epistemological distinctions or various philosophies of mind framing the argument (Adriaans, 2024).

This relationship of information to knowledge, particularly tacit knowledge, becomes ever more complex and as capabilities of information-based technologies expand and complicate what is meant by "human access to information." With the rise of digital technology, almost anything can be informed in media as strings of 0's and 1's. These strings (Collins, 2019) can then be machine transformed into nearly any kind of structure or representation for human perception. The fluid nature of this transforming of digital strings expands the conception of information access to ever-deeper levels of structural coupling between individuals and their information. This digitally enabled, mutually causal, human–machine, knowledge-information coupling of individual minds (embodied, embedded, extended, or otherwise described) to algorithmically sculpted information environments warrant greater attention and further study.

The specific human–machine–knowledge-information coupling explored in this research is centered on the rapidly emerging information access technologies commonly referred to as Extended Reality (XR). An umbrella term, XR today broadly refers to a variety of immersive visual, aural, and haptic information technologies (IT) available in the commercial marketplace. In this research, XR refers to any mix of panmodal technologies employed to provide a user, group, or organization with a desired information access, to include an embodied sense of agency (Clark, 2023) to facilitate learning objectives previously limited to physical mockups (Bailenson, 2019).

The information access provided by XR is posited as a critical new enabler of Knowledge Flow. Specifically, this study sought to characterize and evaluate the ability of XR to enable tacit knowledge flow by presenting a knowledge bearing synthetic experience to overcome the classic problem of experts "knowing more than they can tell" by eliminating altogether the need to "tell" or otherwise make explicit their expertise semantically, but rather by immersing the novice within expertly crafted synthetic experience. By comparing XR-based training to physical mockup training for U.S. Navy (USN) bridge teams, the study sought to empirically test if an information only virtual environment (VE) could measure up to a physical analog of a similar experience carried via a mockup simulator. The overarching hypothesis is that information access provided by XR can supplement or, in some circumstances, replace physical mockup training.

A. BACKGROUND

Human access to information has long been a concern of Information Science (Borko, 1968). Over the centuries, means of processing information for human access has evolved from marking wedge-shaped cuneiform styli impressions on clay tablets, to writing in ink, printing with blocks, telegraphy over wires, emitting signals in the electromagnetic spectrum, computing binary digits for display on computer screens, to multi-sensory immersion technologies of XR. Today more than ever, commercially available computer and network technologies, referred to hereafter as Information Technology (IT), are dramatically transforming how individuals, organizations, and societies become informed.

XR, as noted, is an evolving commercial term roughly encompassing the collection of augmented and virtual display and interaction technologies originally described in 1995 by Milgram et al. in the reality–virtuality continuum (Milgram et al., 1995) (Figure 1).

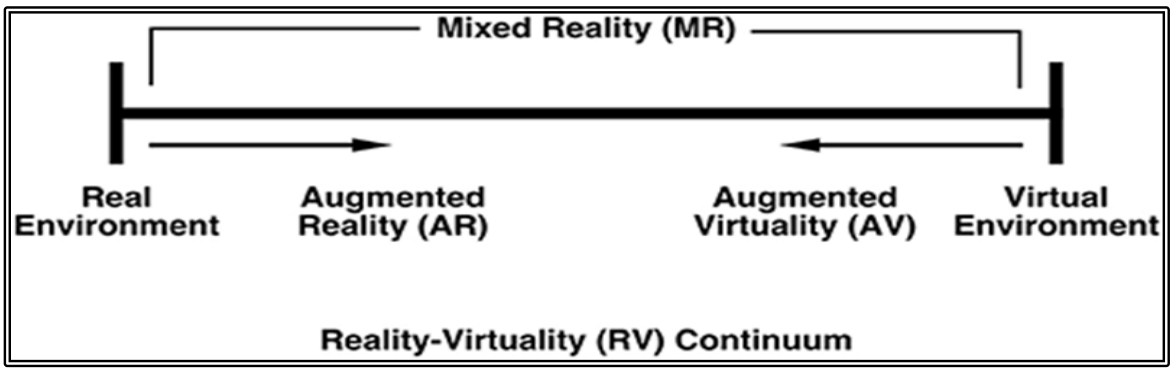


Figure 1. Reality–virtuality continuum. Source: Milgram et al. (1995).

The reality–virtuality (RV) continuum identifies five types of information access:

1. **Reality (R):** Information in the real world as experienced every day.
2. **Augmented Reality (AR):** Information in the real world augmented with visual, aural, and haptic information overlayed into the field of view of the observer.
3. **Augmented Virtuality (AV):** Information from the real world, e.g., the observer's hands, a keyboard, mouse, etc., overlayed in the VE.
4. **Virtual Reality (VR):** Information from the real world occluded and completely replaced with visual, aural, and haptic information composing a VE.
5. **Mixed Reality (MR):** Information access provided by R, AR, VR, and AV cross-modally mixed across the RV continuum.
6. **Extended Reality (XR):** An umbrella term originated after the 1995 RVC concept usually associated with panmodal R, AR, AV, VR, or MR technologies packaged in a single form factor.

XR access to information is sometimes referred to as *spatial computing* (Greenwold, 2003) as it provides information in both binocular and binaural modalities. To better understand the information affordances of XR, Figures 2 through 4 show different information accesses of San Diego Bay, from a nautical chart to a head mounted display (HMD) (Meta Quest 3) enabled VR presentation.



Figure 2. San Diego Bay Nautical Chart 18773 and extended reality head mounted display resting on top of the chart.

The nominal two-dimensional (2D) nautical chart is a static representation with text, figures, lines and colors on paper. The XR HMD, on the other hand, can dynamically interact in multiple modes that provide different means of access to "see" San Diego Bay. It can show solid physical object(s), textual, or graphical references in relation to the real world in AR. In Figure 3, the same chart shown of San Diego Bay, 18773 (Figure 2), is presented electronically via a 2D AR presentation.

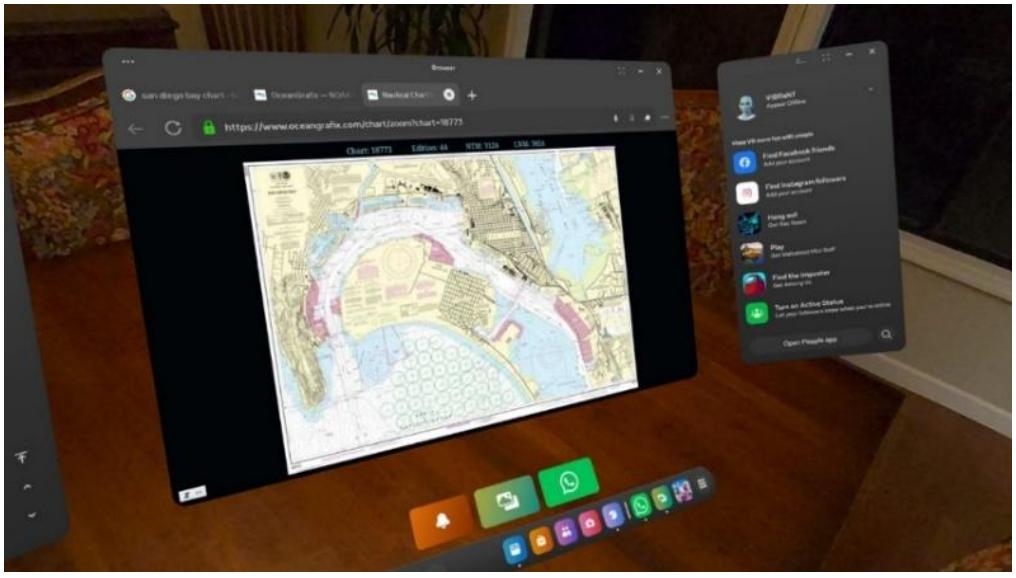


Figure 3. San Diego Bay Chart 18773 displayed in augmented reality via an extended reality head mounted display.

The same XR HMD can also place a user's point of view inside a virtual world. In Figure 4, San Diego Bay, chart 18773, is perceived in the same way as if one was "aboard ship" approaching Point Loma and the San Diego Bay channel entrance.

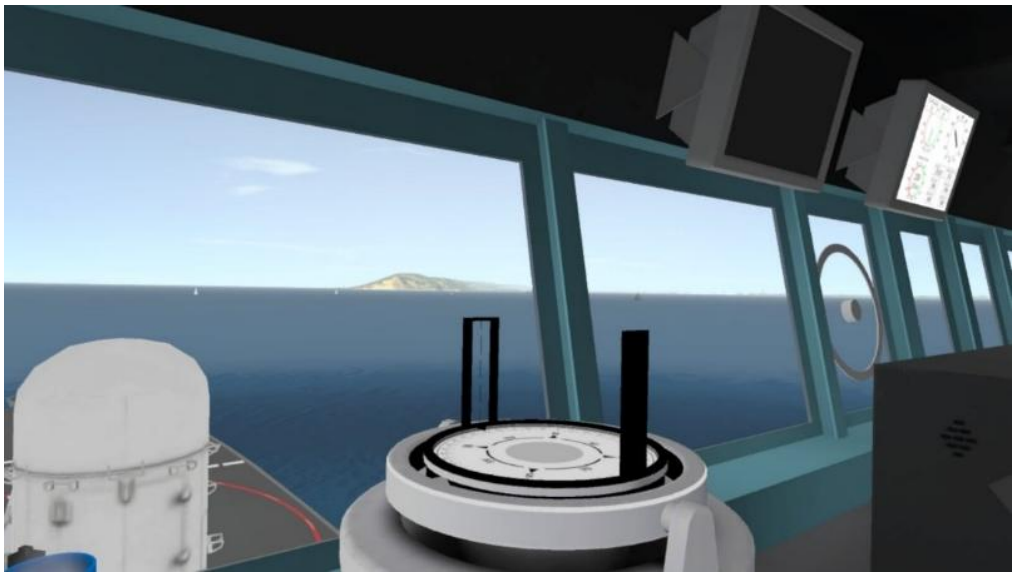


Figure 4. San Diego Bay approach seen in virtual reality via an extended reality head mounted display.

Using an XR HMD, users can switch back and forth between modalities, from seeing a paper chart in a "real location" on a table top via pass-through mode, to observing an AR digital depiction with information overlayed in the same real location, via AR seen in pass-through mode) or seeing San Diego as one might if they were "there in first-person" via VR access.

The study's conceptual frame of knowledge flow refers to the knowledge management (KM) practice of managing organizational learning wherein "knowledge, the information and skills an individual learns through experience or education, can be transferred from person to person; in other words, it can flow" (Shigley, 2021, p. 1). Use of IT to facilitate knowledge flows has been studied for decades (Nissen, 2014). What this research specifically examines is potential of XR use by the USN to flow knowledge needed to train bridge crew recognition and response in a variety of maritime situations. Facilitating this kind of knowledge flow has historically required either on-the-job training or the design, construction, and maintenance of physical mockups (Bailenson, 2019). This research examines the efficacy of supplementing, or in some cases replacing, physical location based mockups with XR to help flow the tacit knowledge required for complex USN operations such as shiphandling.

Although the information access of VR has been studied widely in academia, recent advancements and falling costs of commercially available XR make it a viable option for organizations to enhance knowledge flows and such potential capability warrants further research. Previous problems of simulator sickness have largely been mitigated with frame rates above 72 frames per second, dedicated chip sets, e.g., Qualcomm XR2, Gen2, and video fidelity above 1080 P resolution and increased numbers of pixels per inch. Commercial providers continue to introduce new varieties of XR capability at ever lower costs. These include flying rigs, remote-controlled chairs, and walking surfaces to name but a few, nevertheless the impacts of these applications, particularly from the perspective of organizational management, remain understudied. The complexities of XR-enabled information access are vast; XR not only provides new access modalities, but it can also be coupled to any array of conventional IT, e.g., computers, smart phones, wireless local area networks, Internet, etc., to enable new mixes of different access affordances.

In addition, XR access has the potential to be coupled with artificial intelligence (AI), large (or small) language models, or any other emergent machine-aided process for accessing information across the RV. Yet another reason to conduct this research is cost. As noted, commercially available XR devices continue to fall in cost, while equivalent physical mockup training systems remain expensive to operate. An appreciation of this difference can be seen in Figure 5.



Figure 5. Synthetic experience infrastructure: Navigation, Seamanship, Shiphandling Trainer virtual environment (left) versus Virtual Bridge and Nautical Trainer virtual environment (right).

In the left set of pictures (Figure 5) is the USN's $1,770,000, three-story enclosed Navigation, Seamanship, Shiphandling, Trainer (NSST) mockup. On the right is the XR-based prototype, the Virtual Bridge and Nautical Trainer (VIBRaNT) system of 10 wireless HMDs contained in a pack-up-kit which, at a cost of approximately $20,000 dollars, is a mere fraction of the cost of the physical mockup. Due its low-cost, an XR capability, such as VIBRaNT, can dramatically scale to deliver mission critical training and rehearsal capability anywhere it's needed serving not only to protect valuable ships and harbor assets worth billions of dollars, but also lives of those who serve aboard these ships and countless others who interact with them.

B. RESEARCH FOCUS

XR applications in training have been a topic of recent research in the military (Caso, 2024; Huff, 2023; McCoy-Fisher et al., 2019; Mouloua et al., 2023; Ross & Gilbey, 2023). In the specific case of this research, the employment of XR information access is framed in its capability to provide a synthetic experience wherein the real world is replaced with a VE to enable information flow of spatial-temporal cues in response to real-time human interaction and feedback. This dissertation hypothesizes that an XR facilitated synthetic experience facilitates organizational capability to help flow tacit knowledge needed for complex team-based operations performed on a USN bridge.

This study was performed in the spring of 2025 at the USN Mariner Skills Training Center Pacific (MSTCPAC) in San Diego, CA to compare the facilitation of knowledge flow between the well-established team training of the USN's NSST and an XR VE called VIBRaNT. The comparison was conducted through a post-positive, quantitative, quasi-experiment with human subjects, with the study employing a counter-balanced, within subjects design.

The study was bound to a single ship class, USN's Arleigh Burke Guided Missile Destroyers. This warship class, the most prevalent in the USN, has a bridge team composed of officers and enlisted personnel normally assigned to six watch positions: Officer of the deck (OOD), junior officer of the deck (JOOD), conning officer (CONN), navigator (NAV), helmsman (HELM), and lee helmsman (Lee Helm). This team structure is prevalent for all USN ships and closely matches teams found aboard foreign warships as well. In addition, skills needed for bridge team operations are not for just junior officers or deck seaman but required across an entire career for most mariners. Focusing on a single ship class for the study enabled higher verisimilitude of the XR VE while also controlling for variation in the distribution of control console and instrument laydowns differing from the configurations of other ship classes.

Figure 6 shows a bridge crew training inside the USN physical mockup VE (left picture) compared to the same crew training inside the XR VE (right picture). Both synthetic experiences enable social interaction, marine environmental effects, audio cues,

bridge displays, e.g., radio detection and ranging (RADAR), Voyage Management System (VMS), HELM and Lee HELM controls, and a view of the external world "outside the bridge windows" as it typically appears aboard a ship.



Figure 6. Crew members in Navigation, Seamanship, Shiphandling Trainer (left) and in extended reality Virtual Bridge and Nautical Trainer (right).

The XR participants (Figure 6, right picture) inhabit a virtual world that cannot be captured "from the outside" as was done in the photo taken of the NSST (Figure 6, left picture). To appreciate the XR VE, one must see it "from the inside" as viewed via the XR HMD (Figure 7). Perceiving the VE from this view, one sees participants have agency to choose where to be (Figure 7), bridge equipment to interact with (Figure 8), and can even "use" binoculars to magnify distant objects (Figure 9) as they would in the real world.

In Figure 7, a deck officer uses the Centerline Pelorus (CP) to take a Line Of Bearing (LOB) measurement for a precise direction to another ship approaching in the distance. The officer is represented by a blue avatar whose head movement (gaze direction) and hand positions (grasping or pointing) visually mirror the officer's actual interaction within the XR VE. Performing this task has a high degree of verisimilitude, to include proprioception, with expected team member actions aboard a ship in real world operations.



Figure 7. A watchstander avatar (head and hands) "inside" the bridge virtual environment standing at the pelorus.

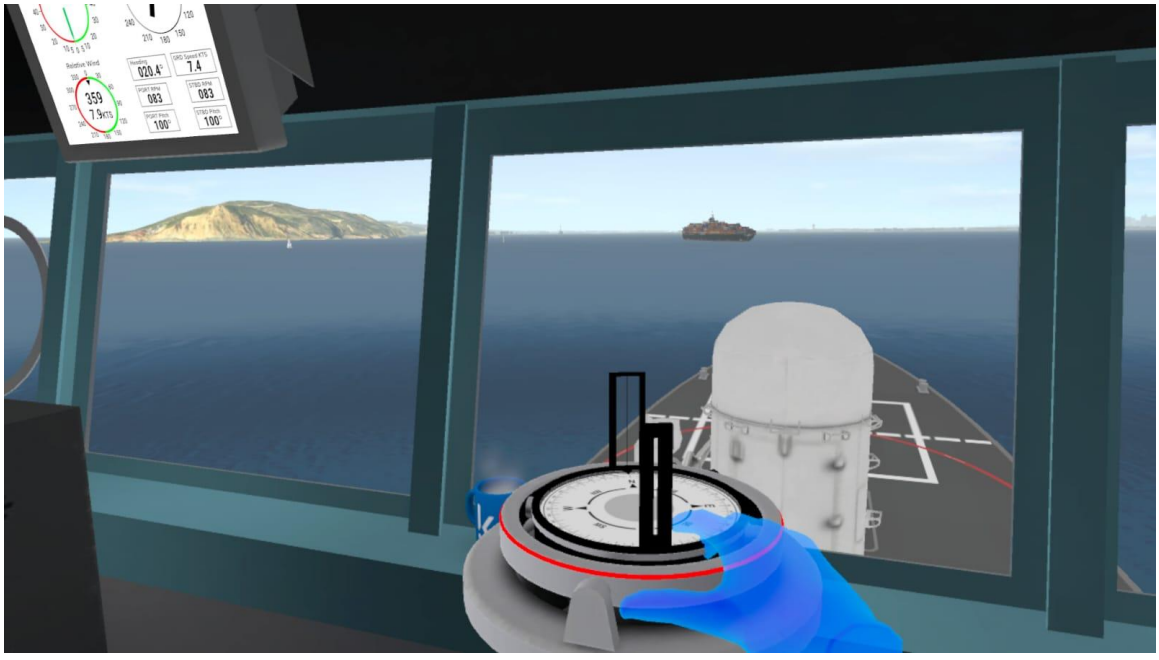


Figure 8. Watchstander's first-person perspective controlling a pelorus "inside" the bridge virtual environment.

In Figure 8, the first-person perspective of the deck officer can be seen. Here the officer turns the CP in XR just like the real world to take an LOB to a ship on the horizon approaching off the port bow. In Figure 9, the same officer uses virtual binoculars to pick out the vessel approaching out of the haze and thus ensure the LOB is correct to the ship of interest. In these figures one can see the degree of spatial-temporal fidelity is presented with sufficient verisimilitude needed for the individual performing the task as well as for the other team members participating within the VE.



Figure 9. Watchstander's first-person perspective looking "out of the windows" using virtual environment binoculars.

In mockup bridge training (Figure 6, right picture), participants have the advantage of physical touch, the proprioception of walking and colliding with other bridge watchstanders, and the ability to discern facial expressions. Both capabilities, mockup NSST and XR VIBRaNT, address the same problem: novice bridge crews lacking experience need knowledge to manage complex, life-endangering situations they are likely to face in the real world while operating at sea.

C. PROBLEM STATEMENT AND PURPOSE

Loss of life and property can result when knowledge-based human performance fails, particularly in military, contested, or other high-reliability operations where knowledge flows lack sufficient facilitation. Examples of poor knowledge flow facilitation include traditional computer-based training (knowledge flow is insufficient), on-the-job training (knowledge flow can be unavailable or too dangerous), and mockup trainers (knowledge flow is limited in availability or requires personnel to travel to specific and limited locations or is deemed too expensive to build and maintain). This is a problem because personnel, equipment, and mission success are placed at risk when dangerous, contested, or high-reliability operations must be performed without the benefit of sufficient prior knowledge. These shortfalls not only occur for individuals, but also occur for teams, particularly when teams are forming and learning the skills of their fellow teammates.

Additionally problematic for USN operations is the fact that prior knowledge must be held not just by a single individual, e.g., the OOD, but by the entire group comprising the bridge team and, as argued in *Cognition in the Wild* (Hutchins, 1995), the watch team's cognitive processes are distributed across the immediate environment of the ship's bridge with all of its various sensor displays and communication systems.

The purpose of this study is to evaluate and characterize *XR facilitated knowledge flows* required for shiphandling, specifically for handling of USN Guided Missile Destroyers. Results of this study are important because characterizing XR's ability to build expertise in USN shiphandling will add to general understanding of XR's capacity to enhance knowledge flow and improve human performance. To this end, the research seeks to evaluate and characterize XR knowledge flow facilitation in comparison to the USN's NSST physical mockup trainer.

D. RESEARCH QUESTION AND SCOPE

This research intends to characterize the extent to which an XR-generated VE can be employed to train USN bridge watch teams in the navigation, seamanship, shiphandling, and bridge resource management skills to safely conduct maritime operations. The primary question being addressed is: For the bounded study of USN shiphandling, to what extent

can knowledge flow be facilitated by an information-based XR VE compared to a physical mockup VE?

The scope of this research is a case study of 28 USN bridge teams sampled from USN Arleigh Burke Class Guided Missile Destroyers stationed at 32nd Street Naval Station in San Diego, CA. Each of the sampled bridge teams consisted of a mix of three to seven bridge crew members (officer and enlisted) with a total of 158 participants performing in the comparison. The composition of each team was held constant in both the experimental (information-based VE) and the control conditions (physical mockup or NSST VE). Eye tracking data of one member of the bridge team, the OOD, was captured in both experimental and control conditions to compare variation in blink rate and gaze vector change rate.

E. LIMITATIONS AND SAFETY CONSIDERATIONS

The research described in this dissertation analyzes Navy Information Warfare Center Pacific's VIBRaNT data collected under Institutional Review Board (IRB) Protocol NIWCP.2024.0003-IR-EP7-A dated November 19, 2024. Per the protocol, all IRB safety measures were followed and none of the 158 study participants reported any form of simulator sickness or other ill effects from the use of the VIBRaNT XR bridge training system. While no safety incidents were noted during data collection, this research was in no way a study of necessary considerations for fleet deployment of XR systems. Testing and data collection were conducted in a large land-based classroom free of obstructions typically found in industrial settings. Additionally, this study did not attempt to replicate various sea states or evaluate the use of XR aboard ships while underway.

This research utilized the Meta Quest 3 with an elevated native frame rate of 72 frames per second, which appears to mitigate simulator sickness problems noted in earlier research studies (Moss & Muth, 2011; Wang et al., 2023; Zhang, 2020). Additionally, each participant completed a simulator sickness questionnaire, and none noted ill effects to a degree requiring removal from the study. It is important to note that the research employed a low-cost, quickly developed prototype XR system on a limited budget, that XR hardware continues to evolve, and commercial safety considerations are not yet fully defined,

particularly in military settings. Individuals have reported sensitivity to the use of such devices, to the point where they could not be used, but overall understanding of this tendency in wider populations has not been identified. References regarding use of such devices aboard ships at sea have not been found, another critical consideration given Navy personnel have additional longitudinal factors of ships motion affecting their vestibular system.

It is recommended that any future effort to deploy VIBRaNT or other XR systems aboard naval vessels be preceded by sufficient study to determine proper safety constraints and set appropriate guidelines for XR shipboard use. Such precautions are necessary even in situations where the ship is tied up pier-side, since pitch, roll, and heave occur from wave action while in port. Further personnel hazards also exist in the unique environments found aboard naval vessels, to include open hatches, electrical switches, high-temperature equipment, RADHAZ and other danger areas, etc. This research comparison was strictly limited to a systems comparison hosted in a highly controlled land-based training environment and MUST NOT be taken as sufficient grounds to employ XR aboard ship.

F. ORGANIZATION OF THE DISSERTATION

Chapter I provides the background context covering XR, VEs, synthetic experience, knowledge flow, research motivation, the specific instance being studied, the research problem, purpose, scope, and the dissertation's structure.

Chapter II is composed of the philosophical foundations and conceptual framework supporting the concept of knowledge flow as it relates to tacit knowledge, embodiment, and the application of XR for generating synthetic experience to flow spatial-temporal tacit knowledge. The current gap in understanding XR facilitation of knowledge flow is reviewed, along with the study's proposed contribution to knowledge.

Chapter III contains variable descriptions, research design, test procedures, control and experimental conditions, hypotheses statements, and test setting. Also included are the VE design and scenario selection, IRB Protocol, logistical consideration in sample selection, survey questionnaires and data collection, preparation, and statistical methods employed to obtain the results.

Chapter IV present the experimentation results and analyses. Findings include statistical analyses of data collected from the following sources: a) de-identified demographic analysis, b) Simulator Sickness Questionnaire, c) NASA Task Load Index Workload Analysis, d) System Usability Survey, e) bridge watchstander self-assessment survey, f) OOD instructor assessment questionnaire, g) bridge watchstander instructor assessment questionnaire, and k) exit interview questionnaire. Data collected is statistically compared between the control and experimental conditions to evaluate the studies alternate hypothesis, H_1 to H_4.

Chapter V explores the central findings and overarching meaning of the results in terms of data findings, research expectations, and the fit to the theoretical constructs of knowledge flow and embodiment. Also included are the challenges encountered in performing the experiment, such as logistics, surprise findings, and researcher impressions.

Chapter VI reiterates the importance of findings from H_1 to H_4, addresses the contribution to of XR facilitation to knowledge flow theory, and addresses how best to employ synthetic experience for organizational advantage. Inferences are made concerning the knowledge reach and life cycle impacts of XR and recommendations are provided for future research.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

This chapter presents the foundation for the value proposition posited in the dissertation's title, *Enhancing Knowledge Flow with XR.* Specifically, it articulates what is meant by "enhancing knowledge flow" and why XR is hypothesized to "enhance" it. The argument, taken from the literature, is built in a series of steps: First, the conception of knowledge flow is built up from general philosophical distinctions of what knowledge is; second, philosophical distinctions are then mapped to delineations held within disciplines of organizational management; and third, with these mappings in mind, the application of IT to support flows of knowledge is described. The argument then turns to problems in flowing tacit knowledge, describes what is tacit in knowledge from the perspective of philosophy of mind and cognitive science, reviews application of physical mockup VEs to flow tacit knowledge, and argues for researching XR VEs to supplement or replace mockup VE's for flowing tacit knowledge needed in shiphandling.

A. KNOWLEDGE FLOW

To understand what is meant by knowledge flow and why XR might be able to enhance knowledge flow, a brief overview of knowledge as defined in epistemology is provided, followed by delineations of knowledge found in organization management literature including: KM, knowledge flows, knowledge flow models, and IT support of organizational knowledge flow.

1. Philosophical Distinctions in Knowledge

Defining knowledge, its true nature and meaning, has been a philosophical concern for centuries. While the philosophical debate continues, epistemologists have traditionally categorized knowledge in three kinds stemming from ancient Greek philosophical terms: *empeiria, episteme*, and *techne* (Coope, 2020; Johansen, 2017; Lorenz & Morison, 2019). *Empeiria,* or acquaintance knowledge, is based in direct experience or familiarity as being "knowledge of"; *episteme,* or propositional knowledge, is based in linguistic or symbolic abstraction as "knowledge that"; and *techne,* or procedural knowledge, is based in skill, true art, or technique as "knowledge how" (Pavese, 2022). Despite this generally held

convention, philosophical debate concerning knowledge continues today in two camps of epistemology: Intellectualism and Anti-Intellectualism. Intellectualists argue "knowledge how" *techne* is a form of "knowledge that" *episteme* (Stanley, 2001, 2017). In their view, Intellectualists posit that both kinds of knowledge stem from the same source: the human capacity to form symbolic abstraction. Anti-Intellectualists refute this view and argue for the clear distinction of "knowledge how" *techne* being separate and distinct from "knowledge that" *episteme* (Dickie, 2012; Fridland, 2017; Pavese, 2017; Ryle, 1945, 1949). They argue "knowledge how" is not a subcategory of "knowledge that," in their view Anti-Intellectualists see "knowledge how" as kind on par with "knowledge that" (Pavese, 2022).

More recently, epistemological conceptions of knowledge have evolved beyond the Intellectualist versus Anti-Intellectualist frame to a virtue-theoretic approach which conceives knowledge as something of value held by the knower (Greco, 2010; Pritchard, 2012; Sosa, 2007, 2009; Zagzebski, 2003). Epistemologist Ernest Sosa posited the virtue-theoretic analogy of an Olympic archer shooting at a target, where the knowledge enabling the archer's gold medal was "Accurate" (arrows hit the bull's eye), "Adroit" (produced by skilled intent) and "Apt" (attributable to the archer's skill). Sosa claims his "AAA" model of knowledge is applicable to any action or object with a characteristic of aim (Sosa, 2007; Sosa, 2020). In this conception, Sosa implies skill, *techne*, as a fundamental component of his "AAA" conception, supporting the Anti-Intellectualist conjecture of "knowledge how" *techne* being distinct from "knowledge that" *episteme.*

Within the last decade, tools for observing brain function, e.g., positron emission tomography scan, computed tomography (CT) scan, high resolution functional magnetic resonance imagining (HR fMRI), etc., have shed light new light on this ancient philosophical debate. HR fMRI has shown that linguistic processes occur in different brain regions from those performing motor skills described in Sosa's archery analogy (Buchweitz & Prat, 2013; Loring et al., 2000; Prat, 2022; Wang & Wang, 2024), which appears to support a distinction of *techne* being apart from *episteme.* Although these findings seem to end the argument, other neuroscience discoveries point to wide neuro-diversity, which may explain the divergent camps found in epistemology. In fact, recent

discoveries in neuroscience that show no two individuals possess the same synaptic architecture (Barrett, 2020; Huang et al., 2025; Wu et al., 2023) and, more to the point of this research, different people have widely divergent proclivities for processing knowledge and accessing information (Gardner, 2011; Grandin, 2023; Huddle, 2024; Sweller, 2024). Grandin's research highlights core aspects of Multiple Intelligence (MI) theory (Gardner, 2011) and points to intelligence being distributed across a spectrum of neurodiversity, thus arguing against institutional bias that she claims that filters out visual motor skills (*techne)* in favor of symbolic abstraction (*episteme).* Without stating this directly, Grandin's conjecture and MI theory parallels the Intellectualist versus Anti-Intellectualist philosophical debate and is directly relevant to information access and the organizational practice of managing knowledge. This is particularly relevant in that organizational theorist have tended to under serve *techne* in preference for concentrating on *episteme.*

2. Knowledge Management

Modern conceptions of KM began in the 1970s after a series of papers, authored by economic theorists Peter Drucker and Paul Strassman, gained wide acceptance among economic theorists researching business management strategies (Geisler & Wickramasinghe, 2015; Sallam et al., 2018; Uriarte Jr, 2008). By the mid-1990s, major corporations around the world embraced the management of knowledge as a critical enabler of competitive advantage (Leitch & Rosen, 2001). Two seminal publications, the Leonard-Barton (Willman, 1996) case study of the KM strategies of the Chaparral Steel Company (Chua, 2002; Verona, 1999; Willman, 1996) and the Nonaka and Takeuchi (Nonaka & Takeuchi, 1995) book, *The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation*, seem to have posited knowledge from the virtue-theoretic "AAA" philosophical stance of Sosa, wherein having knowledge is valuable to the organization and hence must be managed to gain competitive advantage. In this, delineating knowledge to be managed is a central concern (Bapuji & Crossan, 2007; Nonaka, 2009; Small & Sage, 2006).

3. Knowledge Distinctions in Knowledge Management

In their approach to the management of knowledge, KM's practitioners delineate knowledge types apart from distinctions of epistemic kinds. In KM, knowledge conceptions are reduced to the central operationalization that "knowledge…enables direct action" (Nissen, 2014). To the KM practitioner, knowledge is corporate capital, an asset to be organizationally obtained, stored, and distributed, which implies that it must be delineated in terms of its relationship to information. When knowledge is encoded in information, e.g., a recipe, chart, formula, or "how to" description, it is referred to as *explicit*. Knowledge retained in the mind of an individual that can be explained, or otherwise encoded into information, is said to be *implicit.* When an individual knows something, but can't explain exactly how it's done, that knowledge is referred to as *tacit*. Thus, to a knowledge manager, knowledge resides on an axis ranging from knowledge encoded in information, or *explicit knowledge*, to knowledge that can only be obtained, possibly without written or spoken explanation, from an expert, or *tacit knowledge.* The central value proposition in KM is discernment of these knowledge types to provision *knowledge flows* needed to meet organizational goals.

4. Dynamic Knowledge Flow Model

In KM, knowledge *flows* are analogous to water, designed and managed to irrigate organizational activities, e.g., operations, production, sales, marketing, distribution, etc. The American Productivity and Quality Center defines knowledge flow as a seven-step process of creating, identifying, collecting, reviewing, sharing, accessing, and using knowledge (American Productivity Quality Center, 2023). This definition is reflective of Nonaka and Takeuchi's original writings, in which knowledge flows were conceived as a "spiral process" of externalizing what experts know implicitly into "information objects" (explicit knowledge) to be digitized, stored, and endured in IT long after the expert(s) retired or otherwise left their organizations (Nonaka et al., 1996). Nissen's "Knowledge Flow Visualization Model" expands Nonaka and Takeuchi's conception of the "knowledge spiral" (see Figure 10) to visualize knowledge flows occurring in a three-dimensional

vector space delineated across axes of "Life Cycle," "Explicitness," and "Reach" (Nissen, 2002).

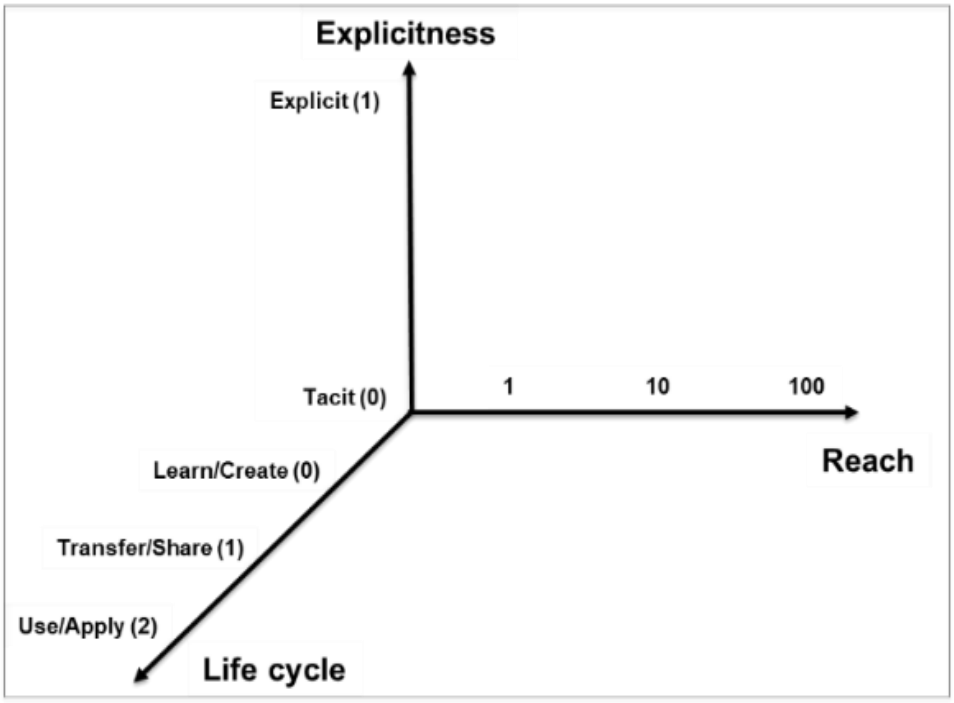


Figure 10. Knowledge flow visualization model. Source: Nissen (2002).

In his knowledge flow theory (KFT) model, Nissen (2006) positioned knowledge flow vectors in a conceptual volume defined in three axes, X, Y, and Z. Along the Z-axis of "Life Cycle," the stages of knowledge flow, e.g., APQC's creating, sharing, applying, etc., are plotted. Along the Y-axis degrees of knowledge "Explicitness," ranging from tacit to fully explicit, are plotted as a function of change over time. On the X-axis of "Reach" degrees of socialization, e.g., from transmission source to the number of receiver individual(s), group(s), organization(s) benefiting from the knowledge are plotted. In this way, Nissen's flow vectors visualize knowledge flow relational dynamics of direction, speed, and power (for a specific flow described) where knowledge held by an individual is made "explicit" (X-axis) through a willful act of "socialization," then is "externalized" into an information medium, "combined" and "refined" along the "Reach" (Y-axis) as a function of "life cycle" (Z-axis). Eventually knowledge is evolved or flows on each axis

until it is documented as "Policies and Procedures." Each vector depicts the direction of the flow, the time anticipated for the flow to take place, and the force of the flow in terms of the complexity or expertise needed to ensure the knowledge is well received and remembered, see Figure 11 (Nissen, 2006).

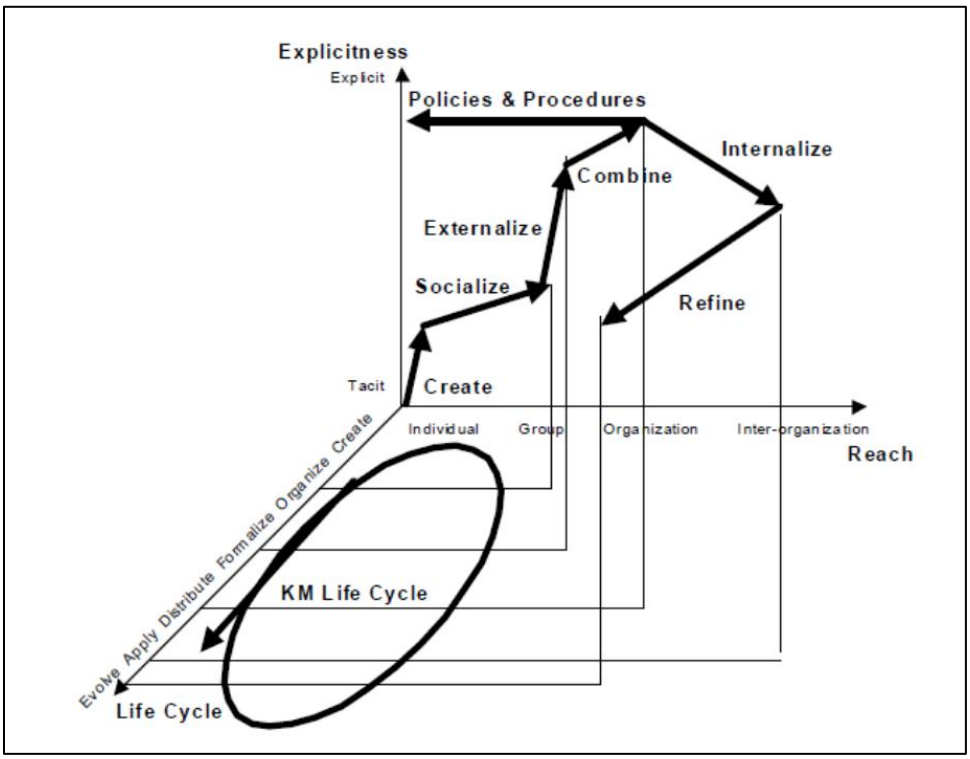


Figure 11. Flow vectors in Nissen's knowledge flow theory model. Source: Nissen (2006).

5. Knowledge Flow and Information Technology

From the beginning, KM practitioners have looked to IT in support of knowledge flow. IT support of knowledge flow stems from the principle that "flows of knowledge" require supplementary "flows of information," and in turn data and signals (Nissen, 2014). The kind of knowledge exploited via IT in KM has primarily been explicit "knowledge that," *episteme* propositional knowledge that can be stored as symbolic or pictorial knowledge objects, or "*strings"* (Collins, 2019). Examples include recipes, instructions, formulas, drawings, videos, etc., and expert capture, or the mapping out of expert locations

or addresses persisted in databases for access, distribution, or live technical exchange (Hutchins et al., 2007; Nissen & Bordetsky, 2011). This KM focus on *episteme* occurs largely because other knowledge kinds, *empeiria* and *techne,* are, to varying degrees, held tacitly by experts wherein articulation of this expertise into "symbolic strings" is either impractical or completely impossible (Collins, 2019). In the case of ship maneuvering, the knowledge in question is largely performative *techne* based on acquaintance *empeiria* where experienced individuals who have built up *techne* in tacit knowledge are able to perform appropriately in dangerous situations.

B. THE PROBLEM OF TACIT KNOWLEDGE FLOW

A common refrain for KM practitioners, including Nissen, is the "problem of tacit knowledge;" how can knowledge flow if it cannot be articulated or otherwise be made explicit (Nissen, 2014)? This problem regarding the mercurial nature of tacit knowledge traces back to Michael Polanyi's famous quote that "one can know more than one can tell" (Polanyi, 1967), and remains central, particularly when attempting to flow *techne* or *empeiria.* Here, epistemological distinctions inform the KM practitioner's needs and point to potential application of XR to flow tacit knowledge by leveraging synthetic experience.

1. Tacit Knowledge and Feedback

Daniel Kahneman, the 2002 Nobel Prize Laureate in Economics, summarized the problem of ship maneuvering as being difficulty in recognizing feedback cues "because of the long delay between maneuvering actions and their noticeable outcomes. Whether professionals (Harbor Pilots) have a chance to develop intuitive expertise depends essentially on the quality and speed of feedback, as well as on sufficient opportunity to practice" (Kahneman, 2011, p. 241). Kahneman's writing on recognition and feedback referred to fellow Economics Nobel Laureate (1978) Herbert Simon's theories of expertise wherein he states that "a major component of expertise is the ability to recognize a very large number of specific relevant cues when they are present in any situation, and then to retrieve from memory information about what to do when those particular cues are noticed" (Simon, 1996). Additionally, Simon stated, "The decision maker of experience has at his disposal a checklist of things (cues, conscious and subconscious) to watch out for before

finally accepting a decision…one of the ways we can acquire a checklist of things to watch out for is to learn how the world really works and adapt ourselves accordingly" (Simon, 1996). In this, both Kahneman and Simon imply that shiphandling is performative, where the pilot leverages tacit knowledge gained by experience over time.

Given the above, it thus appears that explicit knowledge encoded in a book, a video, or even a "how to maneuver a ship" conversation with a harbor pilot (or engagement with other "strings of symbols or sounds") is likely insufficient to flowing this knowledge as much of it resides tacitly within the expert (Collins, 2019). To further improve understanding of the difficulty in flowing knowledge in the case of maneuvering a ship (or any case wherein acquiring spatial-temporal performance skills is needed), it is necessary to consider knowledge flow friction.

2. Tacit Knowledge and Knowledge Flow Friction

The concept of knowledge flow friction, an extension to KFT, posits a "…loss or degradation or delay in the conveyance of knowledge between the source and the receiver(s)" (Shigley, 2021). In this concept, knowledge flow friction is not the same as a knowledge barrier, something blocking the flow between the sending agent and the receiver(s), but rather a "delaying" or "distorting" of knowledge after it has been collected and imparted by the organizational communicator in the transfer link to the designated recipients; it is the cause of the loss of knowledge between the knowledge transmitter and the knowledge receiver. The transfer link can be thought of in terms of "clarity": is the knowledge clearly presented to the receiver? This question reflects Collins' conception that even a long string of symbols may not reproduce tacit understanding held by an expert (Collins, 2019) and begs the question: how can one clearly articulate experiential knowledge to the novice, when clarity of subtle cues is largely held as tacit knowledge by the expert?

3. Procedural Knowledge and Embodied Cognition

Yet another way to consider the problem of knowledge friction in flowing tacit knowledge is from the perspective of cognitive science, specifically the theoretical frames of embodied and extended cognition. A chief concern from this perspective is knowledge

distribution, where knowledge is not singularly held in an individual's brain, but is distributed in the body (embodied cognition) and in the environment (Extended Cognition). From the frame of embodied cognition, procedural knowledge, or *techne,* is held in the sensory-motor constellation of the entire person (Foglia & Wilson, 2013; Lakoff, 2012; Shapiro, 2019; Wilson & Golonka, 2013). Thus, knowledge is not just in a person's head but held in every aspect of their embodied experience, incorporating in situ conscious and subconscious experience in the world in interaction with bodily memory, e.g., proprioception, gut feelings, emotions, etc. In Extended Cognition, procedural knowledge extends beyond the body of the individual and includes the affordances of the interactions and memory components surrounding the body (Chalmers, 2019; Clark, 2010; Clark & Chalmers, 1998; Ludwig, 2015). Thus, Extended Cognition posits knowledge held in the environment e.g., instrument displays inside a ship's bridge, fellow watchstanders, charts and instructions, etc., as critical knowledge components of the procedural knowledge tacitly held and leveraged by the expert in navigating the ship safely (Collins, 2019; Hutchins, 1995; Perry, 2003). Articulation of this tacitly held, procedural knowledge into "symbolic strings" becomes intractable given the extreme variation of bridge team members, details of instrumentation readouts, and spatial-temporal components possible across an infinite number of situations a ship might face.

4. Embodiment and Tacit Knowledge

Theories of embodied and extended cognition have gained greater acceptance in the cognitive sciences and frame human cognition as more than "brain as computer" or "brain performing symbolic representational computation." These theories expand the frame to consider the greater whole of an organic and analog system of brain, human body, and the environment, both physical and social (Chalmers, 2019; Clark, 1999, 2023; Foglia & Wilson, 2013; Lakoff, 2012). Additionally, embodied and extended cognition theory hypothesizes knowledge production as an interaction or structural coupling (Maldonado et al., 1988; Mingers, 2001; Mingers & Willcocks, 2014) of the whole body (not just "a brain") to its environment through whole body sensory-kinesthetic processes (Alibali & Nathan, 2012; Clark, 2015; Joyce, 2007; Kanellopoulos et al., 2021). In yet a further expansion of this idea, multiple theorists postulate that participation of the body in learning

also occurs through social interactivity such as expressions, manners, and emotions (Alexias & Dimitropoulou, 2011; Barrett, 2020; Barrett et al., 2019; Barrett et al., 2007; Clark, 2023).

Concepts of embodiment and extension in cognition have been debated for years in the disciplines of philosophy of science and philosophy of mind (Block, 1993; Chalmers, 2019; Clark, 1999, 2023; Dennett, 2001; Dennett, 2018; Ludwig, 2015; Mingers, 2001; Varela et al., 2017), but recent empirical evidence, particularly discoveries made using the non-invasive brain-monitoring tools of HR fMRI and diffusion tensor imaging (Huang et al., 2025) are beginning to support embodied and extended cognitivists' claims that brain function is indeed "structurally coupled" to its surroundings (Barrett, 2020; Magee & Johnston, 1997; Popp et al., 2024; Prat et al., 2022; Wu et al., 2023). Mingers' diagram in Figure 12 (Mingers, 2001) provides a succinct visual representation of embodied cognition showing the individual as informed by, and simultaneously informing, the external world in a process of structural coupling.

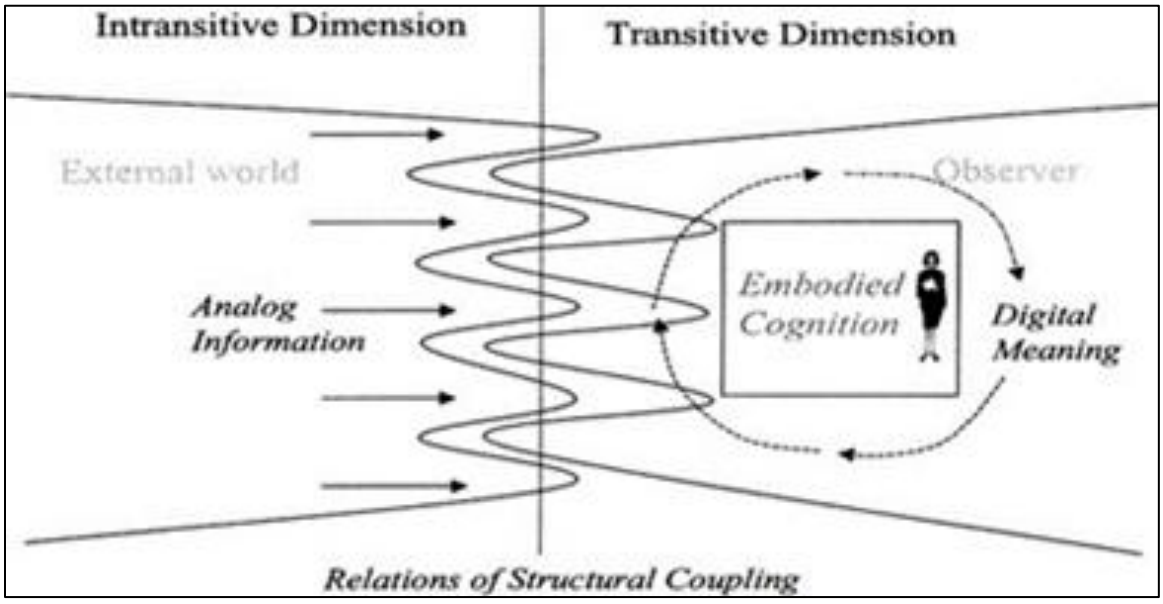


Figure 12. Embodiment and structural coupling. Source: Mingers (2001).

If correct, these tenets of embodied and extended cognition imply that procedural knowledge flow is itself an embodied process, wherein the static presentation of an

information object (explicit knowledge, e.g., instructions) is wholly insufficient to convey tacit knowledge of performance skill, *techne,* gained via experience, *empeiria.* From this point of view, the flow of knowledge needed to recognize subtle cues and feedback, Simon's "checklist of things to watch out for," is gained via the interaction of the structurally coupled "intransitive dimension" of a dynamically changing real world to the "transitive dimension" of the body's adaptive processes of neural connection "tuning and pruning" to improve predictions and motor skills within the individual (Barrett, 2020; Clark, 2015, 2023; Mingers & Willcocks, 2014). Herein lies our source of tacit knowledge friction: how does one clearly flow experiential knowledge when structural coupling to an external world experience is impossible, dangerous, or simply impractical to perform?

C. TACIT KNOWLEDGE FLOW AND SYNTHETIC EXPERIENCE

In this section, employment of physical mockups, used to generate synthetic experiences, is discussed as a means of imparting tacit knowledge to novices. It is then argued that analogous information-based synthetic experience can be generated using XR. The research rationale for employing XR VE's to enhance knowledge flow is then provided, followed by the gaps in current knowledge this research intends to fill.

1. Mockup-Based Synthetic Experience

Employing a danger-free synthetic experience to flow tacit knowledge, even if not stated in this way, has been performed for nearly a century. Edwin Link's 1929 Blue Box Trainer (Allen, 1993) is credited as the first mockup simulator used to train airplane pilots. The term "mockup," originating in 1876 military ship building to mean "make an experimental model" (Brown, 2009), is today applied to any kind of training facility used to provide *empeiria* (acquaintance) for learning an extended embodied process, e.g., piloting an aircraft, navigating a ship, operating a power plant, etc. A mockup system of systems provides a synthetic experience wherein aspects of the real world, e.g., sight and sound, are replaced with physical and electronic emulations enabling a flow of subtle cues and responses to human feedback. By immersing an individual in a mockup VE, an expert can present a synthetic experience akin to what would be expected in dangerous, real world situations, thereby enabling critical flows of performative tacit knowledge as the individual

interacts and learns from the experience. The literature of embodied cognition provides reasoning for how a synthetic experience facilitates this kind of flow by providing levels of immersion and presence near those of actual reality (Bailenson, 2019; Clark, 2010, 2023).

A key tenet of embodied and extended cognition is that of a *connectionism*, the theory that mental processes are composed of synaptic connections between neurons wherein the strength of each connection evolves based on the experience of an individual's body structurally coupled to its environment (Hebb, 2005). A famous conjecture of Donald Hebb that "Neurons that fire together, wire together" has been an accepted tenet of neuroscience, proven empirically, and now referred to as Spike-Timing-Dependent-Plasticity or STDP (Gilson et al., 2010). A ramification of STDP is that neurosynaptic network structures of links, nodes, clusters and hubs form scale-invariant architectures in response to their motor-sensory perceptual interaction (structural coupling) with the environment. Hence, a force like gravity, because it is always and everywhere the same, becomes an invariant representation in the brain. This is relevant in so far as most of human perception is grounded in the invariance of that which is bodily perceived in its interaction with the world, gravity being just one example of extended coupling.

If human perception is grounded in the invariance of bodily experience (Barsalou, 2008, 2010), a critical consideration in extending Nissen's axis of explicitness is the explication of synthetic experience as means to flow tacit knowledge. In keeping with Barsalou, invariances of synthetic experience must nearly match those of the real world, or else the flow of subtle feedback cues will fail to aid novice understanding. In the case of this specific research, given the steadily growing nautical experience of all participants, shiphandling characteristics, ocean dynamics of wind, tide, and current, the bridge layout of controls and information displays (to name but a few environmental details) must be representative of those of the actual ship and how it behaves in the real world. When synthetic experience is created in such a way (invariant matched), then it is possible to extend what is explicit to not just strings of words, numbers, recorded conversations, videos and so on, but to provide rich, scenario-encoded VEs to inform tacit experience. Figure 13

shows a side-by-side comparison of a USN bridge (left picture) and its mockup, the currently accepted VE used to flow performative skill in the USN (right picture).

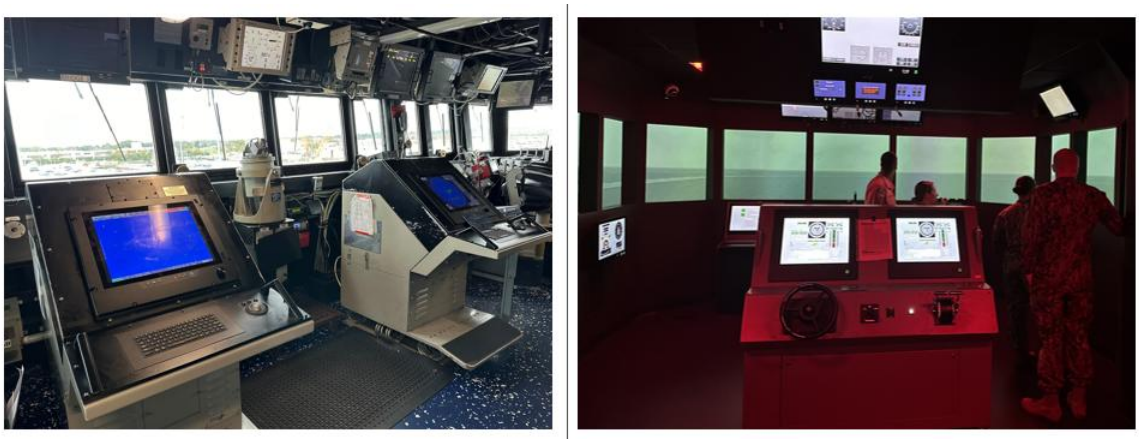


Figure 13. Photograph of an actual U.S. Navy bridge (left) and the representative mockup virtual environment (right).

As seen in Figure 13, the ship-based mockup is not limited to single individuals but designed to facilitate tacit knowledge flow in teams. This is particularly important in the shiphandling case as large ships are not controlled by a single individual, but by a bridge team wherein feedback occurs across an ensemble of watchstanders, their equipment, communication systems, and control interfaces. As subtle and delayed as it may be, feedback must be addressed by the whole team working in concert with each other in a specific ship's bridge environment, where decisions are made within a distributed cognitive frame, unique to the ship, crew, and tasks at hand. This in turn implies that tacit understanding must flow across the social interactions of the team in response to the situation presented in the VE. Mockups provide an opportunity to develop these skills, but as noted in Chapter I, they are expensive, limited in availability, and require a large infrastructure in which to house them (see Figure 5).

2. Extended Reality–Based Synthetic Experience

XR offers a new way to expand knowledge flow in terms of what can be made explicit in information. With an XR-generated VE, information flow is no longer limited

to propositional knowledge, *episteme,* but can also include acquaintance knowledge, *empeiria.* Figure 14 shows a comparison of the physical mockup VE (left picture) and the informational XR VE (right picture); the locational dimensions and spatial relationships of the XR VE can be seen as largely the same as the mockup facility and the actual guided missile destroyer (DDG) bridge in Figure 13 (left picture).

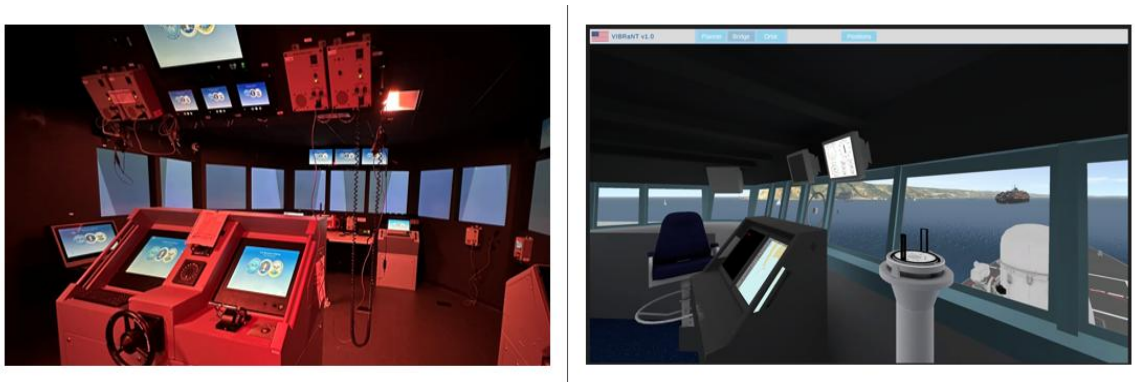


Figure 14. Fixed, physical mockup virtual environment (left) and mobile, information-based virtual environment (right).

If an XR VE can approximate the performance of the mockup, expert skills held as tacit knowledge of an expert's embodied experience can be made explicit in an information object (the scenario-encoded VE) and presented for the purpose of flowing inarticulable understanding needed to build tacit shiphandling knowledge in the novice. This is conjectured because XR VE allows the expert to present the whole of an embodied situation as an experience, rather than an explanation, or story, or other kind of verbal or written information flow. This kind of rich information flow, and the feelings of presence and active agency it brings, freed of physical infrastructure, can now facilitate an experience, not unlike that of a physical mockup. The information-based VE enables structural coupling of the novice to the experience designed by the expert. In this way the VE serves as an analog of Mingers' intransitive dimension, the real world of experience depicted in Figure 12, but in the XR VE only two primary sense vectors, sight and sound, are provided (see Figure 15). In this the novice is immersed in a scenario made explicit by the expert, allowing for critical spatial-temporal tacit knowledge flow as a reaction to the

event sequence provided in the scenario. While the verisimilitude of the information-based VE cannot yet match the mockup, XR fidelity continues to improve and, as it is hypothesized, it already provides the possibility for developing tacit knowledge of the subtle spatial-temporal cues needed in shiphandling.

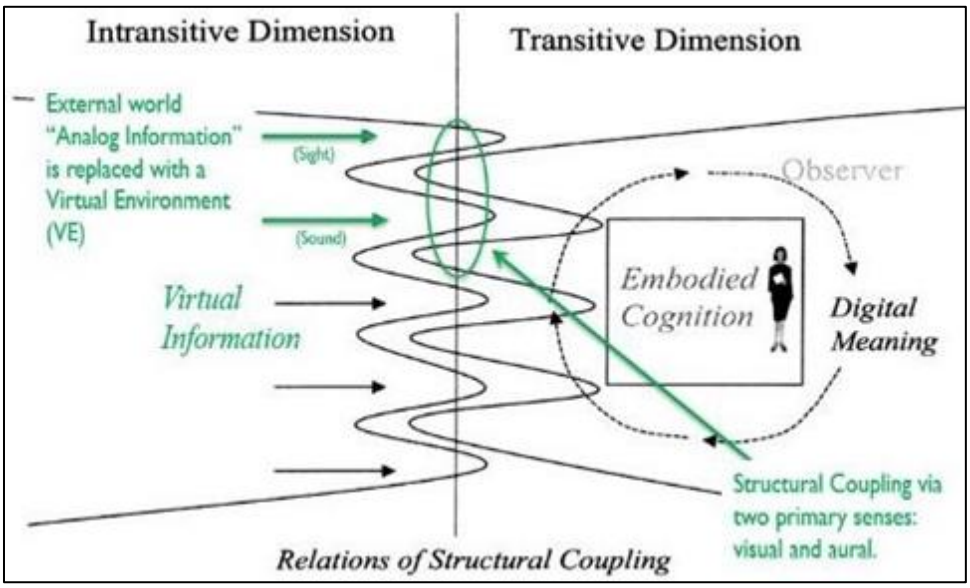


Figure 15. Mingers' structural coupling to an extended reality–based synthetic experience. Adapted from Mingers (2001).

3. Research Rationale

The cost of providing VR-based training has fallen substantially, even as the capability of commercially available equipment has increased dramatically. This market trend makes XR-supported knowledge flow much more achievable for organizations today than ever before, and these trends (falling costs and increasing capability) are accelerating. Additionally, by instantiating knowledge bearing VEs in information, e.g., XR, potential knowledge "reach" as defined by Nissen becomes greatly extended as these technologies are lightweight and mobile, enabling VE participation from any physical location able to connect to the Internet. These factors afford an organization, e.g., the USN, an opportunity to train teams in a common VE, at any time, from any place, together or apart, with or without an instructor or their Commanding Officer. In addition, because the VE is

information-based, it can be easily customized to match any environment desired by the organization to the degree of verisimilitude desired. If realized, XR affords flexibility for the USN to provide every ship in the fleet with tailor made VE customized to the exact configuration of the bridge layout of each ship, which is simply not possible in a physical mockup.

D. RESEARCH GAPS AND HYPOTHESES

Research interest in XR uses for training has greatly expanded and continues to garner interest from a wide sector of academic and commercial interests (Mouloua et al., 2023). Prior research has shown limited efficacy of using XR synthetic experience to facilitate different kinds of military training (Caso, 2024; McCoy-Fisher et al., 2019; Ross & Gilbey, 2023; Smart, 2022) but only one study, *Student Naval Aviation Extended Reality Device Capability Evaluation* (McCoy-Fisher et al., 2019) was relevant to this dissertation. The McCoy-Fisher et al. study focused on a comparison of legacy pilot mockup training compared to a VR presentation; however, it was limited to aviator indoctrination training, analysis of single individual responses, and, having been performed in 2019, did not incorporate the more advanced XR hardware used in this research.

In the specific case of shiphandling, a recent study (Vakil, 2023) surveyed potential XR uses for maritime training in India, but did not go beyond providing general arguments for applying XR for training in the maritime industry. Additionally, two older studies were found: (Grant & Magee, 1998; Hays & Vincenzi, 2000), but these studies were limited single user applications and employed technologies that are dated by today's standards. In the Grant and Magee study, a between-subjects design compared Canadian officer of the watch performance improvements based on VR training in relation to at sea training and thus did not compare bridge team training responses to different training systems: XR and mockup. In the Hays and Vincenzi study, the capacity of exploiting VR to train submarine conning officer skills while operating on the surface was explored, but as mentioned was limited to a single individual using an extremely complex early VR system tethered to an overhead gimble limiting movement and team participation with other watchstanders.

A comparison of multiple participants practicing shiphandling and bridge resource management in a legacy physical mockup trainer to that of an XR facilitated information-based trainer has not been found in the current literature. Theoretical conceptions exploring the replacement of linguistic or symbolic information with spatial-temporal XR VEs to enable an information only transit link for carrying knowledge bearing experience of tacit knowledge for exploitation by teams has also not been found in the literature. As noted, an information only process would expand organizational ability to affordably flow Simon's tacit "checklist of things to watch out for."

To address these gaps, this research evaluates and characterizes how the information access of current XR technologies can be exploited to flow tacit knowledge of maritime skill *techne* from synthetically generated experience *empeiria.* While the specific focus is on USN combatants, the larger goal is to better understand and generalize to other instances demanding knowledge flow of this kind of tacit, team-based spatial-temporal understanding. The following hypotheses were statistically evaluated:

H_1: Knowledge flow facilitated via XR *will* (for mapped skills and tasks) meet or exceed facilitation provided via a mockup trainer.

H_2: Extraneous cognitive load produced by XR *will not* (for mapped skills and tasks) exceed extraneous cognitive load produced by a mockup trainer.

H_3: Facilitation of critical situational awareness (SA) skills via XR *will* meet or exceed SA skills facilitation via a mockup trainer.

H_4: Knowledge acquired via XR *will*, when preceding mockup training sessions, complement and enhance performance observed in the mockup trainer.

H_1 to H_4 serve as alternate hypotheses of the informational XR experimental condition, in comparison to the null hypotheses (H_0) of the physical mockup trainer control condition.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHOD

The following chapter describes the experimental method employed to evaluate and characterize the application of XR to facilitate knowledge flows required for training USN bridge teams skills to effectively navigate and safely maneuver their ships at sea in a variety of different situations. The research design, participant selection, data collection, preparation, and analysis methodology are described, to include specific details of how the XR experimentation was carried out in coordination with USN mockup training.

A. RESEARCH DESIGN

The research design, framed in the philosophical stance of a post-positive, quantitative, quasi-experiment, evaluates and characterizes USN bridge team responses to XR facilitated knowledge flow in comparison to the USN's mockup NSST. Given the novelty of the XR employed, an experimental approach was selected with the NSST serving as the control condition and Office of Naval Research's (ONR) VIBRaNT XR system serving as the experimental condition.

The independent variable, *Information Access,* analogous to a treatment, is provided in two different conditions, one "access" being that of the mockup, or control condition, the other being that of XR or the experimental condition. The dependent variables are the results of the post-event survey results collected after each treatment, mockup and XR respectively (Table 1).

Table 1. Experimental design and independent and dependent variables.

Independent Variables (Information Access) = Treatment Types		Dependent Variables = Survey Responses
Mockup (Treatment A)	XR (Treatment B)	
Mockup (control condition)	XR (experimental condition)	NASA Task Load Index Survey
Mockup (control condition)	XR (experimental condition)	System Usability Scale Survey
Mockup (control condition)	XR (experimental condition)	Bridge Watch Team Assessment Survey
Mockup (control condition)	XR (experimental condition)	OOD self-assessment survey
Mockup (control condition)	XR (experimental condition)	Instructor Assessment Survey *

1. Overarching Research Design

This study employed a within subjects quasi-experimental design to compare participant perceptions of cognitive load, system usability, and maritime training facilitation of a physical mockup, the USN's NSST and an information-based XR system, ONR Tech Solution's VIBRaNT. Following the methodological framework of Cook, Campbell, and Shadish (Cook et al., 2002), the design permitted examination of causal relationships under naturalistic USN operational conditions while controlling for participant-specific differences and order effects through a counter-balanced design.

The model for this design can be expressed as:

$$Yitc = \mu + \alpha i + \beta t + \gamma c + \varepsilon itc,$$

where:

- $Yitc$ is the observed score for individual i at time t in condition c.
- μ is the grand mean.
- αi is the participant (or team) effect which controls for individual differences.
- βt is the time order effect.
- γc is the treatment effect (XR versus NSST).
- εitc is the residual error.

Each participant experienced both treatments or training conditions, XR and NSST, in a counter-balanced sequence. This within subjects configuration effectively controlled for individual differences (αi) and allowed the direct estimation of the treatment effect (γc). Counterbalancing mitigated order effects (βt), such as learning and fatigue, and residual error (εitc) reflected random variance not explained by the model.

Quasi-experimental designs are particularly suitable when random assignment is not feasible or in situations where the sample is one of convenience. Strong design logic and control features can approximate experimental rigor of a fully randomized trail (Cook

et al., 2002). In this study, intact bridge watch teams served as the primary units of analysis, preserving ecological validity while maintaining control through repeated measures within the same participants. The design thus supported generalized causal inference through covariation, temporal precedence, and control of plausible confounds.

The design of analyses utilized several layers and survey instruments to address each of the research hypotheses. Wilcoxon Signed-Rank Tests (WSRTs) were applied to paired item-level data to estimate the treatment effect (γc) under non-parametric assumptions. Hotelling's T^2 test served as the multivariate analog, examining the joint distribution of dependent variables across conditions. Bonferroni corrections protected against inflated Type I errors across multiple comparisons, and principal component analysis (PCA) was used to assess construct validity and confirm the underlying factor structure of the measured items.

a. Threats to Internal Validity and Associated Controls

The research design addressed internal validity by controlling for a number of factors posing threats or bias errors to include: History and Maturation, Testing and Instrumentation, Selection, Statistical Error, and Construct Validity (Table 2).

Table 2. Threats to internal validity and controls.

Threat to Internal Validity	Control and Mitigation Strategy
History and Maturation	Counterbalancing of condition order and matching training scenarios were employed to reduce historical and maturation effects. Participant bridge teams alternated between XR and NSST conditions to minimize cumulative learning or fatigue biases. Scheduling controls ensured equal rest intervals and scenario complexity across both environments.
Testing and Instrumentation	Standardized administration procedures and identical data collection instruments (NASA TLX, System Usability Scale (SUS), bridge watchstander self-assessment, etc.) were used across both conditions. All measures were deployed using identical surveys, time allocation, collection environment, and event facilitation to prevent measurement drift or experimenter bias.
Selection	A within subjects design ensured each participant served as their own control, effectively neutralizing inter-individual variability. All participants met the same inclusion criteria and were drawn from operationally similar Navy watch teams to maintain equivalence.
Statistical Error	Non-parametric statistical methods (WSRTs) were used to accommodate non-normal data distributions. Bonferroni corrections mitigated Type I error inflation from multiple comparisons, while effect size reporting (Cohen's d) provided context for practical significance beyond statistical thresholds.
Construct Validity	PCA was used to confirm the coherence of latent factors across conditions, ensuring construct consistency for TLX, SUS, and self-assessment measures. Stable factor structures indicated that differences observed between XR and NSST reflected true treatment effects rather than measurement artifacts.

b. Threats to External Validity and Associated Controls

Several steps were taken to address external validity and generalizability of the study's findings across persons, settings, treatments, and outcomes (Cook et al., 2002). The design sought strong ecological validity by aligning participants, training settings, and tasks with authentic USN bridge operations. Both the NSST and XR conditions took place in the same Navy schoolhouse environment under the supervision of MSTCPAC Master Mariner instructors and Surface Warfare Schools Command representatives. Task scenarios such as underway replenishment, Contact Management, and channel transit mirrored actual USN curricula provided at MSTCPAC. The XR system's VE was modeled on USS Paul Hamilton DDG-60's bridge with higher visual fidelity than the NSST mockup, and both conditions involved the same complement of watchstanding roles. These

design elements ensured that findings from the study could be confidently generalized to fleet-level operational contexts (Table 3).

Table 3. Threats to external validity and controls.

Threat to External Validity	Control and Mitigation Strategy
Interaction of Selection and Treatment	Participants were qualified Navy bridge watchstanders drawn from an operational training population. Within subjects pairing ensured all participants experienced both NSST and XR conditions, controlling for individual experience levels.
Interaction of Setting and Treatment	Both training conditions occurred at MSTCPAC within a U.S. Navy schoolhouse environment. This maintained consistent organizational culture, supervision, and real world operational norms.
Interaction of History and Treatment	Training sessions for both conditions were scheduled within the same time period and context, avoiding historical or cohort differences. Scenarios were delivered using equivalent temporal spacing and instructional staff.
Interaction of Treatment Implementation	The same Master Mariner instructors supervised both XR and NSST sessions. Similar brief/debrief protocols were followed, ensuring consistency in scenario facilitation and performance expectations.
Interaction of Measurement and Treatment	The same instruments (NASA TLX, System Usability Scale (SUS), and Attachment H) were administered across conditions under standardized conditions. Measures reflected operationally relevant constructs, supporting generalization to real fleet applications.

In summary, the research design provided a rigorous and ecologically valid means to examine facilitation of knowledge flow across physical (mockup) and informational (XR) conditions. The within subjects quasi-experimental approach allowed for the detection of meaningful differences in workload, usability, and team cognition while controlling for individual variability.

2. Participants and Sampling Design

Participants included USN officer and enlisted crew members assigned to perform duties as a ship's bridge watch team. Watch teams were selected as the unit of analysis because the same individuals typically train and serve aboard ship together as a group and hence learn together as a team. A within subjects sampling design, purposive of the USN bridge team organizational construct, collected individual responses from each bridge team

watchstander after being exposed in both control and experimental conditions. Respondents completed surveys in the company of each other immediately after performing in each condition in the same environment (classroom) for each condition. The sample collected was one of convenience from bridge crews conducting training at the MSTCPAC located at 32nd Street Naval Station in San Diego, CA. Bridge Team watchstanders sampled included:

1. OOD
2. JOOD
3. CONN
4. NAV
5. HELM
6. LEE HELM

There were also a few other "non-typical" crew members (HELM safety, boatswain's mate of the watch, etc.).

The within subject's design ensured that each bridge team was observed in both control and experimental conditions on the same day, with the same bridge team composition (number of personnel assigned and their position for the day's training), the same training scenario, and the same MSTCPAC instructor in both conditions. Within subject's observations were counter-balanced such that half the number of bridge teams were observed first in the control condition (NSST), with the other half being observed first in the experimental condition (XR). Data collected included bridge team watchstander responses to survey questionnaires, performance related observations, and eye tracking data collected from bridge team OODs participating in each condition. The same survey questionnaires were used in both conditions, and three written qualitative exit interview questions were posed to all bridge team members to get their impressions of using XR for bridge training.

3. Tasks Requirements Matching in the Experimental Virtual Environment

In both conditions, a VE was created to generate a variety of different nautical situations. The situations designed into the VE depended on the learning objectives desired, including entering a harbor, transiting a crowded sea lane, maintaining station behind an aircraft carrier, etc. Thus, the scenario represented in VE was designed to flow knowledge needed in building the different skills required of each bridge team. The knowledge-carrying agent (the VE) was encoded with MSTCPAC staff expertise wherein each scenario unfolded in space and time to challenge the bridge team's recognition and decision processes. Knowledge flowed as a result of bridge teams *making acquaintance with* or *experiencing* each situation facilitated in the VE as choreographed by the expert.

For this reason, careful consideration of matching scenario in both the mockup (control condition) and XR (experimental condition) presentations was paramount to the research design. To make an appropriate comparison, the XR system needed to be able to present synthetic experiences in situations that matched the mockup trainer's capability, which was not always possible. XR VE limitations were mostly budgetary in nature as the research team did not have sufficient resources to match the full capabilities of the NSST. However, despite this budgetary limitation, there were a significant number of instances in which the XR VE might accommodate scenarios correctly in concert with the NSST, and these were the scenarios selected. In order to match the XR VE to the NSST VE, two principle considerations were made: (1) the required internal bridge system environment, e.g., the bridge equipment and layout details needed for the exercise (see Appendix A), and (2) the external world environment, e.g., land masses, surrounding contact density, nature of operation, etc. (see Appendix B). The selected scenarios included *Contact Management*, *Underway Replenishment, Harbor Transit, Plane Guard, OOD Math,* and *Restricted Maneuvering.*

After determining which MSTCPAC NSST scenarios adequately presented in XR, the corresponding XR VEs were created and validated. Validation of matching XR scenario VEs confirmed details in XR, such as ports represented (i.e., San Diego in NSST's Harbor Transit scenario), contact density and vessel types (i.e., Malacca Strait in NSST's Contact

Management scenario), maneuvering dynamics (i.e., as in NSST's Underway Replenishment scenario), etc., were adequately matched to the situations presented by MSTCPAC instructors in the NSST.

Having determined which XR VE scenarios were adequate, observation periods were selected based on when the ship's training schedules for the NSST aligned to the scenarios mentioned above. This mitigated the logistical problem of requiring bridge crew participation in extra or separate research dedicated events to obtain comparison data needed to meet research goals.

4. Participant Access and Experiment Alignment

To gain access to bridge team participants, a Letter of Approval (Appendix C) was obtained from the Commander of Naval Surface Forces (CNSF) granting permission to conduct research at MSTCPAC. Permission was required before participants could be recruited from ships assigned to Naval Station San Diego. Ships selected were identified based on the MSTCPAC training schedule during the months of February to April 2025. Access to participants and bridge team recruiting was coordinated between the CNSF Fleet Navigator, the CNSF Science Advisor, the MSTCPAC Surface Warfare School Command's (SWSC) representative, and ship's commanding officers (COs). As noted, bridge team participation was designed to be coincident with the ship's normally scheduled NSST bridge training (mockup control condition) at MSTCPAC on a not to interfere basis such that participating bridge teams had 70 minutes available to participate in the XR experimental condition. MSTCPAC instructors were also recruited to participate and observe during the XR sessions and completed instructor surveys when time allowed.

A pre-event coordination email (Appendix D) was sent by CNSF to each ship before their regularly scheduled training requesting bridge team participation in XR. Staff coordinating the crew availability did not have a role in the study (outside coordination) and were not in any position of authority over participant involvement as each participant could at any time withdraw from the study without repercussion. In addition, several bridge teams were recruited impromptu on site without receiving the coordination email but were also informed they could refuse participation without penalty or negative repercussion. In

addition, all participants were required to read a privacy statement (Appendix E), read and sign the informed consent document (ICD) (Appendix F), and complete the pre-screening questionnaire (Appendix G) before participating in the study.

Alignment of the experimental schedule to the routines of MSTCPAC's training schedule was critical to the data collection efforts; without such alignment (and access) the research effort would not have been possible.

B. DATA COLLECTION

The following section details sample size considerations and participants selection criteria, to include limitations and participant population details.

1. Sample Size

G*Power™, a free statistical analysis software developed by Heinrich Heine University Düsseldorf, was used to determine the sample size needed to evaluate the XR condition (H_1 to H_4) in comparison to the Mockup condition (H_0). To perform the calculation, a statistical significance level α (α being the probability of making a Type I error) of $\alpha = 0.05$ (5%) and a power β (β being the probability of making a Type II error) of $1 - \beta = 0.80$ (80%) were selected. Using these values for a within subject's research design, the optimal sample size of n=27 was computed where "n" represented the number of USN bridge teams required to participate in both conditions. As noted, bridge teams were selected from each ship-based on those already scheduled for routine mockup training at MSTCPAC.

2. Sampled Participants

Participants sampled had to be 18 years of age or older and United States citizens serving on active duty in the USN. Participants were excluded if they had a prior history of motion sickness, vertigo, epilepsy, or similar conditions. (The exclusion criterion was required because participants needed to wear a Meta Quest Pro or Meta Quest 3 HMD which has previously shown potential to cause motion sickness.) Additionally, the sample included bridge crews from Arleigh Burke Class Guided Missile Destroyers (DDGs) in order to match the crew's experience to the VE designed for the XR evaluation. Other

bridge VE's were considered and were possible to design, a key advantage to an information only training system, but these environments were not funded for this research.

The number of individual participants sampled assumed seven watchstanders per bridge team, such that n=189 participants initially determined as possible, however, not all teams had that many watchstanders scheduled for training at the same time in their regular rotation of MSTCPAC training. Given this limitation, the sample size in terms of the number of individual participants in each bridge team was allowed to vary from team to team, but the composition of each team sampled was held constant in each condition. An example would be running a crew of four officers through both conditions on one day and running a larger crew (who manned more watch standing positions) on another day. The total sample had 158 participants perform the final comparison, with an additional 40 participants involved in design and pilot studies. All participants were also provided with a questionnaire to inquire as to simulator sickness response (Appendix H), but none of those sampled noted any significant adverse responses to the XR experience.

Individuals making up each bridge team included typical junior grade naval officers, male or female, ranking from Ensign (O1) to Lieutenant (O3). When a bridge team brought along enlisted HELM and Lee HELM watchstanders, they were included in the team observed. Individuals recruited had varying degrees of experience and prior training; however, participants sampled were deemed by MSTCPAC instructors as representative of the larger population of crews assigned to bridge duties. Additionally, multiple bridge teams participated from the same ship, but in these cases the composition of watchstanders was different in each observation period. Because each bridge team was counted as one observation period or sample, the total number of individuals was allowed to vary so long as the threshold of 27 bridge teams was met to complete data collection. Appendix Q details the de-identified demographics, experience levels, age, etc., of those who participated.

3. Setting and Apparatus

The following describes the location of the experiment, the equipment used for both the control and experimental conditions, and details of the setting supports the overall research design.

a. Collection Setting

As noted, all data was collected during normal working hours at MSTCPAC in San Diego, CA. The MSTCPAC facility provided access to both a classroom for the XR experimental condition and the mockup control condition. This arrangement was ideal in that bridge teams could walk from the mockup to the classroom hosting XR equipment in less than a minute. Environmental conditions inside the MSTCPAC building, e.g., temperature, humidity, ambient lighting, acoustics, room and hall colors, flooring, etc., were largely the same in both conditions.

Personnel participating in each condition remained at MSTCPAC without needing to leave during the time between the control condition and the experimental condition. Instructors, management staff, and technical support personnel remained consistent throughout the duration of each event and for the entirety of the experiment.

b. Experimental Condition Apparatus

The primary apparatus of this study consisted of XR, specifically the ONR funded, Navy Information Warfare Center Pacific (NIWCPAC) developed system referred to as VIBRaNT.

VIBRaNT is composed of 11 Meta Quest 3 HMDs, a Meta Quest Pro HMD, a Scenario Creation and Control Personal Computer (SCC PC), a VR Server, and a wireless router for connecting up to 12 individuals into the same VE. Meta Quest 3 headsets are configured with 512GB of storage, 2 color cameras and 3D spatial audio. The system does not require a common facility or room to operate in as it is networked over either a dedicated Wi-Fi enabled local area network (LAN) or an Internet cloud-hosted instance but for the purposes of the research, the system was tested in a single classroom over the dedicated VIBRaNT Wi-Fi LAN, as shown in Figure 16.

A Meta Quest Pro and Pupil Labs *Just Act Naturally* eye trackers were used to collect eye gaze and blink rate data, with the Meta Quest Pro having collected eye tracking data in the XR (experimental condition), and the Pupil Labs eye tracker having collected data in the mockup (control condition).

The VIBRaNT system used the SCC PC to provide researchers (and instructors) with the ability to design and configure scenario parameters for training and to modify scenario VE stimuli during runtime. The SCC PC also recorded Meta Quest Pro eye gaze heat map overlays to aid in determining area of interest (AOI) look-up, gaze vector time data, and blink rates. The SCC PC hardware consisted of a commercial off the shelf (COTS) laptop PC with a Windows 11 operating system, a COTS VE software suite, 3D VE models of San Diego, Norfolk, and the DDG-51 bridge configuration loaded on the COTS VR server and controlled by the SCC PC. Meta Quest 3/Meta Quest Pro headsets communicate to the SCC PC via a VIBRaNT dedicated "off network" LAN using a COTS Wi-Fi network router. The VIBRaNT apparatus is shown inside the pack-up kit (PUK) in Figure 17, but for specific technical details, see Appendix I.

Meta Quest 3 and Meta Quest Pro HMDs were full VR headsets that block out the outside world and present an entirely computer-generated VE into the participant's visual field. These HMDs also had high definition (HD) color-pass-through modes that allowed users to "see" the real world or actual physical environment where the testing was being performed. This pass-through mode also displayed VR boundary circles, system status, and Internet access as AR overlays in the real world. This allowed the facilitator and instructor to pull a student out of VR to aid system set up and highlight important learning points. It also served as an important safety feature, allowing participants to quickly exit the VE without needing to remove the HMD.



Figure 16. Virtual Bridge and Nautical Trainer extended reality testing session in Mariner Skills Training Center, Pacific, Classroom 215.



Figure 17. Virtual Bridge and Nautical Trainer extended reality pack-up kit in Mariner Skills Training Center, Pacific, Classroom 215.

The Meta Quest Pro enabled eye tracking via onboard cameras which were completely passive infra-red (IR) systems that had no impact on user experience. It also provided full color pass-through and even though older than the Meta Quest 3, presented scenarios with nearly identical levels of visual fidelity. OODs, being the most senior watchstander assigned to a bridge team, were selected for using the Meta Quest Pro and were required to sign a release to collect their gaze vector data.

Unity 3D VE-enabling software provided the virtual environment and synthetic experience needed to support the at sea navigation environment. VE details included: (1) simulation of current, wind, and other environmental factors, (2) simulation of own ship hydrodynamic forces, including engine and rudder, (3) simulation of other ships, including tugs, logistics ships for underway replenishment, and other private, commercial, and military ships in the nearby environment, (4) models of common ports, to include San Diego and Norfolk port approaches, (5) a DDG-51 class bridge configuration to include it's helm control systems, control interfaces, space layout, and audio environment, and (6) DDG-51 class shiphandling and maneuver characteristics.

c. Control Condition Apparatus

MSTCPAC NSST simulator systems, bridge mockups, and control and sensor interfaces were used in the control condition. These legacy standalone physical room simulators use video projection to mimic the sea environment, to include other vessels, sea conditions, and land masses. Physical controls and consoles are used for participant input and ship control. Based on the participants' input, the video simulation (projected onto the simulator windows) changes in accordance with programed simulation responses dictated by modeled ship hydrodynamics, environmental conditions, and bridge crew inputs, e.g., HELM, Lee HELM rudder, course, speed inputs, etc.

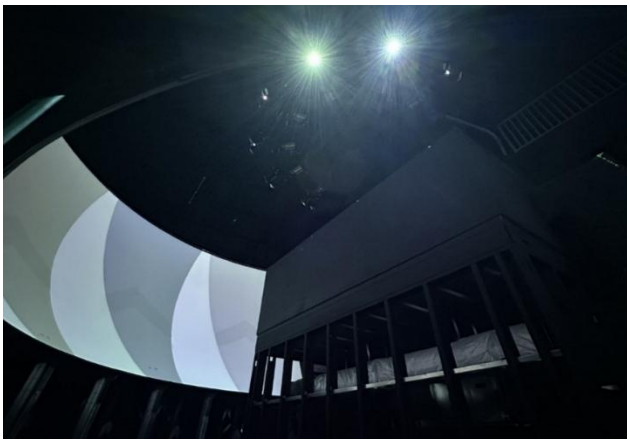


Figure 18. Photograph of Navigation, Seamanship, Shiphandling Trainer 5B from outside and below the bridge mockup windows.

VShips™ Simulation Suite of Hardware and Software powered the mockup environment of the NSST. This system, supported by up to 72 different VE computers and up to 27 projectors, provides instructor-controlled environmental visuals and bridge instrumentation suite of systems, e.g., VMS, RADAR, Automatic Identification System, etc. Figure 19 shows a subset of equipment, seven front facing projectors of the mockup system taken from above the bridge. These systems, combined with emulations of actual bridge hardware, comprised the NSST mockup.

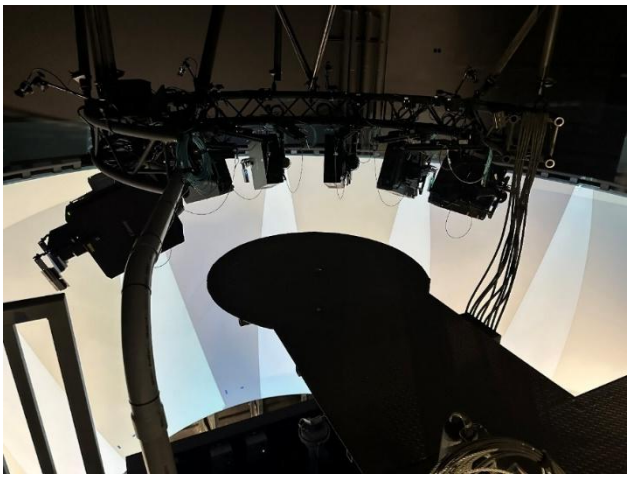


Figure 19. Photograph of Navigation, Seamanship, Shiphandling Trainer 5B portside video projectors taken from outside and above the mockup bridge.

During the control condition (NSST), OODs were provided with Pupil Labs *Just Act Naturally* eye tracking equipment, essentially a pair of glasses, relatively innocuous, to measure OOD response to the trainer by passively tracking eye gaze and blink rate while performing training in the NSST. In this way, eye tracking data was collected in the NSST and analyzed in comparison to the VIBRaNT Meta Quest Pro eye gaze data. Pupil Labs equipment (along with VIBRaNT's Meta Quest 3's, and Meta Quest Pro), was used in accordance with manufacturer's specifications and fact sheets.

4. Surveys

The following details the survey instruments used to collect data from the participants before an after beginning evaluation testing in either the NSST or XR.

a. Pre-Event Survey

(1) Pre-Screening Questionnaire (Appendix G)

Designed specifically for this research, the pre-screening questionnaire was administered to each participant prior to bridge team observations and asked (1) demographic information such as age, qualification level, and shiphandling experience, (2) familiarity with different kinds of technology and, (3) previously experienced health symptoms relevant to the XR equipment, e.g., epilepsy, seizures, etc. If participants noted medical issues in the questionnaire, or if they did not have the appropriate shiphandling experience, they were excluded. (Note: During data collection, no participant was excluded as all met the participation requirements of the study.) Demographic and technology familiarity data were retained for potential use in later post-hoc analysis.

Per the NIWCPAC VIBRaNT IRB Protocol, NIWCP.2024.0003-IR-EP7-A 19NOV24, all data was de-identified such that participants were assigned a code to ensure personally identifiable information (PII) would not be released, see Appendix J.

b. Post-Event Surveys

A battery of surveys was given to gage response to each condition (control and experiment) including: a simulator sickness questionnaire (for XR only), a NASA Task

Load Index (TLX) workload questionnaire, a System Usability Survey (SUS), watchstander self-assessment questionnaire, and an instructor watch team assessment questionnaire. All participants also answered an Exit Interview Questionnaire. Each questionnaire is described in the following sections:

(1) Simulator Sickness Questionnaire (Appendix H)

This questionnaire is a self-report instrument for participants to use if they are feeling simulator sickness and quantified any noted level of discomfort. Each symptom is given a score of 0–3 depending on severity. Based on a total questionnaire tallied score of 24 points or greater, the individual was not allowed to participate (Kennedy et al., 1993). (Note: None of the individuals sampled were rejected due to failing the simulator sickness questionnaire.)

(2) NASA Task Load Index Survey (Appendix K)

Originally designed for NASA astronauts, this questionnaire is a highly empirically validated self-report instrument for participants to record their subjective workload across six workload dimensions. This survey was selected to compare participant workload in both the control and experimental conditions and thus benchmark XR cognitive load to the mockup. The survey was carried out immediately after participating in both conditions.

(3) SUS (Appendix L)

This questionnaire is a self-report instrument that allows participants to report the degree of enjoyment and utility they felt from using the system, how easy it was to use, and whether they were willing to continue using it. This questionnaire enabled researchers to understand which aspects of the system may require additional work and aided the research effort to compare subjective responses to each condition. The same survey was collected immediately after participation in both the control and experimental conditions.

(4) Bridge Watchstander Self-Assessment Rubric (Appendix M)

Designed specifically for this research, this questionnaire was based on the learning rubrics used at the MSTCPAC for student responses to the NSST simulation environment.

This survey was collected from each participant immediately after participation in both the control and experimental conditions. In summary, participants were asked to rate their own performance and SA during the training. The survey is based on a Likert scale wherein the bridge team participant rates how much they agree or disagree with the survey statements provided.

(5) OOD Self-Assessment Rubric (Appendix N)

Designed for this research, this survey was created specifically for the bridge team OOD. The survey is based on the learning rubrics used at MSTCPAC for student OOD responses to the mockup (NSST) simulation environment. The survey used a Likert scale wherein the OOD assessed how much they agreed or disagreed with the statements provided. When available, MSTCPAC instructors also completed this survey evaluation of the OOD and both completed these immediately after participation in each condition, control and experimental.

(6) Instructor Assessment Rubric: Bridge Watch Team (Appendix O)

Designed for this research, this questionnaire was based on MSTCPAC training rubrics used during mockup (NSST) simulator events. This rubric was designed to be collected from instructors at the end of each control and experimental condition; however, instructor participation was inconsistent greatly limiting the sample collected. The instructors were asked to rate the team's performance based on a Likert scale to rate how much they agreed or disagreed with the statements provided. The intent was to collect survey results immediately after participation in the control and experimental conditions.

(7) Exit Interview Questions (Appendix P)

Designed specifically for this research, this questionnaire asked three open-ended, subjective questions related to the XR experience. The intent of these questions was to gather general impressions and probe for possible creative response not anticipated in the other surveys and possibly guide future research and development efforts.

The surveys collected from each participant were identical in both the control and experimental conditions. As noted, all participants completed their surveys immediately

after performing in the observed condition, mockup or XR. To review the details of each survey, see the Appendix for each survey described.

5. Observations

The following details activities that were used to capture eye tracking and video. All data collected has been archived for future analysis in subsequent studies.

a. Eye Tracking

Participants assigned to the watch standing role of OOD were asked to have their eyes tracked in each condition. In the experimental condition, eye tracking was done using the Meta Quest Pro eye tracking system, which recorded data in Meta Quest Pro memory, viewable on the VIBRaNT SCC PC. Data included AOI focus, duration of focus, blink rate, gaze vector, and gaze vector change rate. These same parameters were also collected in the mockup condition via the Pupil Labs system with an added parameter of pupil dilation.

(Note: *The Just Act Naturally* system was compatible with use of regular eyeglasses, even if slightly cumbersome to accommodate.)



Figure 20. Officer of the deck wearing Pupil Labs' eye tracking glasses.

b. Video Recording

Participant video data was also recorded but limited in the following ways: in the XR condition, video of the action taking place in VR was captured using the recording function of the Meta Quest 3 and the Meta Quest Pro. In this, the participants were *not* recorded, but rather their view of the VE was captured. This video captured what participants were looking at and their fellow watchstander VR avatar responses.

In the control condition, video data was recorded via the Pupil Labs system and was limited to what the OOD was looking at during performance in the mockup trainer.

6. Procedures

The following section reviews the detailed procedures used to observe the participants performance in the NSST and VIBRaNT XR respectively. During VIBRaNT XR, the research team also dynamically facilitated the execution of the VE scenarios via the VIBRaNT SCC PC.

a. Control Condition: Mockup (NSST)

As noted, research procedures were aligned with routine training obligations set by the MSTCPAC Scheduler and ship's COs. Bridge team participants and their training scenarios were selected and scheduled by the ship's Navigator in coordination with MSTCPAC a day or two prior to each training event. With the schedule in hand, CNSF notified the ship's COs and Executive Officers (XOs) via email of the XR experimentation request prior to their crew's arrival at MSTCPAC. The recruitment email, generated by the IRB Protocol, explained that participation in the research was completely voluntary, but all COs contacted provided their consent allowing ship's company to participate in the study as they desired. The research team was able to schedule a MSTCPAC classroom for the duration of the research, allowing equipment and survey materials to be set up in the same room for each observation. The classroom and XR system were able to accommodate as many as seven bridge team members in a single session, allowing each participant to interact with their fellow bridge team members (via the XR system's standalone SCC PC and wireless LAN) without the aid of the XR voice over internet protocol function.

Participants were generally observed in the morning after arriving at MSTCPAC, finding their instructor, and either attending an instructor lecture or going directly to the designated NSST mockup. (Note: There are several NSST mockups at MSTCPAC, three of which were used for this study: NSST 3, 4A, and 4B. NSST 4A and 4B were identical to each other, while NSST 3 was slightly smaller. For the purposes of this study, NSST 3 was considered to be functionally the same as NSST 4A and 4B.)

Those participants observed first in the control condition (mockup) were briefed on the project and told participation was completely voluntary and that they were able to withdraw at any time without negative repercussions to themselves or their fellow watchstanders. They were then asked to read the Privacy Act Statement (Appendix E), followed by reading and signing the ICD (Appendix F). Once this step was completed, the senior bridge team officer was asked which individuals were to be assigned to each of the different watch standing roles for the mockup training event, e.g., OOD, JOOD, CONN, etc. These watch standing assignments were noted for matching in the XR condition and participants were subsequently sent to their assigned NSST to prepare for their MSTCPAC-led instruction. During this preparation, the OOD was asked to don the Pupil Labs eye tracking glasses, which were then calibrated for recording at scenario start. OOD eye tracking was captured for approximately the first 30 minutes of the mockup training session. Apart from OOD use of eye tracking glasses, the mockup training session (control condition) was conducted as ordinarily done in the NSST.

After completing their assigned mockup training, participants were escorted to the XR IT classroom and asked to complete the mockup (control condition) post-event surveys including SUS, NASA TLX Survey, and Simulator Sickness Survey. Bridge Team members, e.g., OOD, JOOD, CONN, etc., were asked to complete the Watchstander Self-Assessment Survey, and the OOD was asked to complete an additional OOD self-assessment survey. The MSTCPAC Instructor was asked to complete the Bridge Team Training Instructor Assessment Rubric and an OOD Assessment Rubric. Each of these surveys were collected to baseline participant responses observed during the NSST simulation. Each of the individual survey results were identified via code only. Once

complete, results were collected and secured for comparison to the same survey responses to the XR (experimental) condition.

b. Experimental Condition: XR (VIBRaNT)

Having completed their control condition (NSST) surveys, bridge team participants were asked to complete the pre-screening questionnaire followed by a ten-minute introduction to the XR research goals, methods, and requirements, and were reminded participation was completely voluntary. (Note: none of the participants were screened as ineligible and all remained for the duration of the observation, with a few limited exceptions of bridge team members who, for various reasons, were needed back on the ship to manage emergent tasking.) An additional two-minute introductory video, followed by approximately ten minutes of introductory VIBRaNT XR orientation, controller training for interacting in the VR environment, and XR safety training were provided. The training scenario, having been determined for the mockup training session, was matched and initialized in the XR system SCC PC and set for problem start with the virtual ship's interaction controls set to allow participants to interact with the virtual systems, e.g., the ship's rudder and helm controls, VMS, RADAR, etc. Participants were then briefed on the chosen scenario, e.g., perform Underway Replenishment (UNREP) practice with UK Type 23 Frigate His Majesty's Ship (HMS) Hermes on Romeo CORPEN 090 at 13 Knots. After the briefing, the participants were asked to don their HMD's and "step into" the bridge VE.

In XR, Meta Quest 3 and Meta Quest Pro HMDs provide participants with MR experience. When donning the HMD, participants first observed the MSTCPAC classroom in full color video pass-through mode. This mode provided each participant with a video camera pass-through presentation of the classroom, their fellow bridge team members, the instructor, and all the other classroom contents at a resolution slightly higher than HD television standards (1080 Progressive or 1080 P). Researchers ensured participants had proper focus by checking their ability to read surveys just completed while in video pass-through mode and ensured other classroom objects, e.g., desks, chairs, lectern, etc., were clear and easy to see.

Because the researchers had previously set up the XR environment for observations, each HMD presented the participants with a blue circle painted on the classroom floor as an AR overlay. These blue circles are provided by the vendor to show the boundary between HMD modes of operation, from a pass-through' mode with AR presentation capability to a full VR capability where the real world view is occluded and replaced with a virtual world presentation, the virtual world in this case being the bridge VE. The researchers situated participant blue circles such that each bridge team watchstander might enter and participate in the bridge VE without bumping into other team members participating during the XR observation period. In addition, the orientation and placement of participant VE boundaries were aligned roughly with the orientations expected of bridge crew members standing watch on a USN Arleigh Burke Class Guided Missile Destroyer (DDG).

This "blue circle" alignment procedure was designed after VIBRaNT Pilot Study Number 2 showed that matching body location and orientation to the anticipated bridge VE environment enhanced participants' sense or presence of being aboard ship. All observations were thus performed such that participants stood (as they do on the actual bridge) for the duration of the observation and faced the same classroom wall which, in the bridge VE, was aligned to the front of the bridge or bow of the virtual ship. To prime the participants for this orientation, classroom desks were placed facing the right and left walls. Against the left wall, each movable desk was labeled with the appropriate bridge watch standing assignment, e.g., CONN, OOD, JOOD, CO, and NAV. In this way, the left wall served as the front of the bridge, where officers normally stand and tend to their respective equipment, e.g., Pelorus, bridge control units (BCUs), multi-function control repeaters (MFCRs), RADAR repeaters, VMS displays, etc.

Along the right wall, desks were rolled over facing the opposite direction and labeled with Instructor, HELM, LEE HELM, and XO respectively. This alignment placed participants in the VE towards the back of the bridge where the enlisted watchstanders normally stand and tend their equipment, e.g., Ship Control Console, Rudder, Throttle Controls, and Ship's Whistle. Once in the bridge VE, HELM and Lee HELM watchstanders were turned 180 degrees and the VE was reoriented to have them facing the

same direction as the other watchstanders. This 180-degree rotation roughly matched their body positions to the others by facing the front of the bridge of the virtual ship. Placing Instructors in the rear of the bridge gave them the ability to oversee the entire team's operation, but they tended to move around as needed when participating in the VE.

After stepping into the bridge VE, each watchstander was shown how to orient the VE presentation to their eye height and desired body orientation. Additionally, the research team showed them how to 1) move about the virtual bridge, 2) use their virtual binoculars, 3) use the virtual VMS, 4) use the virtual RADAR, 5) use the virtual radio, 6) activate the ship's virtual whistle, 7) turn the ship's virtual rudders and, 8) move the ship's virtual throttles. This orientation period typically lasted about 5 minutes, depending on how many participants were involved and how adept each participant was in learning how to perform in the VE. While this orientation was proceeding, researchers activated the HMD record function to capture the experience from the participant watchstander point of view to aid analysis of each performance session.

After the OOD confirmed the watch team was ready to commence the problem, i.e. UNREP, a 30-minute observation period was commenced by starting the SCC PC controlled simulation. This started all movement: own ship and other vessels, environmental effects, wind and current, bridge VMS system tracking in the VE, etc. It was then up to the OOD and the watch team to respond to different activities happening (stimuli) in the simulation, e.g., a vessel with a relative movement producing a constant bearing, decreasing range situation threatening the bridge team's ship with collision. In the case of the UNREP scenario, the XR session mirrored the mockup training session by providing the same situation and training goals in both presentations. These goals consisted of, (1) rendezvousing with the replenishment ship, (2) maintaining position in waiting station "L6," (3) performing the steps needed to make the approach alongside the replenishment ship, (4) maintaining proper distance and relative bearing while alongside the replenishment ship while conducting UNREP, (5) tracking, communicating with, and reporting to the ship's CO other vessels in the area and, in particular, those that potentially threatened UNREP activities, (6) warning off other vessels, (7) performing procedural response of performing "CORPEN N" to change course while alongside the replenishment

ship, and (8) performing "EMERGENCY BREAK-AWAY" procedures when risk of collision was eminent.

During the scenario, each participating watchstander could observe the spatial-temporal cues provided and hence (1) giving and receiving orders, (2) observing the feedback from the vantage point of their normal bridge watch station assigned positions, (3) responding to feedback with new orders, and (4) repeating this process as the scenario proceeded in real-time. Furthermore, this practice also allowed each participant to learn how fellow bridge team members reacted under the pressure of the given scenario and all the subtle social cues happening between each watchstander in response to the scenario. Because the kinematics of the simulation mirrored those of the real world, the interplay between the human subject participants and the informational presentation of the XR created a synthetic experience akin to the mockup experience allowing for subsequent within subjects comparison of both informational accesses, mockup and XR. The observation period was run for approximately 30 minutes and stopped to allow participants time to complete their XR response surveys.

After completing the XR observation period, participants were seated and asked to complete the XR (experimental condition) post-event surveys, including the SUS, NASA TLX Survey, and Simulator Sickness Survey. Bridge Team members, e.g., OOD, JOOD, CONN, etc., were asked to complete the Watchstander Self-Assessment Survey, and the OOD was asked to complete an additional OOD self-assessment survey. The MSTCPAC Instructor was asked to complete the Bridge Team Training Instructor Assessment Rubric and an OOD Assessment Rubric. All participants were asked to complete the Exit Interview Questionnaire (for XR only), and these questionnaires were included in the packet of participant responses. Each of these surveys captured participant responses to the XR system, where individual survey results were identified via the same anonymized code used for the NSST system surveys, but signified via "V" to indicate the responses were based on VIBRaNT XR, not the NSST. Once complete, results were collected and secured for comparison to the same survey responses to the mockup (control) condition.

c. Counterbalancing Observations

To counterbalance this sequence of testing, participation in the mockup (NSST) first, followed by XR (VIBRaNT) second, half of the participants experienced the reversed sequence, participating in XR first, followed by mockup second. The primary difference between sequences, beyond the sequencing itself, was the minor difference in how the Pre-Screen Questionnaire was collected. After meeting with their MSTCPAC instructors, and receiving a lecture or other guidance, *all* participants reported to the XR classroom, read the privacy statement, read and signed the ICD, complete the pre-screening questionnaire, and complete the simulator sickness questionnaire. With these forms completed, participants were permitted to participate in XR observations, after which the "post-event" surveys were completed and collected, again as noted above. With the XR observation period complete, participants were escorted to the appropriate mockup NSST where the OOD donned the Pupil Labs' eye tracking glasses and the first 30 minutes of OOD participation in the mockup scenario were recorded. After the bridge team completed the mockup training session, they returned to the XR classroom to complete the post-event responses to the mockup trainer as noted above. Once complete, the surveys were collected, and the bridge team was dismissed to resume duties aboard their ship.

All data collected was anonymized via an assigned participant identifier code, any related identity PII was kept on an encrypted hard drive locked in a cabinet at Naval Information Warfare Center (NIWC) Pacific.

7. Institutional Review Board Considerations

Prior to seeking access to data sources, NIWC Pacific's XR Project VIBRaNT obtained a favorable review of its experimental design protocol. All data collected by Project VIBRaNT was deemed compliant as specified in NIWC Pacific's approved IRB Protocol NIWCP.2024.0003-IR-EP7-A dated November 19, 2024. Per this protocol, all data was de-identified and biospecimens were *not* collected or used for any part of this study, see Appendix J.

This dissertation is a separate, complementary effort that leveraged secondary information collected during the VIBRaNT effort. Documents including a pre-screening

questionnaire, simulator sickness questionnaire, NASA Task Load Index Workload Questionnaire, SUS, Bridge Watchstander Self-Assessment Rubric, OOD Self-Assessment Rubric, Instructor Assessment Rubric-Bridge Watchstander, Instructor Assessment Rubric-OOD, and an Exit Interview Questionnaire. Eye gaze vector data, along with *Time to Task (Recognition)* and *Time to Task (Completion),* were also captured for evaluation. PII data was *not* included in the secondary information and not used for this dissertation. All demographic data was de-identified.

These datasets were collected while scrupulously following the VIBRaNT IRB Protocol and conceivably could be used in future study efforts as well.

C. DATA ANALYSIS

The following section details post experiment data preparation, analysis design, and the statistical tests employed to produce comparison results. It describes the overall approach, the software tools used, and some general data findings. For detailed findings and analysis, refer to Chapter IV. Experimental Results and Analysis.

1. Data Preparation

Post-Event questionnaires consisted of paper packets presented to each participant immediately after completing each observation period. Filling out the packets consisted of participants marking a box or circling a number on each survey, e.g., 1 to 7, under each statement posed to reflect how much they either agreed or disagreed with the statement. As noted above, each packet consisted of multiple surveys, simulator sickness questionnaire, NASA TLX Survey, SUS, and a Watchstander Self-Assessment Survey. The OOD completed an additional OOD self-assessment survey, and MSTCPAC Instructors completed a Bridge Team Training Instructor Assessment Rubric and the OOD Assessment Rubric. Additionally, all participants were asked to complete an XR Exit Interview Questionnaire.

Packet results were transferred by the research team to Microsoft Excel and organized (by questionnaire) into tables with questions across the top (columns), and participant numbers down the left (rows). Using the participant number as the common

key, each survey was combined with the pre-screening questionnaire (demographic data), simulator sickness questionnaire, Watchstander Position Assigned, and a notation as to which condition (control or experimental) the participant was observed in first. Additionally, all response data was coded with higher scores equating to positive agreement. This required reverse coding responses to statements worded in the negative (to avoid positive-biased responses) to positive statements with a corresponding positive value. Data columns were then arranged side-by-side for a paired data comparison of responses to each condition, Control (NSST) and Experimental (XR). In cases where data was missing, e.g., a participant provided a response in one condition (NSST), but not the other (XR), it was noted and the row for that participant was deleted.

Having the data prepared in this way enabled quick filtering in Excel to pull data sets of interest, e.g., NASA TLX Questionnaire responses from "All Participants in the NSST control condition," or "All Participants in the XR experimental condition," or "OOD Only in the NSST Condition after participating in XR," etc. Filtered data were subsequently copied and pasted into Real Options Valuation (ROV) Business Statistics (BizStats®) software (Mun, 2019) for statistical analysis.

2. Statistical Analysis Design

ROV BizStats software was used to allow the researcher to load variables in columns, not unlike Excel, and perform analysis using a wide variety of statistical tests. This provided flexibility to analyze relationships between independent and dependent variables. Here the independent variables were the control and experimental conditions, and survey responses serve as dependent variables. Survey responses were operationalized as correlates of knowledge flow facilitation, e.g., cognitive load as measured via the NASA TLX Survey, system usability as measured via the SUS, and bridge watch team task completion as measured in the various assessment rubrics. By comparing responses in both conditions, a characterization of the experimental condition (XR) to facilitate knowledge flow for bridge training was made relative to the control condition. In this, there was no absolute measure of knowledge flow facilitation, only a comparison relative to the USN's currently employed mockup trainer (NSST). Statistical comparison showed "no significant

difference" in many areas, which is significant, as the XR trainer has added affordances of being much lower in cost, highly portable, and able to provide training from distributed locations, something not possible with a mockup.

To evaluate and characterize XR's capacity to facilitate knowledge flow, hypotheses H_1 to H_4 were evaluated statistically in a three-phase approach. Phase one consisted of analyzing the responses from each survey to characterize the data and determine suitability for parametric or non-parametric statistical comparisons. Phase two consisted of evaluating data reliability, specifically examining the degree to which participant responses accorded with one another and how well the response data held together across the multivariate factors of each questionnaire. Phase three compared control and experimental condition survey responses using multivariate and univariate tests, determined the statistical power of each analysis, performed principal component analyses along with post-hoc tests.

3. Statistical Tests Employed

The specific tests utilized in the ROV BizStats software (Mun, 2014) are listed and briefly described in terms of what was analyzed (Table 4).

Table 4. Summary of survey response statistical analyses.

Survey Response Statistical Tests	Dependent Variable Analyses	
	Test Purpose	Survey Analyzed
Statistical Moments	Data Description	1,2,3,4,5
Distributional Fit Analysis	Data Description	1,2,3,4,5
Shapiro-Wilk, D'Agostino-Pearson (Normality)	Para or Non-Para Tests	1,2,3,4,5
Kendall's W (Inter-Rater Reliability)	Respondent Concordance	1,3,4,5
Guttman's Lambda (Internal Consistency)	Systemic Stability	1,2,3,4,5
Hotelling's T^2 (Multivariate Analyses)	Pooled Comparison	1,2, part 3,4,5
Wilcoxon signed-rank test (Bivariate Analyses)	Item Comparison	1,2,3,4,5
Cohen's *d* (Effect Size)	Difference Magnitude	1,2,3,4,5
Bonferroni Correction (WSRT Assessment)	Pooled Comparison Test	1,2,3,4,5
Mann-Whitney Test (order effects Comparison)	Order Impact-Control	1,2,3,4,5
PCA with Varimax Rotation	Factor Analysis	1,2,3,4,5
Statistical Power Analysis (Corner Case Analysis)	Sample/Test Assessment	1,2,3,4,5

a. Statistical Analysis: Data Description

ROV BizStats descriptive statistics tests were used to evaluate several different statistics for each set of participant responses, i.e., NASA TLX control condition (NSST) *All Participants*. This included the arithmetic mean, median, upper and lower confidence intervals (CI), standard deviation, skew, kurtosis, etc. Results data were then explored using a distributional fitting analysis of participant response data by examining the best ten distributional model fits for each test of the data sets presented. For example, a distributional fit test of *All Participant* responses to the NASA TLX control condition (NSST) factor of Mental Demand. To perform this specific test, ROV BizStats determined the first best fit was a PERT distribution using an Akaike model, the second best Akaike model fit was triangular, etc. After exploring the best ten Akaike model fits, the next best set of model fits, e.g., Anderson–Darling, were determined, and so forth for each of the NASA TLX subfactors, e.g., *Mental Demand (Q1)* for each of the groups, e.g., *All Participants.* After completing the distributional fitting analysis, each data set was evaluated for normality using Shapiro–Wilk-Royston tests and in some cases D'Agostino–Pearson tests. While a few of the data sets, i.e., NASA TLX control condition (NSST) *Temporal Demand*, were found to be plausibly normal, many were not. **Based on these findings, the decision was made to proceed using non-parametric tests**.

b. Statistical Analysis: Data Reliability

Data reliability was evaluated using internal consistency reliability (ICR) and inter-rater reliability (IRR) tests. ROV BizStats Guttman's Lambda and Split Half Model was employed to evaluate ICR, and the non-parametric Kendall's W (With Ties) test evaluated IRR. Although some of the data was found to be plausibly normal, the inter-class correlation test was not selected as the survey data was neither interval nor ratio, but ordinal based on subjective participant responses with no way to judge interval distances from person to person. Cronbach's Alpha was also rejected as the data was not dichotomous. These tests determined the data to be sufficiently reliable to warrant statistical comparison analysis.

c. Statistical Analysis: Data Comparison

To compare each condition, an ROV BizStats™ Hotelling T-Square: 2 Variable Dependent Pair with Related Measures test was run to evaluate the multivariate pooled statistical difference (across all question responses) of each survey, e.g., NASA TLX, SUS, Bridge Watch Team Self-Assessment, etc. Tests compared *All Participant* responses to each survey as well as each subgroup with n > 25, e.g., *OOD Only, JOOD Only, CONN Only*, etc. Following this, a Non-Parametric Wilcoxon signed-rank test (two variables) was employed to compare the control condition (NSST) responses to each individual question to those of the experimental condition (XR). As with the Hotelling T-Square comparison, tests were first run for *All Participants* and then for each subgroup, e.g., *OOD Only, JOOD Only, CONN Only*, etc., for all surveys including NASA TLX Workload, SUS, Bridge Watch Team Self-Assessment, OOD Assessment, Instructor Bridge Watch Team and OOD Assessments. A Bonferroni Test (Two Variables with Repetition) was then run on the same response data sets to correct potential p-value errors (produced by the independently run WSRT tests), along with a simultaneous confidence interval test. Following this, PCA, statistical power analysis, to include effect size (Cohen's Delta), were also run.

THIS PAGE INTENTIONALY LEFT BLANK

IV. EXPERIMENTAL RESULTS AND ANALYSES

The following chapter presents the analysis of survey responses characterizing XR training facilitation in comparison to the NSST. Results are presented in order of the surveys analyzed including the NASA TLX analyses, SUS analyses, Bridge Watchstander Self-Assessment Survey, OOD self-assessment survey, and Instructor Assessment Survey. Statistical analyses were determined using BizStats® Software (Mun, 2014).

A. NATIONAL AERONAUTICS AND SPACE ADMINISTRATION TASK LOAD INDEX SURVEY RESULTS

Participants were asked to rate their workload perception in response to six NASA TLX subscale questions on a scale from 0 to 10 in increments of 0.5 (Table 5).

Table 5. National Aeronautics and Space Administration Task Load Index subscale question key.

NASA TLX Subscale	NASA TLX Subscale Question (Full Text)
Mental Demand (Q1)	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the mission easy or demanding, simple or complex, exacting or forgiving?
Physical Demand (Q2)	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the mission easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand (Q3)	How much time pressure did you feel due to the rate or pace at which the mission occurred? Was the pace slow and leisurely or rapid and frantic?
Performance (Q4)	How successful do you think you were in accomplishing the goals of the mission? How satisfied were you with your performance in accomplishing these goals?
Effort (Q5)	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration (Q6)	How discouraged, stressed, irritated, and annoyed versus gratified, relaxed, content, and complacent did you feel during your mission?

NASA TLX scale increments and questions were derived from the literature (Hart, 2006; Hart & Staveland, 1988), with higher scores, apart from *Performance (Q4),* indicating greater workload for the participant to complete their given mission. For this research, "mission" equated to individual bridge watchstander performance needed to

successfully navigate the ship through the given scenario. To make meaningful comparisons, the ship's mission was kept identical such that NASA TLX responses differentiated the experimental condition (XR) experience from the benchmark of the control condition (NSST). As an example, if participant 22–119 gave a *Mental Demand (Q1)* score of 5.5 in the NSST, and 7.0 in XR, then the inference made (for participant 22–119) was that XR is more demanding mentally than the NSST. For *Performance (Q4),* if participant 22–119 gave a *Performance* score of 6.0 in the NSST, and 8.5 in XR, the inference made was that participant 22–119 perceived their performance in XR as being higher than in the NSST. After collection and paring of all participants' results in Excel, the total number of paired results was N = 129. As noted, this was a subset of the total number of possible pairings as some individuals, for a variety of reasons, failed to complete both the NSST and XR post-event NASA TLX surveys.

NASA TLX statistical analyses and results are organized as follows: Findings Summary, Descriptive Statistics, Distributional Fit, and Normality, Reliability and Agreement, NSST vs. XR Comparisons, order effects, PCA, Power Analysis, and the Summary of Key Findings.

1. Findings Summary

The combined analytic strategy provided a robust, multi-angular finding that respondents rated **only two out of six NASA TLX subfactors as significantly different ($p<0.05$) in XR compared to the NSST.** These two subfactors were *Performance (Q4)* and *Frustration (Q6).* Based on exit interviews, variance in these subfactors may have been attributable to XR novelty, as none of the participants had used XR for bridge training prior to the study. Effect sizes were consistent across full sample and subgroup analyses, though smaller subgroups lacked power to reliably detect weaker effects. PCA suggested differences in XR responses were attributable to an affective/emotive component, rather than cognitive. **These findings suggest XR facilitation of training nearly met the NSST benchmark, but an XR novelty penalty may have held it back slightly as none of the participants had ever used the VIBRaNT XR system.** Order effects, the impact of experiencing a scenario in XR right before training in the NSST, were insignificant,

suggesting further study is needed to characterize how XR impacts subsequent NASA TLX subfactors in NSST training sessions.

2. Descriptive Statistics, Distributional Fit, and Normality

The following section details analysis of NASA TLX results in terms of their descriptive statistics, distributional fit, and normality.

a. Descriptive Statistics

Survey response data was analyzed to determine four primary statistical moments: distribution center, spread, skew, and kurtosis. Findings are tallied below for both the control condition (NSST) and experimental condition (XR) in Table 6.

Table 6. National Aeronautics and Space Administration Task Load Index survey, descriptive statistics of Navigation, Seamanship, Shiphandling Trainer control and extended reality experimental conditions for *All Participants* (N = 129).

NASA TLX Subscale	Condition	Mean	95% CI Lower–Upper	Median	SD	Skew	Kurt
Mental Demand	NSST	6.04	5.66–6.42	6.00	2.18	-0.40	-0.43
Mental Demand	XR	5.87	5.50–6.24	6.00	2.09	-0.37	-0.27
Physical demand	NSST	2.50	2.13–2.87	2.00	2.08	1.04	0.53
Physical Demand	XR	2.56	2.22–2.91	2.00	1.95	0.95	0.29
Temporal Demand	NSST	4.76	4.37–5.15	5.00	2.20	0.13	-0.58
Temporal Demand	XR	4.83	4.45–5.22	5.00	2.18	-0.02	-0.38
Performance	NSST	7.47	7.13–7.80	8.00	1.91	-0.79	-0.05
Performance	XR	5.62	5.17–6.08	5.50	2.57	-0.16	-0.88
Effort	NSST	5.33	4.92–5.74	5.50	2.32	-0.08	-0.65
Effort	XR	4.95	4.55–5.36	5.50	2.31	-0.24	-0.79
Frustration	NSST	3.33	2.89–3.76	2.50	2.49	0.67	-0.55
Frustration	XR	4.29	3.81–4.76	4.50	2.70	0.18	-0.94

b. Distributional Fit Analysis

Across all subscales, the best distributional fit was rarely Normal. Instead, skewed or bounded distributions (PERT, Triangular, Logistic, Beta, Lognormal, Pearson families, etc.) **dominated in a pattern consistent with task-anchored ratings where practical minima and maxima often piled up in an asymmetric skew toward an anchor. This analysis supported the use of non-parametric statistical tests** (Table 7).

Table 7. National Aeronautics and Space Administration Task Load Index survey, distributional fit analysis of Navigation, Seamanship, Shiphandling Trainer and extended reality control and experimental conditions for *All Participants* (N = 129).

NASA TLX Subscale	Condition	Best Fits (by criteria)	Normal Fit (overall)
Mental Demand	NSST	PERT; Triangular; Logistic (Normal sometimes under A-D)	No
Mental Demand	XR	Beta-4; Logistic; PERT; Double Log	No
Physical demand	NSST	Lognormal; Gumbel Max; Beta/ Erlang	No (strong skew)
Physical Demand	XR	Pearson V; Erlang; Lognormal; Rayleigh	No (strong skew)
Temporal Demand	NSST	PERT; Cosine; Beta; Triangular (A-D Normal once)	Borderline
Temporal Demand	XR	Triangular; PERT; Beta-4 (Normal appears mid-rank)	Borderline
Performance	NSST	Beta; PERT; Gumbel-Min; Laplace	No
Performance	XR	PERT; Triangular; Cosine; Beta-4	No
Effort	NSST	Beta; PERT; Triangular (Normal occasionally)	No
Effort	XR	Triangular; Cosine; Beta-4 (A-D Triangular)	No
Frustration	NSST	Lognormal; Erlang; Pearson; Gamma	No (strong skew)
Frustration	XR	Pearson VI; Gamma; Arcsine; Weibull	No

Note. NSST = control condition; XR = experimental condition. Results reflect model fits determined by best distributional alignment criteria (A-D = Anderson-Darling).

c. Normality Tests

Shapiro–Wilk (SW) and D'Agostino–Pearson (DAP) tests were used to confirm that NASA TLX response data distributions were non-normal. Standard statistical conventions were followed for indicating probabilities where $p<0.10$ reflected a probability of α (α = cut-off value) being 1 in 10, a single Asterix is added, i.e., $p<0.10$*, for α of 1 in 20, $p<0.05$ a double Asterix is added and the value is bolded, i.e., p<**0.05****, and finally for an α of 1 in 100, $p<0.001$ a triple Asterix is added and the value is bolded, i.e., p<**0.001***** **Normality test findings of the NASA TLX response data supported the use of non-parametric statistical comparison testing** (Table 8).

Table 8. National Aeronautics and Space Administration Task Load Index survey, normality checks, Shapiro–Wilk and D'Agostino–Pearson of Navigation, Seamanship, Shiphandling Trainer and extended reality control and experimental conditions for *All Participants* (N = 129).

NASA TLX Subscale	Condition	SW p-value	DAP p-value	Normal? (overall)
Mental Demand	NSST	**0.0089****	0.1035	Mixed → Not overall
Mental Demand	XR	**0.0282****	0.1813	Mixed → Not overall
Physical demand	NSST	**0.0001*****	**0.0000*****	Not normal
Physical Demand	XR	**0.0001*****	**0.0000*****	Not normal
Temporal Demand	NSST	0.1089	0.3214	Normal
Temporal Demand	XR	0.1370	0.6599	Normal
Performance	NSST	**0.0000*****	**0.0010*****	Not normal
Performance	XR	**0.0025****	0.0868	Mixed → Not overall
Effort	NSST	0.0553	0.2858	Normal
Effort	XR	**0.0020****	0.0936	Mixed → Not overall

Note: "SW" = Shapiro–Wilk; "DAP" = D'Agostino–Pearson. Only "Temporal Demand" consistently met assumptions of normality in both conditions. Other subscales showed mixed results, but were all treated as non-normal. These findings supported "Non-parametric" testing.

3. Data Reliability and Agreement

Analysis of IRR and ICR across the NASA TLX subscales **revealed data agreement and consistency required for subsequent analysis.** Participants rated NASA TLX workload subscales consistently within groups, with significant concordance across raters. Internal consistency was moderate, with some variation spread across subgroups, e.g., the *CONN Only* subgroup, and across each condition, NSST or XR. Taken together, the **findings supported internal validity of NASA TLX as a workload assessment tool in both XR and traditional NSST contexts**. Findings also suggested subtle differences in conditions, with NSST yielding more uniform agreement. This uniformity may point to novelty of effects of XR, where first time reactions to the XR VE were less-uniform across participants. The bridge crews sampled for the research had limited to no VR experience, but did have extensive experience interacting with the NSST, which likely impacted expectations with the NSST condition meeting what was expected. XR results, on the other hand, suggest uneven expectations of performing bridge training in an XR VE.

a. Kendall's W (IRR–Concordance)

Across all participant groups (All, OOD, JOOD, etc.), Kendall's W values were statistically significant, indicating **participants had a reliable level of concordance**

(Table 6). W values tended to be higher in the NSST compared to XR, suggesting slightly greater agreement in workload assessments under the traditional training environment. The strongest agreement was observed in JOOD (NSST, W = 6.79), while lower—but still significant—agreement was observed in HELM/LEE HELM (XR, W = 3.20). Of note, all test p-values were p < **0.001***** (Table 9).

Table 9. National Aeronautics and Space Administration Task Load Index survey, Kendall's W (with ties) of Navigation, Seamanship, Shiphandling Trainer and extended reality control and experimental conditions for *All Participants* and *Subgroups*.

Subgroup	Condition	N (Reviewers)	Kendall's W	p-value
All Participants	NSST	129	5.272	< 0.001***
All Participants	XR	129	3.99	< 0.001***
OOD	NSST	26	5.082	< 0.001***
OOD	XR	26	4.579	< 0.001***
JOOD	NSST	25	6.788	< 0.001***
JOOD	XR	25	5.442	< 0.001***
CONN	NSST	26	6.183	< 0.001***
CONN	XR	26	3.709	< 0.001***
NAV	NSST	16	5.769	< 0.001***
NAV	XR	16	4.55	< 0.001***
HELM/LEE HELM	NSST	26	4.311	< 0.001***
HELM/LEE HELM	XR	26	3.196	< 0.001***

b. Guttman's Lambda (Internal Consistency)

Lambda values fell in the 0.53–0.71 range, reflecting that **participant responses had moderate internal consistency across subscales.** As might be expected with subjective workload measures, reliability was lower than classical benchmarks (e.g., $\lambda > .70$) but remained largely within an acceptable range. Split-half and odd-even correlations were positive across all groups, though variable, indicating **no evidence of systematic instability in subscale structure.** NSST conditions typically produced slightly higher Lambda values than XR with the highest reliability of $\lambda = 0.71$ for the JOOD in the NSST (Table 10).

Table 10. National Aeronautics and Space Administration Task Load Index survey, Guttman's Lambda of Navigation, Seamanship, Shiphandling Trainer and extended reality control and experimental conditions for *All Participants* and *Subgroups*.

Subgroup	Condition	Lambda (λ)	Split-Half (SB)	Odd-Even (SB)
All Participants	NSST	0.615	0.709	0.655
All Participants	XR	0.605	0.730	0.645
OOD	NSST	0.591	0.667	0.605
OOD	XR	0.574	0.666	0.627
JOOD	NSST	0.711	0.753	0.768
JOOD	XR	0.532	0.738	0.585
CONN	NSST	0.606	0.898	0.614
CONN	XR	0.602	0.819	0.627
NAV	NSST	0.492	0.419	0.640
NAV	XR	0.681	0.661	0.699
HELM/LEE HELM	NSST	0.571	0.586	0.692
HELM/LEE HELM	XR	0.641	0.740	0.674

4. Statistical Comparisons: Navigation, Seamanship, Shiphandling Trainer versus Extended Reality

Multivariate Hotelling T^2 tests, bivariate WSRTs, Cohen's *d* effect size, and Bonferroni tests were used to compare XR and NSST responses. Hotelling's T^2 pooled results showed a strong overall difference for *All Participants, JOOD Only,* and *CONN Only* subgroups. *OOD Only* and *NAV Only* showed a difference trend, but *HELM/LEE HELM Only* did not show a multivariate difference.

WSRT largely showed similarity in NSST and XR condition responses but did pinpoint a consistent difference in *Performance (Q4)* (across all groups), *Frustration (Q6)* for *All Participants* and *Effort (Q5)* for *CONN Only*. Cohen's *d* analysis highlighted *Performance (Q4)* effects as larger than others, with comparatively moderate effect sizes in subscales *Effort (Q5)* and *Frustration (Q6)*. **Bonferroni testing confirmed a significant difference in *Performance (Q4)* in just two groups, *All Participants* and *JOOD Only*** (Table 11).

Table 11. National Aeronautics and Space Administration Task Load Index summary of comparisons of Navigation, Seamanship, Shiphandling Trainer versus extended reality for *All Participants* and *Subgroups*.

Subgroup	Hotelling's T^2 (omnibus)	Wilcoxon Significant Subscales ($p < 0.05$)	Cohen's *d* (effect size)	Bonferroni Correction
All Participants (N=129)	T^2 = 68.75 F = 11.01 p < **0.001*****	Q4 p < **0.001***** Q6 p = **0.005****	Q4 *d* = 0.81 Q6 *d* = 0.37	Q4 High, All others within CI
OOD (n=26)	T^2 = 17.64 F = 2.35 p = 0.070	Q4 p = **0.010****	Q4 *d* = 0.83	All within CI
JOOD (n=25)	T^2 = 35.74 F = 4.72 p = **0.004****	Q4 p = **0.001*****	Q4 *d* = 1.21	Q4 High, All others within CI
CONN (n=26)	T^2 = 20.60 F = 2.75 p = **0.041****	Q4 p = **0.043** Q5 p = **0.034****	Q4 *d* = 0.63 Q5 *d* = 0.62	All within CI
NAV (n=16)	T^2 = 29.17 F = 3.24 p = 0.050	Q4 p = **0.036****	Q4 *d* = 0.79	All within CI
HELM/LEE HELM (n=26)	T^2 = 13.68 F = 1.82 P = 0.145	Q4 p = **0.010****	Q4 *d* = 0.75	All within CI

5. Order Effects

Research hypothesis H_4 sought to determine the training impacts of using XR before attending bridge training in the NSST. To characterize hypothesized XR impacts, NSST-TLX survey responses taken after participants performed XR training (labeled "NSST 2") were compared responses taken by those who went directly into NSST training first (labeled "NSST 1"). **Comparison analysis showed that order effects did not bias NSST responses in either a positive or negative direction. Participants' prior exposure to XR did not significantly alter their responses when completing NSST afterwards.**

Median differences (from Mann–Whitney tests) hovered around zero, and effect sizes (Cohen's *d*) remained below the conventional threshold for a small effect (d = 0.2). Given these results, evaluating order effects would likely require a greater sample in follow-on tests to properly characterize (Table 12).

Table 12. National Aeronautics and Space Administration Task Load Index of order effects analysis after exposure to the extended reality condition of Navigation, Seamanship, Shiphandling Trainer for *All Participants* (N = 129).

NASA TLX Subscale	NSST 2 (XR 1st?)	Med	Mann-Whitney p-value	Cohen's *d*	95% CI (Lower–Upper)
Mental Demand	Yes	6.5	p = 0.867	*d* = 0.05	-0.66 0.85
Mental Demand	No	6.0			
Physical demand	Yes	2.0	p = 0.518	*d* = 0.20	-0.33–1.16
Physical demand	No	2.0			
Temporal Demand	Yes	4.5	p = 0.452	*d* = 0.11	-1.04–0.54
Temporal Demand	No	5.0			
Performance	Yes	8.0	p = 0.622	*d* = 0.09	-0.86–0.51
Performance	No	8.0			
Effort	Yes	6.0	p = 0.921	*d* = 0.08	-0.64–1.01
Effort	No	5.5			
Frustration	Yes	2.5	p = 0.573	*d* = 0.07	-1.06–0.72
Frustration	No	3.0			

Note: "XR 1st?" column indicates whether participants completed XR training before going to the NSST. Yes = XR-first condition; No = XR not completed before NSST condition.

6. Principal Component Analysis

The PCA with Varimax rotation revealed two consistent components across both NSST and XR conditions: 1) a cognitive workload component and 2) an Affective/Performance component. The cognitive workload component was dominated by NASA TLX responses in *Mental Demand, Temporal Demand, Effort,* and, in some cases, *Physical Demand.* This component captured the mental and resource-related aspects of the NASA TLX workload responses. The *Affective and Performance* component was defined primarily by *Performance* responses and *Frustration*, with some cross-loadings from the other subscales. This *Affective and Performance* component reflects that participant judged their effectiveness and emotional response to task demands and performance ratings as better in the NSST, lower in XR, and may align with an XR novelty deficit due to first time use.

Results found are consistent with prior literature examining NASA TLX empirical validity (Braarud, 2021; Hart & Staveland, 1988). NASA TLX responses have typically been shown to cluster into broader latent factors, reflecting cognitive workload (e.g., *Mental Demand, Temporal Demand, Effort*) and affective or performance related dimensions (Aksu, 2025). Current findings reinforce this pattern, suggesting that observed

differences between NSST and XR training conditions reflect meaningful shifts in affective and performance related perceptions rather than measurement artifacts.

Eigenvalues greater than one show that the first two components explain approximately 65% of the variance in the NSST and 67% in XR. The similarity of this factor structure across both conditions suggests that differences observed in workload ratings are not due to measurement artifacts but rather represent genuine differences in how participants experienced the training systems. Minor cross-loadings (e.g., frustration overlapping with demand variables) in the XR analysis suggest slightly higher emotional salience when participants engaged with the immersive XR environment. The experimental condition components, as mentioned, were similar, with the notable exception of *Mental Demand (Q1).* This too is likely a reflection of the novelty effect of participants using XR (VIBRaNT) for the first time (Tables 13 and 14).

Table 13. National Aeronautics and Space Administration Task Load Index principal component analysis of Navigation, Seamanship, Shiphandling Trainer condition for *All Participants*.

PCA Factor	Eigenvalue	Proportion of Variance	Cumulative Proportion
Factor 1	**2.66**	**44.3%**	**44.3%**
Factor 2	**1.24**	**20.7%**	**65.0%**
Factor 3	0.71	11.8%	76.8%
Factor 4	0.64	10.7%	87.5%
Factor 5	0.40	6.0%	93.4%
Factor 6	0.36	6.6%	100.0%

Table 14. National Aeronautics and Space Administration Task Load Index principal component analysis of the extended reality condition for *All Participants*.

PCA Factor	Eigenvalue	Proportion of Variance	Cumulative Proportion
Factor 1	**2.80**	**46.7%**	**46.7%**
Factor 2	**1.21**	**20.1%**	**66.9%**
Factor 3	0.82	8.2%	75.2%
Factor 4	0.46	7.7%	82.9%
Factor 5	0.41	6.9%	89.8%
Factor 6	0.30	10.2%	100.0%

PCA rotated factor loadings highlighted which NASA TLX subscales aligned most strongly with the cognitive workload and affective and performance components. Findings

replicated typical NASA TLX two-factor structure reported in the literature of cognitive load and emotional response (Aksu, 2025; Braarud, 2021).

For the NSST control condition, responses to NASA TLX Subscale questions for *Mental* and *Physical Demand*, along with effort, clustered in the Cognitive Workload component, where *Temporal Demand*, *Performance*, and *Frustration* loaded consistently under *Affective and Performance* component, accounting for 65.0% of the total variance.

PCA rotated factor loadings for the XR condition showed similar results, with Mental and Physical Demand loading into the Cognitive Workload component, but effort moved over to the *Affective and Performance* component with *Temporal Demand*, *Performance*, *Effort*, and *Frustration* for 66.9% of the total variance (Tables 15 and 16).

Table 15. National Aeronautics and Space Administration Task Load Index rotated principal component analysis of the Navigation, Seamanship, Shiphandling Trainer condition responses for *All Participants*.

NASA TLX Subscale	Cognitive Workload	Affective/Performance
Mental Demand	**-0.90**	0.26
Physical demand	**-0.96**	0.18
Temporal Demand	-0.03	**0.99**
Performance	-0.25	**0.89**
Effort	**-0.90**	0.24
Frustration	-0.12	**0.97**

Note: The negative signs are not meaningful; magnitude and clustering identify the factors.

Table 16. National Aeronautics and Space Administration Task Load Index rotated principal component analysis of the extended reality condition responses for *All Participants*.

TLX Subscale	Cognitive Workload	Affective/Performance
Mental Demand	**0.91**	-0.25
Physical demand	**0.98**	-0.09
Temporal Demand	-0.03	**-0.97**
Performance	0.27	**-0.90**
Effort	0.26	**-0.90**
Frustration	0.17	**0.88**

Note: The negative signs are not meaningful; magnitude and clustering identify the factors.

7. Power Analysis

A statistical power analysis evaluated how well the response data differentiated distinctions reported in the results. The following is a sample of statistical power findings based on more extreme components' findings determined in the PCA analysis. These corner-cases were *All Participant* and *OOD Only* responses to the NASA TLX subscales of *Mental Demand* and *Performance* (Table 17).

Table 17. National Aeronautics and Space Administration Task Load Index power analysis of compared control and experimental condition findings of mental demand and performance for *All Participants* and *OOD Only*.

Subgroup	NASA TLX Subscale	Effect Size *d*	Wilcoxon Result	Post-Hoc Power	Interpretation
All Participants	Mental Demand	$d = 0.08$	$p = 0.45$	10%	No meaningful difference; trivial effect
All Participants	Performance	$d = 0.98$	$p <$ **0.00*** **	100%	Strong difference
OOD Only	Mental Demand	$d = 0.35$	$p <$ **0.00*** **	24%	Underpowered; null finding not reliable
OOD Only	Performance	$d = 0.82$	$p <$ **0.01** **	83%	Strong effect even with small sample

For both the *All Participants* pool and the *OOD Only* subgroup, *Mental Demand* showed a trivial effect ($d = 0.08$) with very low power (10%), indicating that analysis of the difference between the NSST and XR (for this NASA TLX subscale of *Mental Demand*) should be considered as exploratory. In contrast, the *All Participants* responses showed a significant difference in the NASA TLX *Performance* subscale with a large effect ($d = 0.98$) and a very high statistical power for the test (100%). This confirmed a strong difference in the bridge team perceptions of *Performance*, with a preference being for the NSST. In the *OOD Only* subgroup (n=26), results for *Mental Demand* showed a small to moderate effect ($d = 0.35$), but power was only 24%, making the null result less reliable. However, *OOD Only* responses for *Performance* showed a large effect ($d = 0.82$) with strong statistical power (83%), demonstrating that robust effects can still be detected even with small sample sizes.

Together, the **power analysis highlighted that the analytic framework was sensitive to real differences**, i.e., results found for *Performance*, while correctly showing negligible effects where they exist, i.e., findings for *Mental Demand*. Subgroup analyses were limited by reduced statistical power, indicating that null findings in smaller groups should be interpreted cautiously. Performance preference found for the NSST was not surprising as almost all those who participated in the study had never performed in XR, impacting the perception of their performance during the experimental condition. In aggregate, the **power analysis indicated that follow-on studies with greater sample sizes could contribute to higher statistical power in order to better differentiate NASA TLX findings for bridge team subgroups.**

8. Summary of Key Findings

Overall, NASA TLX survey findings showed that XR responses were, to a large degree, no different from those in the NSST.

- **XR is a viable training method that did not impose extraneous cognitive load on bridge teams participating in the study.**
- **XR seemed to incur a novelty penalty in the NASA TLX subscales of *Performance* and *Frustration*.** As noted, this may be due to first time use of XR by participants.
- **XR system design enhancements for some watch stations, e.g., JOOD, are recommended** to more closely mirror system interactions of equipment used aboard ship.
- **Enhanced XR system "on-ramping" (e.g., training familiarity) may be needed to minimize frustration and enhance perceived performance.** Follow-on studies could control for the nuisance variable of "on-ramping" to understand how to best familiarize bridge crews to XR interfaces and training features to optimize system usability and better facilitate knowledge flow in XR VEs.

B. SYSTEM USABILITY SCALE SURVEY RESULTS

Participants were asked to evaluate system usability in response to 10 System Usability Scale (SUS) subscale questions rated from 1 to 5 in increments of 1 (Table 18).

Table 18. System Usability Scale question key.

SUS Statement	Full SUS Question/Statement
Q1 (Frequent Use)	I think that I would like to use this system frequently.
Q2 (Complexity)	I found the system unnecessarily complex. (R)
Q3 (Easy to Use)	I thought the system was easy to use.
Q4 (Need Support)	I think that I would need the support of a technical person to be able to use this system. (R)
Q5 (Well Integrated)	I found the various functions in this system were well integrated.
Q6 (Inconsistency)	I thought there was too much inconsistency in this system. (R)
Q7 (Learn Quickly)	I would imagine that most people would learn to use this system very quickly.
Q8 (Cumbersome)	I found the system very cumbersome to use. (R)
Q9 (Confidence)	I felt very confident using the system.
Q10 (Learning Needed)	I needed to learn a lot of things before I could get going with this system. (R)

Note: This table lists the SUS statements with shorthand labels alongside the exact wording. Responses for all questions followed a "5 Integer" Likert Scale: 1 = Strongly disagree, 3 = Neutral, 5 = Strongly agree. SUS statements marked (R) were reverse-scored.

SUS measurement increments and questions were derived from the literature (Klug, 2017; Lewis, 2018) with a 1 indicating "Strongly Disagree," 3 being "Neutral," and a 5 indicating "Strongly Agree." Thus, if participant 01–003 provided a 5 for "Q3 (Easy to Use)" in the experimental condition (XR), then participant 01–003 "Strongly Agreed" with the statement "I thought the system (XR) was easy to use." This result was paired with and compared to participant 01–003's response to the same survey question for the control condition (NSST). If the control condition (NSST) response to "Q3 (Easy to Use)" was a 4, then a comparison inference was made that participant 01–003 found XR easier to use than the NSST. If participant 01–003 provided a 3 for Q9 (Confidence) in XR, and a 5 for the same question in NSST, the inference made was that participant 01–003 "felt more

confident using the (NSST) system" than XR, and so forth. After collection and pairing in Excel, the total number of paired results was N = 130.

As provided in the NASA TLX results, a SUS Findings Summary is provided along with SUS response Descriptive Statistics, Distributional Fit, and Normality, Data Reliability and Agreement, NSST vs. XR Comparisons, order effects, PCA, Power Analysis, and a Summary of Key Findings.

1. Findings Summary

Respondents rated XR "usability" largely on par with the NSST. Differences found often depended on the watch station and the SUS statement response, but overall perceptions noted in the comparison analysis (Table 20) of the XR experimental condition were near those of the NSST control condition. Difference largely reflected subgroup responses pointing to a lack of interaction fidelity for specific system interfaces in XR. For example, *JOOD Only* responses for SUS "Q3 Ease of Use" reflected the difficulty of interacting with the primary ship system decision aid, the VMS and RADAR, which was not as strongly reflected in other watch stations as their responsibilities did not require VMS or RADAR interaction. **Statistical analysis showed the SUS survey instrument to be internally consistent, reflecting actual differences and not merely test artifacts.**

2. Descriptive Statistics, Distributional Fit, and Normality

The following section details SUS response analyses for descriptive statistics, distributional fit, and normality.

a. Descriptive Statistics

SUS survey responses were analyzed to determine the four primary statistical moments of center, spread, skew, and kurtosis. The midpoint of the SUS scale was 3, neither agreeing nor disagreeing with the SUS statement (Table 19).

Table 19. System Usability Scale descriptive statistics of Navigation, Seamanship, Shiphandling Trainer and extended reality condition responses for *All Participants* (N = 130).

SUS Statement	Cond.	Mean	95% CI (Lower–Upper)	Med	SD	Skew	Kurt
Q1 (Frequent Use)	NSST	4.28	4.13–4.43	4.00	0.84	-1.51	2.97
Q1 (Frequent Use)	XR	3.59	3.39–3.79	4.00	1.15	-0.54	-0.51
Q2 (Not Complex)	NSST	4.38	4.25–4.51	5.00	0.75	-1.08	0.78
Q2 (Not Complex)	XR	3.94	3.76–4.11	4.00	1.00	-0.86	0.15
Q3 (Easy to Use)	NSST	4.28	4.14–4.42	4.00	0.79	-1.31	2.47
Q3 (Easy to Use)	XR	3.92	3.76–4.09	4.00	0.95	-0.66	-0.17
Q4 (No Tech Support)	NSST	3.14	2.89–3.39	3.00	1.43	-0.06	-1.40
Q4 (No Tech Support)	XR	3.35	3.12–3.57	4.00	1.30	-0.35	-1.03
Q5 (Well Integrated)	NSST	4.16	4.03–4.30	4.00	0.77	-0.60	-0.11
Q5 (Well Integrated)	XR	3.72	3.55–3.88	4.00	0.93	-0.62	0.08
Q6 (Consistency)	NSST	4.22	4.09–4.35	4.00	0.75	-0.39	-1.13
Q6 (Consistency)	XR	3.95	3.80–4.11	4.00	0.89	-0.58	0.00
Q7 (Learn Quickly)	NSST	4.22	4.08–4.35	4.00	0.76	-0.49	-0.14
Q7 (Learn Quickly)	XR	3.98	3.82–4.15	4.00	0.96	-0.92	0.47
Q8 (Not Cumbersome)	NSST	3.97	3.79–4.15	4.00	1.02	-0.97	0.62
Q8 (Not Cumbersome)	XR	3.71	3.52–3.89	4.00	1.07	-0.40	-0.61
Q9 (Confidence)	NSST	4.27	4.15–4.39	4.00	0.68	-0.69	0.59
Q9 (Confidence)	XR	3.62	3.42–3.81	4.00	1.12	-0.43	-0.74
Q10 (Learning Need)	NSST	3.88	3.68–4.09	4.00	1.17	-0.82	-0.31
Q10 (Learning Need)	XR	3.54	3.33–3.75	4.00	1.19	-0.61	-0.46

Note: Cond. = Condition: NSST being Control and XR being Experimental. C.I. = Confidence Interval.

NSST response distributions were more consistent than those of XR, with the average standard deviation for XR being higher than the NSST. NSST responses average mean was higher (4.08) with XR's average determined as 3.73. On an individual statement level, *All Participant* responses for SUS Q4 (No Tech Support) favored XR, with the responses for the other responses trending higher for the NSST for the other statements.

b. Distributional Fit

Across all SUS responses, the best distributional fit was rarely Normal. Instead, the Logistic family (with occasional Normal and Weibull T alternatives), yielding S-shaped cumulative distribution functions, supporting the use of non-parametric statistical test methods (Table 20).

Table 20. System Usability Scale distributional fit analysis of Navigation, Seamanship, Shiphandling Trainer and extended reality condition responses for *All Participants* (N = 130).

SUS Statement	Condition	"Best Fits" (by criteria)	Normal? (overall)
Q1 (Frequent Use)	NSST	Logistic; Triangular; Beta 4 (Normal mid-rank under some criteria)	No
Q1 (Frequent Use)	XR	Lognormal 3/Normal; Gumbel-Min; Logistic (Normal tops AD & Kuiper)	Borderline
Q2 (Complexity)	NSST	Lognormal 3; Gumbel Max; Beta 4/ Erlang (clear skew)	No
Q2 (Complexity)	XR	Beta 4; Weibull; Gumbel-Min/ Logistic (Normal lower-rank)	No
Q3 (Ease of Use)	NSST	Logistic; Weibull; Beta 4 (Normal occasionally mid-rank)	No
Q3 (Ease of Use)	XR	Logistic; T-Shifted; Weibull (Normal lower-rank)	No
Q4 (Support Needed)	NSST	Logistic; T-Shifted; Weibull (Normal mid-rank once)	Borderline
Q4 (Support Needed)	XR	Cosine; Weibull; Gumbel Max/ Gamma (Normal best on AD)	Borderline
Q5 (Integration)	NSST	Beta 4; Logistic; Weibull	No
Q5 (Integration)	XR	Logistic; Beta 4; T-Shifted/ Weibull (Normal lower-rank)	No
Q6 (Consistency)	NSST	Logistic; Weibull; Laplace	No
Q6 (Consistency)	XR	Logistic; T-Shifted; Lognormal 3 (Normal mid-rank once)	No
Q7 (Learn Quickly)	NSST	Logistic; T-Shifted; Weibull (Normal mid-rank)	No
Q7 (Learn Quickly)	XR	Weibull; Lognormal 3; Normal (Normal AD-best once)	Borderline
Q8 (Cumbersome)	NSST	Logistic; Gumbel-Min; Beta 4	No
Q8 (Cumbersome)	XR	Weibull; Lognormal 3; Beta 4/ Logistic (Normal mid-rank)	No
Q9 (Confidence)	NSST	Logistic; Weibull; Triangular (Normal mid-rank)	No
Q9 (Confidence)	XR	Weibull; Normal; T (Logistic mid-rank)	No
Q10 (Learning Needed)	NSST	Logistic; Triangular; Gumbel-Min (KS favored GenPareto artifact)	No
Q10 (Learning Needed)	XR	Triangular; Logistic; Normal/ Beta 4 (Normal mid-rank)	No

Notes: "Borderline" indicates that "Normal" ranked best under one goodness-of-fit criterion (e.g., Anderson–Darling or Kuiper) but was not "Normal" across other fit testing criteria.

Unlike the TLX distributions, where many of the best-fitting subscale distributions were PERT and triangular-like (e.g., Beta-PERT or triangular), SUS distributions often favored the Logistic family with a strong central mass, and heavier tails than a Normal distribution. Interpreted together, these results suggest that **TLX responses behaved like**

constrained workload judgments with sharper peaks depending on subscale, whereas SUS responses reflected a smoother, more continuous spread of perceived usability across the scale. This finding implies progressive changes in SUS response probabilities, rather than abrupt jumps, with room for a few low- or high-outlier ratings without dominant or extreme skew.

c. Normality Tests

Shapiro–Wilk (SW) and D'Agostino–Pearson (DAP) tests confirmed **data response distributions were non-normal, justifying non-parametric comparison testing procedures** (Table 21).

Table 21. System Usability Scale Shapiro–Wilk and D'Agostino–Pearson normality checks of Navigation, Seamanship, Shiphandling Trainer and extended reality condition responses for *All Participants* (N = 130).

SUS Statement	Condition	SW p-value	DAP p-value	Normal? (overall)
Q1 (Frequent Use)	NSST	< 0.0001***	< 0.0001***	No
Q1 (Frequent Use)	XR	< 0.0001***	0.0194**	No
Q2 (Complexity)	NSST	< 0.0001***	< 0.0001***	No
Q2 (Complexity)	XR	< 0.0001***	0.0003***	No
Q3 (Ease of Use)	NSST	< 0.0001***	< 0.0001***	No
Q3 (Ease of Use)	XR	< 0.0001***	0.0072**	No
Q4 (Support Needed)	NSST	< 0.0001***	0.0040**	No
Q4 (Support Needed)	XR	< 0.0001***	0.0132**	No
Q5 (Integration)	NSST	< 0.0001***	0.0179**	No
Q5 (Integration)	XR	< 0.0001***	0.0141**	No
Q6 (Consistency)	NSST	< 0.0001***	0.0052**	No
Q6 (Consistency)	XR	< 0.0001***	0.0230**	No
Q7 (Learn Quickly)	NSST	< 0.0001***	0.0650*	Mixed
Q7 (Learn Quickly)	XR	< 0.0001***	< 0.0001***	No
Q8 (Not Cumbersome)	NSST	< 0.0001***	< 0.0001***	No
Q8 (Not Cumbersome)	XR	< 0.0001***	0.0594*	Mixed
Q9 (Confidence)	NSST	< 0.0001***	0.0018**	No
Q9 (Confidence)	XR	< 0.0001***	0.0283**	No
Q10 (Learning Needed)	NSST	< 0.0001***	0.0005***	No
Q10 (Learning Needed)	XR	< 0.0001***	0.0085**	No

Note. "SW" = Shapiro–Wilk; "DAP" = D'Agostino–Pearson. "Normal" indicates both test results found $p \geq 0.05$. "Mixed" indicates one test found data normality ($p \geq 0.05$) and one did not ($p \leq 0.05$); only NSST Q7 and XR Q8 fail to reject normality.

3. Data Reliability and Agreement

Although Kendall's W was used to assess the degree of concordance in NASA TLX workload rankings, this test was not appropriate for SUS, where responses reflect discrete statements that aggregate into a single score. Guttman' Lambda, on the other hand, was able to confirm SUS data **ICR was sufficient in both control and experimental conditions to support subsequent comparison analyses.** Although participants rated SUS statements consistently, some variation was found in different subgroups in different conditions.

a. Guttman's Lambda and Spearman–Brown (Internal Consistency)

Lambda values fell in the 0.41–0.95 range, reflecting moderate to high internal consistency across SUS responses. In the NSST (control condition), $\lambda = 0.654$ (moderate) and Split-half (Spearman–Brown) = 0.779. In the XR (experimental condition), $\lambda = 0.798$ (good) and Split-half (Spearman–Brown) = 0.909, showing stronger internal consistency in XR than NSST (+0.14 in λ). Spearman–Brown corrections corroborate high split-half reliability for XR in *CONN Only* and *NAV Only* (> 0.94), supporting consistency.

The only notable outlier was *OOD Only* in XR with low λ, which may be attributable to the OOD being the only watch station to use a Meta Quest Pro (all others used Meta Quest 3 HMDs). **Split-half and odd-even correlations were variable but positive across all groups, indicating no evidence of systematic instability in subscale structure** (Table 22).

Table 22. System Usability Scale Guttman's Lamda and split-half internal consistency tests of Navigation, Seamanship, Shiphandling Trainer and extended reality condition responses for *All Participants* and *Subgroups*.

Station	Sample	Condition	Guttman's λ	Split-half (r)	Split-half (SB)	Odd-even (r)	Odd-even (SB)
All Participants	N = 130	NSST	0.654	0.638	0.779	0.492	0.659
All Participants	N = 130	XR	0.798	0.834	0.909	0.664	0.798
OOD	n = 26	NSST	0.619	0.563	0.721	0.457	0.628
OOD	n = 26	XR	0.407	0.786	0.880	0.264	0.418
JOOD	n = 25	NSST	0.570	0.636	0.778	0.400	0.571
JOOD	n = 25	XR	0.840	0.809	0.894	0.741	0.851
CONN	n = 26	NSST	0.775	0.593	0.745	0.632	0.775
CONN	n = 26	XR	0.952	0.860	0.925	0.911	0.953
NAV	n = 16	NSST	0.652	0.489	0.657	0.511	0.676
NAV	n = 16	XR	0.834	0.896	0.945	0.733	0.846
HELM/LEE HELM	n = 27	NSST	0.504	0.614	0.761	0.346	0.514
HELM/LEE HELM	n = 27	XR	0.661	0.816	0.899	0.495	0.663

Notes: As rules of thumb for SUS internal consistency, Guttman's λ ≥0.70 is often interpreted as "acceptable," ≥ 0.80 as "good," and ≥ 0.90 as "excellent."

4. Statistical Comparisons: Navigation, Seamanship, Shiphandling Trainer versus Extended Reality

Multivariate Hotelling T^2 tests, WSRTs, Cohen's *d* effect size, and Bonferroni tests were used to compare SUS XR vs. NSST responses. Hotelling's T^2 showed a strong overall difference for *All Participants*, *JOOD Only*, and *HELM/LEE HELM Only* subgroups. Notably three subgroups: *OOD Only*, *CONN Only*, and *NAV Only* did not show a significant difference between conditions (Table 23).

WSRT differences for *All Participants* were statistically significant for nearly all SUS statements, favoring NSST, with the only exceptions being "Q4 Need Support" and "Q7 Learn Quickly" being insignificantly different. **Subgroup analyses showed far fewer significant differences between NSST and XR than *All Participants*. While NSST was rated higher when a statistical difference was found, it was found for only a minority of SUS statements** (Table 23).

Cohen's *d* effect sizes were typically in the medium range, with moderate effects concentrated in "Q1 Frequent Use," "Q2 Complexity," and "Q9 Confidence," depending on the subgroup analyzed; in particular *HELM/LEE HELM Only* (Q2 Complexity), *NAV Only* (Q9 Confidence), and *JOOD Only* (Q3 Ease of Use) noted large *d* values (Table 23).

Bonferroni Correction found that only the *All Participant* responses for "Q1 Frequent Use" and "Q9 Confidence" fell outside expected CIs (High); all other comparisons fell within expected limits. In this finding, only Q1 "I would like to use this system frequently" and Q9 "I felt very confident in using the system" were noted as impactful differences between NSST and XR. These differences favored NSST and may be indicative of an XR first use/novelty penalty (Table 23).

Table 23. System Usability Scale Hotelling's T^2, Wilcoxon signed-rank, Cohen's *d*, and Bonferroni comparison tests of Navigation, Seamanship, Shiphandling Trainer versus extended reality condition responses for *All Participants* and *Subgroups*.

Subgroup	Hotelling's T^2 (omnibus)	Wilcoxon Significant Items ($p < .05$)	Cohen's *d* (effect size)	Bonferroni Correction
All Participants (N = 130)	$T^2 = 76.59$ $F = 7.12$ $p <$ **0.001*** **	Q1, Q2, Q3, Q5, Q6, Q8, Q9, Q10	Q1 *d* = 0.68 Q2 *d* = 0.50 Q5 *d* = 0.52 Q9 *d* = 0.70	Q1,Q9 "High" All others within CI
OOD (N = 26)	$T^2 = 22.47$ $F = 1.44$ $p = 0.249$	Q1, Q5, Q9	Q1 *d* = 0.72 Q5 *d* = 0.70 Q9 *d* = 0.72	All within CI
JOOD (N = 25)	$T^2 = 58.83$ $F = 3.68$ $p =$ **0.012****	Q3, Q9	Q3 *d* = 0.80 Q9 *d* = 0.66	All within CI
CONN (N = 26)	$T^2 = 18.92$ $F = 1.21$ $p = 0.353$	Q7, Q8, Q9	Q7 *d* = 0.42 Q8 *d* = 0.38 Q9 *d* = 0.50	All within CI
NAV (N = 16)	$T^2 = 25.68$ $F = 1.03$ $p = 0.511$	Q1, Q6, Q9	Q9 *d* = 1.23 Q6 *d* = 0.87 Q1 *d* = 0.81	All within CI
HELM/LEE HELM (N = 27)	$T^2 = 70.26$ $F = 4.59$ $p =$ **0.003****	Q1, Q2, Q5, Q9	Q2 *d* = 1.09 Q1 *d* = 0.79 Q5 *d* = 0.68 Q9 *d* = 0.77	All within CI

5. Order Effects

Research hypothesis H_4 sought to determine the training impacts of using XR before training in the NSST. To characterize XR training impacts, a comparison was made between NSST survey responses taken after participants performed XR training (labeled "NSST 2") and those who had not yet been exposed to XR (labeled "NSST 1").

Comparison showed that order effects were minimal, but statistically significantly different for SUS statements "Q1 Frequent Use," "Q2 Complexity," and "Q3 Ease of Use," with the results being higher for those who experienced XR first. Effect sizes (Cohen's *d*) corresponding to these differences were found to be small to medium, with overall effect sizes being small to insignificant for the other statements (Q4–Q10), revealing little to no order bias. **This finding indicates that prior exposure to XR did not systematically alter usability ratings of NSST, and for a minority of statements (Q1–Q3) may have elevated NSST SUS perceptions.** Trends suggested other elevations in NSST ratings after using XR, but Cohen's *d* values for these findings were generally too small to draw conclusions, underscoring a need for further research (Table 24).

Table 24. System Usability Scale order effects analysis of Navigation, Seamanship, Shiphandling Trainer response after exposure to extended reality for *All Participants* (N = 130).

SUS Statement	XR 1st?	Median	MW p-value	Cohen's *d*	95% CI (Lower–Upper)
Q1 (Use Frequently)	Yes	5.0	**0.040****	*d = 0.32*	-0.03–0.58
Q1 (Use Frequently)	No	4.0			
Q2 (Not Complex)	Yes	5.0	**0.001*****	*d = 0.55*	0.14–0.66
Q2 (Not Complex)	No	4.0			
Q3 (Easy to Use)	Yes	5.0	**0.002****	*d = 0.48*	0.10–0.65
Q3 (Easy to Use)	No	4.0			
Q4 (No Tech Support)	Yes	3.0	0.101	*d = 0.32*	-0.05–0.98
Q4 (No Tech Support)	No	2.5			
Q5 (Functions Integrated)	Yes	4.0	0.280	*d = 0.17*	-0.14–0.40
Q5 (Functions Integrated)	No	4.0			
Q6 (Not Inconsistent)	Yes	4.0	0.299	*d = 0.19*	-0.12–0.41
Q6 (Not Inconsistent)	No	4.0			
Q7 (Quick to Learn)	Yes	4.0	0.060	*d = 0.34*	-0.01–0.53
Q7 (Quick to Learn)	No	4.0			
Q8 (Not Cumbersome)	Yes	4.0	0.390	*d = -0.20*	-0.58–0.16
Q8 (Not Cumbersome)	No	4.0			
Q9 (Confident)	Yes	4.0	0.357	*d = 0.16*	-0.13–0.36
Q9 (Confident)	No	4.0			
Q10 (Learning Needed)	Yes	4.0	0.055	*d = 0.40*	0.06–0.88
Q10 (Learning Needed)	No	4.0			

Note: "XR 1st?" indicates whether participants completed XR training before NSST. Yes = XR-first condition; No = no XR prior to NSST. MW = Mann–Whitney Test.

6. Principal Component Analysis

PCA with Varimax rotation revealed two consistent components across SUS findings for both NSST and XR conditions: 1) Ease and Integration and 2) Confidence and Attitude. The first component, **Ease and Integration,** was dominated by statements related to ease of use, integration of functions, consistency, and the ability to quickly learn the system. The second component, **Confidence and Attitude,** was defined primarily by statements concerning confidence in using the system, reduced need for support, and willingness to use the system frequently.

Together, these two components capture both the practical usability of the system and the affective stance of participants toward using it. Eigenvalues showed these first two components explain approximately 53–56% of the variance in NSST and XR usability perceptions. **The similarity of this factor structure across both conditions suggests that the differences observed in SUS ratings are not due to measurement artifacts but reflect genuine differences in how participants experienced the training systems.** The experimental condition components were broadly similar, with the notable exception of certain confidence-related statements, likely an impact of participants using XR for the first time (Tables 25 and 26).

Table 25. System Usability Scale principal component analysis of Navigation, Seamanship, Shiphandling Trainer condition responses for *All Participants*.

NSST PCA Factor	Eigenvalue	Proportion of Variance	Cumulative Proportion
Factor 1	3.25	32.5%	32.5%
Factor 2	2.05	20.5%	53.0%
Factor 3	1.45	14.5%	67.5%
Factor 4	0.92	9.2%	76.7%
Factor 5	0.71	7.1%	83.8%
Factor 6	0.62	6.2%	90.0%
Factor 7	0.50	5.0%	95.0%
Factor 8	0.45	4.5%	99.5%
Factor 9	0.05	0.5%	100.0%
Factor 10	0.00	0.0%	100.0%

Table 26. System Usability Scale principal component analysis of extended reality condition responses for *All Participants*.

XR PCA Factor	Eigenvalue	Proportion of Variance	Cumulative Proportion
Factor 1	3.40	34.0%	34.0%
Factor 2	2.15	21.5%	55.5%
Factor 3	1.55	15.5%	71.0%
Factor 4	0.90	9.0%	80.0%
Factor 5	0.70	7.0%	87.0%
Factor 6	0.60	6.0%	93.0%
factor 7	0.45	4.5%	97.5%
Factor 8	0.25	2.5%	100.0%
Factor 9	0.00	0.0%	100.0%
Factor 10	0.00	0.0%	100.0%

Rotated PCA factor loadings highlighted which SUS statements aligned most strongly with each factor, *Ease and Integration* and *Confidence and Attitude*, respectively. Findings replicated the classic SUS two-factor structure of *Usability* and a smaller factor of *Learnability* (Lewis, 2017).

For the NSST control condition, responses to SUS statements Q1, Q2, Q3, Q5, Q6, Q7, and Q9 loaded strongly, representing general system *Usability*. SUS statements Q4 and Q10 loaded consistently with system *Learnability*. SUS statements Q7 and Q8 showed moderate cross-loadings in both *Usability* and *Learnability*. In aggregate, *Usability* and *Learnability* accounted for 52.5% of the total NSST variance (Table 27).

Table 27. System Usability Scale rotated principal component analysis of Navigation, Seamanship, Shiphandling Trainer condition responses for *All Participants*.

SUS Statement	Ease and Integration (Usability)	Confidence and Attitude (Learnability)
Q1 (Use Frequently)	**-0.677**	-0.235
Q2 (Not Complex)	**-0.591**	0.426
Q3 (Easy to Use)	**-0.721**	0.117
Q4 (No Tech Support)	-0.080	**0.659**
Q5 (Functions Integrated)	**-0.799**	0.064
Q6 (Not Inconsistent)	**-0.537**	0.419
Q7 (Quick to Learn)	**-0.578**	0.480
Q8 (Not Cumbersome)	-0.027	0.472
Q9 (Confident)	**-0.761**	0.359
Q10 (Learning Needed)	-0.152	**0.776**

Note: The negative signs of loadings are not meaningful in this context; what matters is the magnitude (**bolded**) and clustering of items within each factor.

Rotated PCA results for the XR experimental condition also revealed *Usability* and *Learnability* as primary components. SUS statement results from Q1, Q3, Q5, Q7, and Q9 loaded strongly for *Usability*, with additional moderate contributions from Q2, Q6, and Q8, accounting for approximately 58% of the variance. Results from SUS statements Q4 and Q10 loaded very strongly for *Learnability*, consistent with SUS literature findings (Bangor, 2008; Khan, 2025; Lewis, 2009, 2017). Unlike SUS findings in the NSST, moderate cross-loading was found in XR PCA results, e.g., Q2, Q3, Q7, Q9, suggesting that learning aspects were more salient in the XR condition (Table 28).

Table 28. System Usability Scale rotated principal component analysis of extended reality condition responses for *All Participants*.

SUS Statement	Ease and Integration (Usability)	Confidence and Attitude (Learnability)
Q1 (Use Frequently)	**-0.834**	0.093
Q2 (Not Complex)	-0.367	0.634
Q3 (Easy to Use)	**-0.550**	0.569
Q4 (No Tech Support)	-0.153	**0.766**
Q5 (Functions Integrated)	**-0.797**	0.079
Q6 (Not Inconsistent)	-0.399	0.464
Q7 (Quick to Learn)	**-0.601**	0.566
Q8 (Not Cumbersome)	-0.432	0.342
Q9 (Confident)	**-0.671**	0.422
Q10 (Learning Needed)	-0.057	**0.841**

As an additional exploration of the data, PCA analysis of results filtered for the "OOD Only" subgroup was conducted. In this analysis, SUS Q4 (Technical Support) and Q10 (Learning Needed) were removed to improve clarity of the factor structures. **By trimming Q4 and Q10, NSST results tended to disperse across multiple smaller factors, indicating a fragmented perception of usability. In contrast, XR responses clustered into two coherent factors of *Adoption* and *Usability*.** The adoption factor clustered SUS Q1 and Q8, representing "a willingness to use" and "a lack of cumbersomeness." The usability factor clustered SUS Q3 and Q9, reflecting ease of use and confidence. Other items (Q2, Q5, Q6, Q7) loaded onto separate smaller factors, suggesting fragmentation likely due to small sample size (Tables 29 and 30).

Table 29. System Usability Scale principal component factor loadings of extended reality condition responses for *OOD Only*.

SUS Statement	Factor							
	1	2	3	4	5	6	7	8
Q1 (Use Frequently)	0.92	-	-	-	-	-	-	-
Q2 (Not Complex)	-	-	-	-	-	0.90	-	-
Q3 (Easy To Use)	-	-	-	-	-	-	0.89	-
Q5 (Functions Integrated)	-	-	-0.95	-	-	-	-	-
Q6 (Not Inconsistent)	-	-	-	-0.94	-	-	-	-
Q7 (Quick To Learn)	-	-	-	-	-0.96	-	-	-
Q8 (Not Cumbersome)	0.95	-	-	-	-	-	-	-
Q9 (Confident)	-	-	-	-	-	-	-0.95	-

Table 30. System Usability Scale principal component analysis of extended reality conditions to statement summary for *OOD Only*.

PCA Factor "OOD Only"	SUS Statement Correlation
Factor 1 (Adoption)	Q1 and Q8: willingness to use, lack of cumbersomeness
Factor 7 (Usability)	Q3 and Q9: ease of use, confidence
Factor 6 (Simplicity)	Q2: perceived simplicity, stands alone
Factors 3, 4, 5 (Fragmented Usability)	Q5, Q6, Q7 split into separate factors due to small sample size

Factor loadings for the *OOD Only* subgroup suggests that XR delivers a more unified and interpretable experience for OODs, with usability perceptions strongly linked to confidence. In contrast, findings from the NSST condition appear more fragmented, implying that OODs weigh individual aspects of usability separately in the NSST rather than as an integrated whole in XR (Table 31).

Table 31. System Usability Scale statement to factor principal component analysis summary of Navigation, Seamanship, Shiphandling Trainer and extended reality conditions for *OOD Only*.

SUS Full Statement	NSST (Control) Factor Loadings	XR (Experimental) Factor Loadings
Q1: I think that I would like to use this system frequently.	Factor 5 (≈0.92)	Factor 2 (≈0.92)
Q2: I found the system NOT unnecessarily complex.	Factor 6 (≈0.90)	Factor 6 (≈0.90)
Q3: I thought the system was easy to use.	Factor 7 (≈0.89)	Factor 7 (≈0.89)
Q5: I found the various functions in this system were well integrated.	Factor 3 (≈0.95)	Factor 3 (≈0.95)
Q6: I thought there was NOT too much inconsistency in this system.	Factor 5 (≈0.87)	Factor 4 (≈0.94)
Q7: I imagine that most people would learn to use this system very quickly.	Factor 7 (≈0.95)	Factor 5 (≈0.95)
Q8: I found the system NOT very cumbersome to use.	Factor 4 (≈0.98)	Factor 1 (≈0.95)
Q9: I felt very confident in the system.	Factor 9 (≈0.95)	Factor 2 (≈0.92)

Note: Factor loadings shown represent each item's **highest loading** following Varimax rotation. Values are rounded to two decimals and denoted as approximate (≈) to indicate that small cross-loadings on other factors were suppressed for clarity.

Unfortunately, the sample size for *OOD Only* was small and resulted in some factors splitting into single-item components. These items would be expected to merge into broader usability factors in larger samples. Given this small sample size, these results must be considered exploratory only until further samples corroborate these findings.

7. Power Analysis

A statistical power analysis was conducted to evaluate how effectively the SUS data distinguished differences in perceived usability between the NSST and XR conditions. As with the NASA TLX analysis, the analyses examined corner-cases with a focus on the *All Participants* and *OOD Only* subgroups.

a. SUS All Participants Power Analysis

Following the same corner case logic applied in the NASA-TLX analysis, representative extremes derived from the SUS PCA results were selected. The analysis examined dominant SUS PCA components, *Ease and Integration* and *Confidence and Attitude*, for statistical power based on paired SUS statements that loaded strongly or

weakly on their respective factors. The *Ease and Integration* component was represented by SUS statements Q2 (Not complex) and Q6 (Not inconsistent), while the *Confidence and Attitude* component was represented by SUS statement Q9 (Confident) and Q4 (No tech support). **This approach allowed a targeted, but limited, examination of measurement sensitivity across PCA component usability dimensions** (Table 32).

Table 32. System Usability Scale power analysis of "ease and integration" and "confidence and attitude" components of Navigation, Seamanship, Shiphandling Trainer and extended reality conditions for *All Participants*.

Subgroup	SUS Corner-Case and PCA Component	Effect Size (*d*)	Power (1-β)	Interpretation
All Participants	Q2 (Not complex)	0.50	97.8%	Medium effect; strong power and reliable result
All Participants	Q6 (Not inconsistent)	0.33	74.9%	Small effect; marginal power (just below 80%)
PCA Component Ease and Integration	Composite: Ease/ Integration (Q2 + Q6)	—	—	Overall moderate sensitivity: small effects may be missed
All Participants	Q9 (Confident)	0.70	99.99%	Large effect; excellent power
All Participants	Q4 (No tech support)	0.15	22.9%	Very small effect; underpowered
PCA Component Confidence and Attitude	Composite: Confidence/Attitude (Q9 + Q4)	—	—	Strong difference overall, driven by Q9 confidence ratings

Note: Interpretation of effect sizes, e.g., "moderate," are relative to effect sizes found in the overall population and do not reflect an absolute measure.

For the Ease and Integration component, effect sizes ranged from d=0.33 to d=0.50, with a statistical power between approximately 75% and 98%, respectively. **These results indicated that the SUS data were sufficiently sensitive to detect medium usability differences between NSST and XR systems.** By contrast, Confidence and Attitude component results revealed one relatively high-powered statistical difference (Q9

[Confidence], $d = 0.70$; 1-β=99.99%) alongside one of low-power (Q4 [No tech support], $d = 0.15$; 1-β=22.9%). Together, these patterns mirror the NASA-TLX findings: the analytic framework effectively detects real differences while properly indicating trivial ones. As with the NASA TLX subscales, **SUS subgroup-specific analyses were limited by reduced power, suggesting that future studies with larger samples may help resolve the subtle usability differences between physical mockups, like the NSST, and XR.**

b. SUS OOD Only Power Analysis

Statistical power analyses for the *OOD Only* subgroup evaluated the same PCA components of *Ease and Integration* (with contributions from SUS statements Q5, Q6, and Q7) and *Confidence and Attitude* (contributions from Q1, Q8, and Q9). Given the smaller sample size (n = 26), **these analyses should be considered with caution. While they provide insight, e.g.− where the measurement framework remains robust, where limited power may constrain interpretation, etc.− they suggest further research is needed with a larger sample** (Table 33).

Table 33. System Usability Scale power analysis "ease and integration" and "confidence and attitude" components of Navigation, Seamanship, Shiphandling Trainer and extended reality conditions for *OOD Only*.

Subgroup	SUS Corner-Case Component	Effect Size (*d*)	Power (1-β)	Interpretation
OOD Only	Q5 (Functions integrated)	0.70	69.1%	Medium-large effect; moderate power
OOD Only	Q6 (Not inconsistent)	0.14	7.9%	Small effect; underpowered
OOD Only	Q7 (Quick to learn)	0.00	0%	Small measurable difference; effectively equivalent
PCA Component	Composite: Ease/Integration (Q5 + Q6 + Q7)	—	—	Highly variable sensitivity; small effects undetectable
OOD Only	Q1 (Use frequently)	0.72	71.8%	Medium-large effect; moderately strong difference
OOD Only	Q8 (Not cumbersome)	0.24	13.3%	Small effect; underpowered
OOD Only	Q9 (Confident)	0.72	71.6%	Medium-large effect; moderate power
PCA Component	Composite: Confidence/Attitude (Q1 + Q8 + Q9)	—	—	Strong differences driven by Q1 and Q9 confidence/ adoption signals

Note: Interpretations of effect sizes, e.g., "moderate," are relative to effect sizes found in the overall population and do not reflect an absolute measure.

Results for *OOD Only* SUS statements Q6 and Q7 power analysis showed small to negligible effects ($d \leq 0.14$), with power levels well below 20 %, indicating that null results in this domain are inconclusive due to limited sensitivity. However, Q5 revealed a moderate-to-large effect (d=0.70; 1-β=69%), suggesting some OOD participants did perceive differences in system integration. In contrast, the confidence and attitude component analysis found relatively medium-to-large effects for Q1 and Q9 (d=0.72; 1-β=72%); **even with the small sample, these findings support the *OOD Only* subgroup willingness to use and confidence in the XR system**. Q8 remained underpowered (d=0.24; 1-β=13%). Together, these results indicate that, **for OOD participants, the SUS measurement retained sensitivity to large effects in perceived adoption and confidence but cannot reliably detect small differences in fragmented usability items.** This finding mirrors the pattern seen in the *All Participant* analysis, where XR effects were concentrated in confidence-related dimensions, reinforcing that perceptual adoption cues, rather than minor usability nuances, drive OOD differences. Future studies with larger subgroup samples would further clarify the smaller-magnitude usability effects observed here.

8. Summary of Key Findings

Comparative analyses of SUS responses indicated that, overall, XR usability ratings approached, and for some subgroups and SUS statements were statistically the same, as those of the NSST.

- As with the TLX findings, this research supports XR facilitation of knowledge flow in terms of perceived system usability responses of the bridge teams who participated in the study.
- Group-level analyses and effect sizes showed small to moderate differences, with XR often rated on par with the NSST, particularly by junior roles (e.g., *CONN Only*). Certain subgroups, however, demonstrated variation: OODs tended to show mixed or fragmented usability perceptions in NSST, while XR responses were more coherent and

confidence linked and suggested a greater sense of immersion in the XR scenario than in the NSST.

- Bonferroni-corrected tests further highlighted significant differences that were concentrated in specific roles or SUS statements rather than across the full sample. A possible reason for this difference may have been that **each watchstander had a different set of interaction requirements, e.g., JOOD interacted more with VMS than CONN, leading to a different subjective experience of the XR simulation, particularly when the XR interaction didn't meet expectations of the VMS interaction provided in the NSST.**
- Order effect analyses (XR before NSST vs. NSST without XR exposure) showed that experiencing XR before NSST did not negatively impact usability ratings of NSST. In certain cases, SUS Q1, Q2, and Q3 XR first seemed to meaningfully improve NSST user perceptions. This suggests that the systems were evaluated on their own merits rather than biased by sequence.
- Taken together, these SUS findings suggest that XR delivers a unified and interpretable "usability experience" nearly on par with NSST benchmarks. In addition, findings also reflect subgroup differences favoring the NSST, e.g., *JOOD Only*, *HELM/LEE HELM* Only. This variance may have been attributable in cases where XR interaction expectations were not met or suffered a novelty penalty.

C. MARITIME KNOWLEDGE AND PERFORMANCE RESULTS

In addition to the NASA TLX and SUS questionnaires, participants were asked to complete additional maritime knowledge questionnaires based on guidance from MSTCPAC staff and the USN bridge training requirements outlined in Commander Naval Surface Forces Pacific and Commander Naval Surface Forces Atlantic Instruction 3505 1C N7 08FEB22 guidance. The questionnaires included a Bridge Watchstander Self-Assessment Questionnaire issued to all participants, an OOD Self-Assessment

Questionnaire issued only to participating OODs, and an Instructor Bridge Watch Team Assessment issued only to the contract instructor staff. The following section covers each of these survey analyses and details the findings for each.

1. Bridge Watchstander Self-Assessment Survey

Participants were asked to evaluate their bridge watchstanding performance in response to 29 self-assessment survey questions rated from one to seven in increments of one (Table 34).

Table 34. Bridge watchstander self-assessment question key.

Self-Assessment Statement	Full Question Statement
Q1 (Reaction Time)	I was able to react in sufficient time to commands given.
Q2 (Decision Time) (R)	I could not make decisions with sufficient time for the action to be carried out.
Q3 (Mission Success) (R)	I was not able to accomplish the scenario's mission.
Q4 (Comm Combat Information Center [CIC])	I was able to consistently communicate with CIC.
Q5 (Comm Central Control Station (CCS)) (R)	I could not effectively exercise communication procedures with Central Control Station.
Q6 (Bridge to Bridge)	I was able to communicate with other vessels via Bridge to Bridge radio when needed.
Q7 (Focus) (R)	I could not focus on the correct AOI for the situation at hand.
Q8 (Rules of Road)	I was able to correctly apply Rules of the Road.
Q9 (OOD Backup) (R)	I was not able to provide forceful OOD backup.
Q10 (Notify OOD)	I was able to notify the OOD of potential violations of Rules of the Road.
Q11 (Collision Avoidance)	I was able to identify and point out potential collision/near miss situations.
Q12 (Perceive Danger) (R)	I was not able to perceive potential dangers.
Q13 (Team Commands)	My bridge team used standard commands to effectively conn the ship.
Q14 (Set and Drift) (R)	My bridge team did not perceive set and drift conditions to conn the ship.
Q15 (Situational Foresight)	My bridge team foresaw situational dynamics needed to conn the ship.
Q16 (COLREGS) (R)	My bridge team was not able to operate our ship in accordance with COLREGS.
Q17 (System Ease)	It was easy for me to use the training system.
Q18 (VMS Trouble) (R)	I had trouble interacting with the simulated VMS.
Q19 (RADAR Ease)	I could easily interact with the simulated RADAR.
Q20 (HELM Trouble) (R)	I had trouble interacting with the simulated HELM/Lee HELM controls.
Q21 (Team Interaction)	It was easy to interact with my fellow bridge team members during the training scenario.
Q22 (System Interference) (R)	The training system interfered with learning of the content in the scenario.

Self-Assessment Statement	Full Question Statement
Q23 (Skills Development) (R)	The training system will not help refine my seamanship/shiphandling skills.
Q24 (Immersion)	After the training began, I forgot about the training system and was immersed in the scenario.
Q25 (Bridge Realism) (R)	The training system was not a good representation of a ship's bridge environment.
Q26 (Team Cohesion)	Using this training system helps build bridge team cohesion.
Q27 (COLREGS Practice) (R)	The training system does not provide a sufficient way to practice proper execution of COLREGS.
Q28 (Motion Set and Drift) (R)	The training system did not give the motion sense needed in assessing Set & Drift.
Q29 (Motion Advance & Transfer)	The training system gave the motion sense needed in assessing Advance & Transfer.

Note: This table lists the bridge watchstander self-assessment statements with shorthand labels alongside the exact wording. Responses include 1=Strongly Disagree, 4=Neither Agree nor Disagree, 7=Strongly Agree, with N/A available for non-applicable situations. Items marked (R) were reverse-coded.

The self-assessment questions were derived from Commander Naval Surface Forces Pacific and Commander Naval Surface Forces Atlantic Instruction 3505 1C N7 08FEB22 with Likert scale integer measurement of 1 indicating "Strongly Disagree," 4 being "Neutral," and 7 indicating "Strongly Agree." As with the previous two surveys, if participant 16–078 provided a 6 for "Q24 Immersion" in the experimental condition (XR), then participant 16–078 "Agreed" with the statement "After the training began, I forgot about the (XR) training system and was immersed in the scenario." This result was paired and compared to participant 16–078's response to the same survey question for the control condition (NSST). If the control condition (NSST) response to "Q24 Immersion" was a 5, then a comparison inference was made that participant 16–078 found XR more immersive than the NSST. If participant 16–078 provided a 4 for "Q15 Situational Foresight" in XR, and a 5 for the same question in NSST, the inference made was that participant 16–078 felt that "My bridge team foresaw situational dynamics needed to conn the ship" was better facilitated in the NSST than in XR, and so forth. After paring responses in Excel, the total number of paired results varied for each question. This was largely because "N/A" or "Not Applicable" was a possible answer for every question, leaving the choice up to every watchstander to decide whether to answer the question or circle "N/A" which impacted "n" values for each analysis performed.

As with the previous survey analyses, this section provides a Findings Summary along with analyses for Descriptive Statistics, Distributional Fit, and Normality; Data Reliability and Agreement; NSST vs. XR Comparisons; order effects; PCA; Power Analysis; and a Summary of Key Findings.

a. Findings Summary

In 23 of 29 BWS self-assessment questions, NSST ratings were statistically significantly higher than XR, reflecting greater familiarity and habitual confidence in the physical mockup environment. However, when examined by watchstanding subgroup, the number of significant NSST–XR differences dropped sharply; **most subgroup-level comparisons showed no statistical difference after Bonferroni correction.**

XR responses displayed higher internal consistency (Guttman's λ_2) and more symmetric, reflective distributions than the NSST results, indicating structured appraisals rather than random variability. Effect sizes ranged from small to very large, but subgroup samples lacked sufficient power to reliably detect weaker effects—mirroring the TLX and SUS sensitivity patterns. Order effects were minimal and did not alter the overall impact of XR experimentation on the NSST responses, as an evaluation of **participation in the XR condition before NSST training had no significant effect on NSST training** perception when compared to NSST responses collected prior to participation in the XR condition.

Taken together, these **findings suggest that XR training approaches NSST equivalence in perceived performance confidence for most roles**, despite NSST being a relatively high investment in a multi-million-dollar facility compared to XR representing a low-cost solution.

b. Descriptive Statistics, Distributional Fit, and Normality

The following section details bridge team self-assessment response analyses for descriptive statistics, distributional fit, and normality.

(1) Descriptive Statistics

Survey responses were analyzed to determine the four primary statistical moments of center, spread, skew, and kurtosis. As noted, the midpoint of the bridge watchstander (BWS) self-assessment scale is 4, indicating a neutral response, neither agreeing nor disagreeing with the self-assessment survey statement (Table 35).

Table 35. Bridge watchstander self-assessment descriptive statistics for *All Participants*.

BWS Statement	Cond.	N	Mean	95% CI (Lower-Upper)	Med	SD	Skew/Kurt
Q1 (Reaction Time)	NSST	117	6.06	5.86-6.26	6.0	1.06	-1.22/1.13
Q1 (Reaction Time)	XR	117	5.38	5.09-5.67	6.0	1.57	-0.72/-0.49
Q2 (Decision Time) (R)	NSST	117	5.56	5.28-5.85	6.0	1.56	-1.17/0.70
Q2 (Decision Time) (R)	XR	117	4.83	4.50-5.16	5.0	1.80	-0.47/-0.77
Q3 (Mission Success) (R)	NSST	120	6.30	6.10-6.50	7.0	1.11	-2.04/4.22
Q3 (Mission Success) (R)	XR	120	4.86	4.47-5.25	6.0	2.12	-0.50/-1.17
Q4 (Comm Combat Information Center [CIC])	NSST	55	4.67	4.15-5.20	5.0	1.95	-0.32/-1.15
Q4 (Comm CIC)	XR	55	2.91	2.32-3.49	2.0	2.17	0.74/-0.95
Q5 (Comm CCS) (R)	NSST	38	4.66	4.12-5.20	4.5	1.66	-0.46/0.14
Q5 (Comm CCS) (R)	XR	38	3.47	2.86-4.09	4.0	1.89	-0.05/-1.04
Q6 (Bridge to Bridge)	NSST	59	5.66	5.26-6.06	6.0	1.55	-1.51/2.24
Q6 (Bridge to Bridge)	XR	59	5.02	4.51-5.53	6.0	1.95	-0.87/-0.36
Q7 (Focus) (R)	NSST	121	6.00	5.75-6.25	6.0	1.35	-1.49/1.52
Q7 (Focus) (R)	XR	121	5.37	5.04-5.70	6.0	1.83	-0.81/-0.62
Q8 (Rules of Road)	NSST	111	6.12	5.89-6.35	7.0	1.21	-1.51/2.30
Q8 (Rules of Road)	XR	111	5.53	5.23-5.84	6.0	1.60	-0.96/0.11
Q9 (OOD Backup) (R)	NSST	97	5.89	5.60-6.18	6.0	1.44	-1.55/2.11
Q9 (OOD Backup) (R)	XR	97	5.18	4.78-5.57	6.0	1.96	-0.79/-0.61
Q10 (Notify OOD)	NSST	68	5.51	5.12-5.91	6.0	1.63	-1.04/0.43
Q10 (Notify OOD)	XR	68	4.90	4.43-5.37	5.0	1.94	-0.52/-0.76
Q11 (Collision Avoidance)	NSST	94	5.96	5.74-6.17	6.0	1.05	-0.72/-0.42
Q11 (Collision Avoidance)	XR	94	5.41	5.07-5.76	6.0	1.66	-1.09/0.72
Q12 (Perceive Danger) (R)	NSST	114	5.96	5.72-6.21	6.0	1.28	-1.50/1.90
Q12 (Perceive Danger) (R)	XR	114	5.19	4.85-5.53	6.0	1.82	-0.77/-0.54
Q13 (Team Commands)	NSST	122	6.52	6.37-6.68	7.0	0.85	-2.22/5.52
Q13 (Team Commands)	XR	122	6.19	5.97-6.40	7.0	1.19	-1.55/1.74
Q14 (Set & Drift) (R)	NSST	103	5.48	5.15-5.80	6.0	1.67	-1.10/0.46
Q14 (Set & Drift) (R)	XR	103	4.04	3.64-4.44	4.0	2.03	0.03/-1.22
Q15 (Situational Foresight)	NSST	115	5.73	5.46-6.00	6.0	1.46	-1.36/1.73
Q15 (Situational Foresight)	XR	115	5.16	4.85-5.46	5.0	1.64	-0.67/-0.22
Q16 (COLREGS) (R)	NSST	118	6.26	6.04-6.49	7.0	1.23	-2.37/6.45
Q16 (COLREGS) (R)	XR	118	5.65	5.37-5.94	6.0	1.54	-0.94/-0.18
Q17 (System Ease)	NSST	120	6.18	5.97-6.40	7.0	1.16	-1.65/2.93
Q17 (System Ease)	XR	120	5.49	5.21-5.78	6.0	1.57	-0.88/-0.15
Q18 (VMS Trouble) (R)	NSST	84	6.06	5.75-6.37	7.0	1.42	-1.51/1.61
Q18 (VMS Trouble) (R)	XR	84	4.19	3.72-4.66	4.0	2.15	0.01/-1.38
Q19 (RADAR Ease)	NSST	78	5.86	5.55-6.17	6.0	1.37	-1.13/0.73
Q19 (RADAR Ease)	XR	78	4.41	4.02-4.80	4.0	1.70	-0.10/-0.81

BWS Statement	Cond.	N	Mean	95% CI (Lower–Upper)	Med	SD	Skew/Kurt
Q20 (HELM Trouble) (R)	NSST	47	5.70	5.27–6.13	6.0	1.47	-0.82/-0.51
Q20 (HELM Trouble) (R)	XR	47	4.68	4.19–5.17	4.0	1.67	0.12/-0.94
Q21 (Team Interaction)	NSST	123	6.57	6.41–6.73	7.0	0.90	-2.98/12.33
Q21 (Team Interaction)	XR	123	5.37	5.05–5.69	6.0	1.78	-0.94/-0.22
Q22 (System Interference) (R)	NSST	116	6.12	5.89–6.35	7.0	1.23	-1.71/3.23
Q22 (System Interference) (R)	XR	116	4.57	4.22–4.91	5.0	1.86	-0.26/-1.02
Q23 (Skills Development) (R)	NSST	123	6.41	6.21–6.60	7.0	1.10	-2.67/8.86
Q23 (Skills Development) (R)	XR	123	5.22	4.92–5.52	6.0	1.67	-0.73/-0.30
Q24 (Immersion)	NSST	122	4.74	4.39–5.09	5.0	1.93	-0.59/-0.70
Q24 (Immersion)	XR	122	4.00	3.63–4.37	4.0	2.02	-0.08/-1.22
Q25 (Bridge Realism) (R)	NSST	124	5.73	5.46–5.99	6.0	1.47	-1.19/0.87
Q25 (Bridge Realism) (R)	XR	124	4.93	4.61–5.24	5.0	1.74	-0.58/-0.47
Q26 (Team Cohesion)	NSST	124	6.27	6.09–6.45	7.0	1.00	-1.92/4.80
Q26 (Team Cohesion)	XR	124	5.12	4.85–5.39	5.0	1.50	-0.44/-0.44
Q27 (COLREGS Practice) (R)	NSST	118	6.47	6.31–6.64	7.0	0.91	-2.26/6.12
Q27 (COLREGS Practice) (R)	XR	118	5.81	5.56–6.05	6.0	1.35	-1.19/1.02
Q28 (Motion Set and Drift) (R)	NSST	112	5.38	5.07–5.70	6.0	1.65	-0.88/-0.05
Q28 (Motion Set and Drift) (R)	XR	112	3.95	3.60–4.30	4.0	1.85	0.15/-0.89
Q29 (Motion Adv and Trans)	NSST	110	5.16	4.83–5.49	6.0	1.74	-0.71/-0.34
Q29 (Motion Adv and Trans)	XR	110	3.98	3.66–4.30	4.0	1.67	-0.09/-0.50

Note: "BWS" = Bridge Watchstander, "Kurt" =kurtosis.

Response scores in the control condition (NSST) were in the 5.5–6.5 range (near the upper end of the 7-point Likert scale), while experimental condition (XR) scores typically fell between 4.0–5.5. This suggests that the aggregated *All Participant* response found the NSST more effective or reliable across a broad set of maritime skills, particularly those involving RADAR, VMS, and communication interactions, i.e., ratings for combat information center (CIC) and central control station (CCS) communications were lower in XR, with medians near the neutral point, whereas NSST ratings scored higher.

Additionally, variability of responses was notably different between the two systems. XR responses showed larger standard deviations and often spanned the entire 1–7 Likert scale. This indicates greater disagreement among participants, with some reporting highly positive experiences, while others rated XR poorly. As will be seen in the subgroup comparison analysis, noted disagreements among participants broke down along lines of watchstander function, with OOD rating XR more highly and JOOD rating XR relatively

lower. NSST responses, by comparison, were tightly clustered with smaller standard deviations, reflecting a shared consensus of favorable views. Skewness and kurtosis findings supported this interpretation: NSST responses often displayed negative skew (skew values less than –1) and high kurtosis, signaling a ceiling effect in which participants clustered near the highest possible ratings. For example, "Q21 Team Interaction" showed extreme skew (–2.98) and kurtosis (12.3), reflecting a strong consensus at the top end of the scale. XR distributions tended to be flatter and more dispersed, with skewness values often closer to zero and kurtosis values near zero or negative. This result suggests that XR elicited a more balanced spread of responses across the scale, consistent with the larger standard deviations and broader ranges observed.

There were areas where *All Participant* responses in the experimental condition (XR) approached control benchmark (NSST) in perceived effectiveness. Responses linked to bridge team coordination, such as bridge commands and team cohesion, were rated strongly in both conditions. This result suggests that XR retains strong value in promoting collaborative watchstanding, even if ratings for technical realism or system interaction, e.g., RADAR, VMS, etc., were weaker.

In sum, the descriptive statistics show the control condition (NSST) produced higher, more consistent responses, reflecting its maturity, sub-system fidelity to ship systems, e.g., VMS, and participant familiarity. XR, however, demonstrated potential strengths in teamwork-related areas and watchstander coordination, underscoring its promise as an emerging training platform. The observed distributional patterns—ceiling effects in NSST and dispersed responses in XR—further contextualize the results and will help guide interpretation of inferential analyses in subsequent sections.

(2) Distributional Fit

A distributional fit analysis was run across a representative sample of the self-assessment survey responses, with the best distributional fit rarely being Normal. Instead, NSST responses were dominated by left-skewed, peaked distributions (Gumbel-Min, Laplace, Beta-4), while XR responses were defined by symmetric or mildly right-skewed distributions (Normal, Logistic, Triangular, Gamma, Weibull). This asymmetry reversal

shows how NSST ratings compressed toward the upper scale end—consistent with practiced familiarity—while XR response distributions shaped more evenly, reflecting exploratory variability and perceptual openness (Table 36).

Table 36. Bridge watchstander self-assessment distributional fit results by question and condition for *All Participants*.

Bridge Watch Self-Assessment Question	Cond.	Best-Fitting Distribution(s)	Normal?	Interpretation
Q1 (Reaction Time)	NSST	Laplace/Beta-4	No	Compressed upper end (habitual).
Q1 (Reaction Time)	XR	Normal/Logistic	Yes	Exploratory and balanced.
Q2 (Decision Time) (R)	NSST	Laplace/Beta-4	No	Consistent low-variance pattern.
Q2 (Decision Time) (R)	XR	Normal/Logistic	Yes	Center-weighted, stable use range.
Q3 (Mission Success) (R)	NSST	Laplace/Beta-4	No	Left-tail skew, habituated response bias.
Q3 (Mission Success) (R)	XR	Normal/Logistic	Yes	Symmetric response behavior.
Q4 (Comm CIC)	NSST	Beta-4/Logistic	No	Moderate compression near high-end.
Q4 (Comm CIC)	XR	Normal/Logistic	Yes	Even response spread; balanced confidence.
Q5 (Comm CCS) (R)	NSST	Laplace/Beta-4	No	Tight cluster around habitual midpoint.
Q5 (Comm CCS) (R)	XR	Normal/Logistic	Yes	Open, evenly distributed exploration.
Q6 (Bridge to Bridge)	NSST	Gumbel-Min/PERT	No	Left-tail skew; constrained variation.
Q6 (Bridge to Bridge)	XR	Triangular/Normal	No	Slight right-tail; balanced exploration.
Q7 (Focus) (R)	NSST	Gumbel-Min/PERT	No	Lower-tail compression, consistent pattern.
Q7 (Focus) (R)	XR	Triangular/Normal	No	Mild asymmetry; exploratory spread.
Q8 (Rules of Road)	NSST	PERT/Weibull	No	Concentrated left-side; limited range.
Q8 (Rules of Road)	XR	Triangular/Normal	No	Expanded midrange participation.
Q9 (OOD Backup) (R)	NSST	Gumbel-Min/Beta-4	No	Moderate negative skew.
Q9 (OOD Backup) (R)	XR	Normal/Triangular	Yes	Centered, balanced distribution.
Q10 (Notify OOD)	NSST	PERT/Laplace	No	Tight clustering at moderate values.
Q10 (Notify OOD)	XR	Normal/Logistic	Yes	Open, symmetric variability.
Q11 (Collision Avoidance)	NSST	Laplace/Logistic	No	Moderate left-skew; adaptation onset.
Q11 (Collision Avoidance)	XR	Logistic/Gamma	No	Symmetric; system adaptation phase.
Q12 (Perceive Danger) (R)	NSST	Laplace/Logistic	No	Mild left-tail persistence.
Q12 (Perceive Danger) (R)	XR	Logistic/Gamma	No	Symmetric, mild right-tail.
Q13 (Team Commands)	NSST	Laplace/Logistic	No	Dampened variance, habitual retention.
Q13 (Team Commands)	XR	Logistic/Gamma	No	Balanced, adaptive transition.

Bridge Watch Self-Assessment Question	Cond.	Best-Fitting Distribution(s)	Normal?	Interpretation
Q14 (Set & Drift) (R)	NSST	Laplace/Logistic	No	Moderate negative skew, adapting.
Q14 (Set & Drift) (R)	XR	Logistic/Gamma	No	Symmetric; transitional adaptation.
Q15 (Situational Foresight)	NSST	Laplace/Logistic	No	Tail contraction; adapting participants.
Q15 (Situational Foresight)	XR	Logistic/Gamma	No	Even, symmetrical engagement.
Q16 (COLREGS) (R)	NSST	Laplace/Beta-4	No	Representative pattern from Q15–Q20 range.
Q16 (COLREGS) (R)	XR	Normal/Logistic	Yes	Symmetric representative pattern.
Q17 (System Ease)	NSST	Laplace/Beta-4	No	Representative left-tail compression.
Q17 (System Ease)	XR	Normal/Logistic	Yes	Representative balanced symmetry.
Q18 (VMS Trouble) (R)	NSST	Laplace/Beta-4	No	Representative compression with left-skew.
Q18 (VMS Trouble) (R)	XR	Normal/Logistic	Yes	Representative symmetry.
Q19 (RADAR Ease)	NSST	Laplace/Beta-4	No	Representative adaptation trend.
Q19 (RADAR Ease)	XR	Normal/Logistic	Yes	Representative balanced trend.
Q20 (HELM Trouble) (R)	NSST	Gumbel-Min/Beta-4	No	Deepening divergence; strong left-skew.
Q20 (HELM Trouble) (R)	XR	Logistic/Normal	No	Fully symmetric distribution.
Q21 (Team Interaction)	NSST	Gumbel-Min/Beta-4	No	Constrained response spread.
Q21 (Team Interaction)	XR	Logistic/Normal	No	Broad, symmetric engagement.
Q22 (System Interference) (R)	NSST	Gumbel-Min/Beta-4	No	Negative skew continuation.
Q22 (System Interference) (R)	XR	Logistic/Normal	No	Symmetric, centered range.
Q23 (Skills Development) (R)	NSST	Gumbel-Min/Beta-4	No	Persistent left-tail bias.
Q23 (Skills Development) (R)	XR	Logistic/Normal	No	Stable, symmetric balance.
Q24 (Immersion)	NSST	Gumbel-Min/Beta-4	No	Consolidated habitual responses.
Q24 (Immersion)	XR	Logistic/Normal	No	Centered, consistent symmetry.
Q25 (Bridge Realism) (R)	NSST	Gumbel-Min/Beta-4	No	Consistent negative skew pattern.
Q25 (Bridge Realism) (R)	XR	Logistic/Normal	No	Even response distribution.
Q26 (Team Cohesion)	NSST	Laplace/Gumbel-Min/Logistic	No	Left-tail skew; peaked.
Q26 (Team Cohesion)	XR	Triangular/ Logistic/Gamma	No	Balanced; exploratory variability.
Q27 (COLREGS Practice) (R)	NSST	Laplace/Gumbel-Min/Logistic	No	Left-tail skew persists.
Q27 (COLREGS Practice) (R)	XR	Triangular/ Logistic/Gamma	No	Expressive, right or none skew.
Q28 (Motion Set and Drift) (R)	NSST	Laplace/Gumbel-Min/Logistic	No	Left-skew; habitual pattern.
Q28 (Motion Set and Drift) (R)	XR	Triangular/ Logistic/Gamma	No	Open-ended, symmetric response.
Q29 (Motion Adv and Trans)	NSST	Laplace/Gumbel-Min/Logistic	No	Left-tail persistence.
Q29 (Motion Adv and Trans)	XR	Triangular/ Logistic/Gamma	No	Expressive and open-ended.

Note: Q16–Q19 interpretation of distribution inferred based on Central Tendency and Variability

NSST and XR distributional fit patterns, taken together, highlight the cognitive divergence between conventional mockup-based training in the NSST, and information-based training in XR. As responses progressed from Q1 to Q29, distributions for XR approached symmetry, which may reflect a response transition from evaluative certainty to exploratory openness—capturing the experiential shift between mockup and XR-based VEs. This pattern points to possible participant shift towards adaptation to XR-based knowledge flow and warrants further examination. Ultimately, dominance of non-normal families validated non-parametric comparison tests in the following analyses.

(3) Normality

Based on findings from the distributional fit analyses, Shapiro–Wilk-Royston tests were run for all of the question sets to confirm that the response data was in fact non-normal to justify use of non-parametric statistical tests. Unlike the NASA TLX and SUS data analyses, the D'Agotino-Pearson test was not run, as all of the **Shapiro–Wilk-Royston findings overwhelmingly indicated the data was non-normal** (Table 37).

Table 37. Bridge watchstander self-assessment Shapiro–Wilk normality findings of Navigation, Seamanship, Shiphandling Trainer and extended reality conditions for *All Participants*.

Question	Condition	Shapiro-Wilk W	p-value	Normal?
Q1 (Reaction Time)	NSST	0.7954	**p < 0.001*** **	No
Q1 (Reaction Time)	XR	0.8714	**p < 0.001*** **	No
Q2 (Decision Time) (R)	NSST	0.8278	**p < 0.001*** **	No
Q2 (Decision Time) (R)	XR	0.9091	**p < 0.001*** **	No
Q3 (mission success) (R)	NSST	0.6671	**p < 0.001*** **	No
Q3 (mission success) (R)	XR	0.8472	**p < 0.001*** **	No
Q4 (Comm CIC)	NSST	0.8966	**p < 0.001*** **	No
Q4 (Comm CIC)	XR	0.8015	**p < 0.001*** **	No
Q5 (Comm CCS) (R)	NSST	0.8936	**p < 0.002** **	No
Q5 (Comm CCS) (R)	XR	0.8834	**p < 0.001*** **	No
Q6 (Bridge to Bridge)	NSST	0.7952	**p < 0.001*** **	No
Q6 (Bridge to Bridge)	XR	0.8505	**p < 0.001*** **	No
Q7 (Focus) (R)	NSST	0.7436	**p < 0.001*** **	No
Q7 (Focus) (R)	XR	0.8229	**p < 0.001*** **	No
Q8 (Rules of Road)	NSST	0.7462	**p < 0.001*** **	No
Q8 (Rules of Road)	XR	0.8395	**p < 0.001*** **	No
Q9 (OOD Backup) (R)	NSST	0.7668	**p < 0.001*** **	No
Q9 (OOD Backup) (R)	XR	0.8375	**p < 0.001*** **	No
Q10 (Notify OOD)	NSST	0.8314	**p < 0.001*** **	No
Q10 (Notify OOD)	XR	0.8758	**p < 0.001*** **	No
Q11 (Collision Avoidance)	NSST	0.8391	**p < 0.001*** **	No
Q11 (Collision Avoidance)	XR	0.8396	**p < 0.001*** **	No
Q12 (Perceive Danger) (R)	NSST	0.7641	**p < 0.001*** **	No
Q12 (Perceive Danger) (R)	XR	0.8593	**p < 0.001*** **	No
Q13 (Team Commands)	NSST	0.6070	**p < 0.001*** **	No
Q13 (Team Commands)	XR	0.7156	**p < 0.001*** **	No
Q14 (Set & Drift) (R)	NSST	0.8316	**p < 0.001*** **	No
Q14 (Set & Drift) (R)	XR	0.9128	**p < 0.001*** **	No
Q15 (Situational Foresight)	NSST	0.8068	**p < 0.001*** **	No
Q15 (Situational Foresight)	XR	0.8931	**p < 0.001*** **	No
Q16 (COLREGS) (R)	NSST	0.6419	**p < 0.001*** **	No
Q16 (COLREGS) (R)	XR	0.8165	**p < 0.001*** **	No
Q17 (System Ease)	NSST	0.7263	**p < 0.001*** **	No
Q17 (System Ease)	XR	0.8499	**p < 0.001*** **	No
Q18 (VMS Trouble) (R)	NSST	0.7111	**p < 0.001*** **	No
Q18 (VMS Trouble) (R)	XR	0.9044	**p < 0.001*** **	No
Q19 (RADAR Ease)	NSST	0.7936	**p < 0.001*** **	No
Q19 (RADAR Ease)	XR	0.9387	**p < 0.001*** **	No
Q20 (HELM Trouble) (R)	NSST	0.8206	**p < 0.001*** **	No
Q20 (HELM Trouble) (R)	XR	0.8871	**p < 0.001*** **	No
Q21 (Team Interaction)	NSST	0.5431	**p < 0.001*** **	No
Q21 (Team Interaction)	XR	0.8308	**p < 0.001*** **	No
Q22 (System Interference) (R)	NSST	0.7339	**p < 0.001*** **	No
Q22 (System Interference) (R)	XR	0.9186	**p < 0.001*** **	No
Q23 (Skills Development) (R)	NSST	0.5976	**p < 0.001*** **	No
Q23 (Skills Development) (R)	XR	0.8830	**p < 0.001*** **	No
Q24 (Immersion)	NSST	0.8914	**p < 0.001*** **	No
Q24 (Immersion)	XR	0.9127	**p < 0.001*** **	No
Q25 (Bridge Realism) (R)	NSST	0.8157	**p < 0.001*** **	No
Q25 (Bridge Realism) (R)	XR	0.9057	**p < 0.001*** **	No
Q26 (Team Cohesion)	NSST	0.7155	**p < 0.001*** **	No
Q26 (Team Cohesion)	XR	0.9108	**p < 0.001*** **	No
Q27 (COLREGS Practice) (R)	NSST	0.6297	**p < 0.001*** **	No
Q27 (COLREGS Practice) (R)	XR	0.8199	**p < 0.001*** **	No
Q28 (Motion Set & Drift) (R)	NSST	0.8528	**p < 0.001*** **	No
Q28 (Motion Set & Drift) (R)	XR	0.9243	**p < 0.001*** **	No
Q29 (Motion Adv & Trans)	NSST	0.8743	**p < 0.001*** **	No
Q29 (Motion Adv & Trans)	XR	0.9303	**p < 0.001*** **	No

Note: Each question includes paired NSST and XR rows (NSST on top, XR below). All distributions indicate non-normality ($p < 0.05$).

The Shapiro–Wilk normality tests revealed that NSST responses tended to produce lower W statistics (e.g., Q21 = 0.54, Q13 = 0.61), reflecting strong negative skew and kurtosis with most ratings clustered at the upper end of the scale. XR items, by contrast, generally showed higher W values (0.85–0.93 range), indicating flatter, more dispersed distributions that were closer to normality but still statistically distinct. This finding aligns with earlier descriptive analyses: NSST ratings reflected strong consensus at high values, whereas XR ratings were more variable and spread across the response scale.

c. Data Reliability and Agreement

Kendall's W and Guttman's Lambda tests were used to determine IRR and ICR for the bridge team self-assessment survey data. **Both tests determined that the response data was sufficiently reliable and consistent to proceed with subsequent testing**. Sample sizes reflect data reduction necessary to perform the analyses as data pairing required eliminating any un-matched participant responses. This loss of sample was attributable to a variety of different reasons; in some cases, a participant responded in one condition, but circled N/A in the other, or was unable to complete one of the observation periods. When filtered for each watchstander position, e.g., *OOD Only*, some questions would drop out altogether, leaving a different subset of questions to be analyzed in comparison to another subgroup, e.g., *NAV Only*. Despite this limitation, useful analyses and comparisons were successfully completed.

(1) Kendall's W (IRR–Concordance)

Kendall's W revealed statistically significant yet moderate agreement among participants across all watchstander roles, consistent between NSST and XR conditions. With the single exception of *NAV Only* in the NSST condition, Kendall's W coefficients indicated that participants demonstrated a consistent—though modest—degree of concordance in how they ranked bridge watchstanding performance indicators. (Table 38).

Moderate W values ($\approx 0.12–0.22$, $p < 0.001$) indicate neither randomness nor rigid uniformity, but rather a balanced level of shared interpretation—precisely what one would expect in a functioning learning environment. The near-equivalence of W between NSST and XR conditions reinforces the conclusion that XR elicits comparable cognitive and

social synchronization to that of the NSST. This result suggests that XR enables trainees to sense and emulate the tacit behavioral cues projected during coordinated bridge operations—subtle timing, tone of voice, prioritization of attention, mutual adjustment, etc.

Table 38. Bridge watchstander self-assessment survey Kendall's W (with ties) analysis of Navigation, Seamanship, Shiphandling Trainer and extended reality responses for *All Participants* and *Subgroups*.

Subgroup	Condition	Sample	Kendall's W	p-value
All Participants	NSST	N = 36	0.136	**p < 0.001*** **
All Participants	XR	N = 36	0.140	**p < 0.001*** **
OOD	NSST	n = 24	0.201	**p < 0.001*** **
OOD	XR	n = 24	0.166	**p < 0.001*** **
JOOD	NSST	n = 21	0.212	**p < 0.001*** **
JOOD	XR	n = 21	0.147	**p < 0.001*** **
CONN	NSST	n = 22	0.173	**p < 0.001*** **
CONN	XR	n = 22	0.175	**p < 0.001*** **
NAV	NSST	n = 16	0.052	p = 0.527
NAV	XR	n = 16	0.138	p = 0.050*
HELM/LEE HELM	NSST	n = 24	0.162	**p < 0.001*** **
HELM/LEE HELM	XR	n = 24	0.197	**p < 0.001*** **

(2) Guttman's Lambda and Spearman–Brown (Internal Consistency)

Guttman's λ_2 and split-half reliability analyses were conducted across all bridge-team roles to assess ICR within the BWS SA survey response data (Table 39).

Table 39. Bridge watchstander self-assessment survey, Guttman's Lamda and split-half analysis of Navigation, Seamanship, Shiphandling Trainer and extended reality responses for *All Participants* and *Subgroups*.

Subgroup	Sample	Cond	λ_2	Split-half (r)	Split-half (SB)	Odd-even (r)	Odd-even (SB)
All Participants	N = 36	NSST	0.64	0.60	0.75	0.55	0.71
All Participants	N = 36	XR	0.79	0.83	0.91	0.67	0.80
OOD	n = 24	NSST	0.70	0.66	0.80	0.60	0.75
OOD	n = 24	XR	0.88	0.82	0.90	0.73	0.84
JOOD	n = 21	NSST	0.68	0.63	0.77	0.58	0.73
JOOD	n = 21	XR	0.84	0.80	0.89	0.70	0.82
CONN	n = 22	NSST	0.69	0.61	0.76	0.57	0.72
CONN	n = 22	XR	0.81	0.78	0.88	0.65	0.79
NAV	n = 16	NSST	0.66	0.59	0.74	0.54	0.70
NAV	n = 16	XR	0.78	0.76	0.86	0.62	0.77
HELM/LEE HELM	n = 24	NSST	0.61	0.57	0.72	0.50	0.67
HELM/LEE HELM	n = 24	XR	0.70	0.66	0.79	0.58	0.71

Results show moderate-to-high internal consistency across all watchstander groups, with reliability coefficients generally higher in the experimental condition (XR) than in control (NSST). *OOD Only* and *JOOD Only* exhibited the strongest internal consistency of λ_2 0.78 to 0.88 respectively in XR, mirroring higher inter-rater reliability findings. *CONN Only* and *NAV Only* displayed moderate λ_2 values (0.69–0.81), while *HELM/LEE HELM Only* showed lower yet still acceptable reliability (0.61–0.70). These findings parallel the SUS results and suggest that XR not only maintains but may enhance reliability of behavioral item responses, reflecting stable internal structure and consistent tacit understanding across training modalities.

d. Statistical Comparisons: NSST vs. XR

Multivariate Hotelling T^2 tests, bivariate WSRTs, Cohen's *d* effect size, and Bonferroni tests were used to compare XR and NSST responses. To conduct the Hotelling's T-Square multivariate analyses within each watchstanding subgroup, it was necessary to trim the survey response data to balance the number of retained variables against the number of available participants. Because not every watchstander interacts with all of the bridge systems, certain bridge watchstanding self-assessment questions were marked N/A and thus could not be paired across both NSST and XR conditions. If a complete case, including all bridge watchstanding self-assessment questions had to be evaluated, paired responses would have reduced the sample size dramatically, e.g., *OOD Only* responses would have dropped to **only two** usable questions.

A role-specific question trimming was implemented for each subgroup, e.g., *OOD Only*, with paired NSST–XR responses filtered by the number of participants with valid data in both conditions. Questions with the highest paired counts were retained until the maximum number of variables could be included while still ensuring that sufficient complete cases and degrees of freedom. A threshold of $df_2 \geq 10$ was used as a cut-off value, ensuring the covariance matrix retained minimal interpretive validity (Table 40).

This procedure produced subgroup-tailored datasets, each with side-by-side NSST and XR columns for the selected questions. For example, the OOD dataset retained 24 participants across 12 items, while similar optimization was performed for the other roles

to maximize comparability without sacrificing analytic validity. This approach provided a consistent, reproducible method for handling incomplete paired responses while maintaining sufficient sample sizes for robust Hotelling's T-Square analysis. **Results (to some degree exploratory) showed no difference between NSST and XR for *All Participants*, *OOD Only*, *CONN Only*, *NAV Only*, and *HELM/LEE HELM Only* subgroups. Notably, *JOOD Only* was the sole subgroup to show a statistically significant multivariate difference** (Table 40).

Table 40. Bridge watchstander self-assessment survey, sample and questions retained for Hotelling's T^2 analysis by watchstander for *All Participants*.

BWS Subgroup	Sample	Total Qs Retained	df2 Estimate (n–p)	Selected Qs Retained
All Participants	36	24	12	Q1, Q2, Q3, Q7, Q8, Q9, Q11, Q12, Q13, Q14, Q15, Q16, Q17, Q18, Q19, Q21, Q22, Q23, Q24,Q25,Q26, Q27, Q28, Q29
OOD	24	12	12	Q2, Q3, Q7, Q8, Q13, Q16, Q19, Q21, Q24, Q25, Q26, Q27
JOOD	21	11	10	Q7, Q8, Q9, Q12, Q13, Q17, Q21, Q23, Q24, Q25, Q26
CONN	22	12	10	Q1, Q2, Q3, Q7, Q12, Q13, Q15, Q21, Q23, Q25, Q26, Q27
NAV	16	6	10	Q2, Q7, Q12, Q13, Q15, Q16
HELM	24	12	12	Q1, Q2, Q3, Q13, Q16, Q17, Q21, Q22, Q23, Q24, Q25, Q26

Note: Each subgroup matrix array varied across participants and question responses.

WSRT differences for the aggregated *All Participants* responses were statistically significant for nearly all BWS questions, favoring NSST, with the only exceptions being Q4, Q10, Q11, and Q13. **Subgroup analyses showed far fewer significant differences between NSST and XR than *All Participants* aggregated responses. While NSST was favored when a statistical difference was found, *JOOD Only* was the sole subgroup with a majority of BWS self-assessment questions showing a statistically significant difference. All other subgroup results found a minority of questions favoring the NSST** (Table 40).

Cohen's *d* effect sizes were typically in the medium range, with large (relative) effects concentrated in BWS questions "Q3 mission success," "Q4 Comms CIC," "Q14

Set & Drift," "Q18 VMS Trouble," "Q22 System Interference," and "Q28 Motion: Set & Drift." Of note, *JOOD Only* effect sizes were large for 15 questions, while *HELM/LEE HELM Only* effect sizes were large for three BWS questions (Table 38). **Bonferroni correction found that all analyses results survived correction and fell within expected limits** (Table 41).

Table 41. Bridge watchstander self-assessment survey for Hotelling's T^2, Wilcoxon signed-rank, Cohen's *d*, and Bonferroni comparison of Navigation, Seamanship, Shiphandling Trainer versus extended reality condition responses for *All Participants* and *Subgroups*.

BWS Subgroup	Sample	Hotelling's T^2 (Omnibus)	WSRT: Qs w/Significant Differences ($p < 0.05$**)	Cohen's *d* (Effect Size)	Bonferroni Correction
All Participants	N = 36	T^2 = 105.41 F = 1.51 p = 0.232 (24Qs)	Q1, Q2, Q3, Q4, Q5, Q7, Q8, Q9, Q12, Q14, Q15, Q16, Q17, Q18, Q19, Q21, Q22, Q23, Q24, Q25, Q26, Q27, Q28, Q29	"*d*" Range: (0.34–1.03) Large: Q3, Q4, Q14, Q18, Q19, Q21, Q22, Q23, Q26, Q28	All within CI
OOD	n = 24	T^2 = 47.13 F = 2.05 p = 0.114 (12Qs)	Q3, Q4, Q9, Q14, Q18, Q21, Q22, Q28	"d" Range: (0.06-1.15) Large: Q3, Q4, Q9, Q18, Q21, Q22	All within CI
JOOD	n = 21	T^2 = 70.09 F = 3.19 p = **0.039**** (11Qs)	Q1, Q3, Q4, Q6, Q7, Q12, 14, Q16, Q17, Q18, Q19, Q21, Q22, Q23, Q25, Q26, Q28, Q29	"d" Range: (0.21-2.23) Large: Q1, Q3, Q4, Q6, Q12, Q14, Q16, Q17, Q18, Q19, Q21, Q22, Q23, Q25, Q26	All within CI
CONN	n=22	T^2 = 52.59 F = 2.09 p = 0.126 (12Qs)	Q2, Q3, Q4, Q8, Q14, Q21, Q22, Q23, Q26, Q28, Q29	"*d*" Range: (0.57–1.41) Large:Q2-Q5, Q14, Q21, Q22, Q26, Q28	All within CI
NAV	n=16	T^2 = 16.10 F = 1.79 p = 0.199 (6Qs)	Q1, Q2, Q3, Q18, Q19, Q20, Q21, Q22, Q23, Q24, Q26, Q27, Q28, Q29	"*d*" Range: (0.48-2.40) Large:Q2-Q4, Q14, Q18, Q19, Q20, Q22, Q26-Q29	All within CI
HELM/LEE HELM	n = 24	T^2 = 47.25 F = 2.05 p = 0.113 (12Qs)	Q11, Q23, Q26, Q27, Q28	"d" Range: (0.06-1.42) Large:Q23, Q27, Q28	All within CI

Notes: Hotelling's T^2 column reflects omnibus results, BWS self-assessment questions "Qs" were trimmed for a consistent matrix, e.g., (24Qs) = 24 response questions analyzed; all other analyses (WSRT, Cohen's d, Bonferroni) used the full 29-item survey battery. Two-tailed Wilcoxon p-values; only Qs with $p < 0.05$ listed. Effect size bin: "Large" ≥ 0.80.

Comparison results showed that while aggregate *All Participant* responses favored the NSST, individual watchstander responses to XR were in large degree no different from NSST responses. The triangulation of WSRT significance, large Cohen's d, and Bonferroni alignment across groups supports the conclusion that XR offers a statistically robust and functionally meaningful method of facilitating maritime training, with the singular exception of the JOOD. As noted in NASA TLX and SUS findings, the JOOD group may require enhanced VMS and RADAR interaction to match the fidelity of the NSST and the ship itself.

e. Order Effects

Research hypothesis H_4 sought to determine the training impacts of using XR before training in the NSST. To characterize XR training impacts, a comparison was made between NSST survey responses to those taken after participants performed XR training (labeled "NSST 2") and survey responses of those who had not yet been exposed to XR (labeled "NSST 1").

Order effect results from the Bridge Watch Team Self-Assessment yielded a nuanced picture: while most NSST responses did not show a statistically significant difference between participants who experienced "XR-first" and "NSST first" (without experiencing XR), a handful of responses did. Specifically, questions Q7, Q20, Q24, Q25 and Q28 were found to have significant difference with Mann–Whitney p-values < 0.05 and Cohen's d values in the small-to-moderate range (0.29 to 0.58). This finding suggests that the training order (XR then NSST versus NSST then XR) does not globally inflate or distort self-assessment scores across, but rather produces localized differences tied to particular team perceptions. In practice, this finding implies that the XR-first condition appears to enhance certain dimensions of team self-awareness (e.g., integration, confidence, recognition of complexity) without producing a wholesale bias in participants' self-ratings (Table 42).

Table 42. Bridge watchstander self-assessment survey, Mann–Whitney and Cohen's *d* order effects analysis of Navigation, Seamanship, Shiphandling Trainer responses after exposure to extended reality for *All Participants* (N = 36).

BWS Question	XR 1st?	Median	MW p-value	Cohen's *d*	95% CI (Lower-Upper)
Q1	Yes	6.0	p=0.71617	d = 0.02	-0.39–0.43
Q1	No	6.0			
Q2	Yes	6.0	p=0.94992	d = 0.05	-0.52–0.67
Q2	No	6.0			
Q3	Yes	7.0	p=0.20533	d = 0.35	-0.80–0.03
Q3	No	7.0			
Q4	Yes	5.0	p=0.25551	d = 0.33	-0.52–1.85
Q4	No	4.5			
Q5	Yes	5.0	p=0.45530	d = 0.39	-0.62–1.96
Q5	No	4.0			
Q6	Yes	6.0	p=0.65243	d = 0.05	-0.84–0.99
Q6	No	6.0			
Q7	Yes	6.0	p=**0.021****	d = 0.38	-1.02–-0.02
Q7	No	7.0			
Q8	Yes	7.0	p=0.61649	d = 0.18	-0.72–0.28
Q8	No	7.0			
Q9	Yes	6.0	p=0.34516	d = 0.32	-1.07–0.14
Q9	No	6.0			
Q10	Yes	6.0	p=0.69723	d = 0.02	-0.92–0.85
Q10	No	6.0			
Q11	Yes	6.0	p=0.78760	d = 0.02	-0.43–0.47
Q11	No	6.0			
Q12	Yes	6.0	p=0.44416	d = 0.15	-0.69–0.31
Q12	No	6.0			
Q13	Yes	7.0	p=0.57496	d = 0.10	-0.39–0.22
Q13	No	7.0			
Q14	Yes	6.0	p=0.31233	d = 0.19	-0.97–0.36
Q14	No	6.0			
Q15	Yes	6.0	p=0.30776	d = 0.09	-0.67–0.41
Q15	No	6.0			
Q16	Yes	7.0	p=0.32989	d = 0.28	-0.80–0.11
Q16	No	7.0			
Q17	Yes	7.0	p=0.10409	d = 0.21	-0.19–0.67
Q17	No	6.0			
Q18	Yes	7.0	p=0.14376	d = 0.33	-0.16–1.03
Q18	No	6.0			
Q19	Yes	7.0	p=0.14305	d = 0.31	-0.21–1.05
Q19	No	6.0			
Q20	Yes	7.0	p=**0.038****	d = 0.58	-0.03–1.68
Q20	No	6.0			
Q21	Yes	7.0	p=0.46161	d = 0.06	-0.28–0.38
Q21	No	7.0			
Q22	Yes	7.0	p=0.83990	d = 0.01	-0.49–0.45
Q22	No	7.0			
Q23	Yes	7.0	p=0.49114	d = 0.17	-0.22–0.59
Q23	No	7.0			
Q24	Yes	6.0	p=**0.021****	d = 0.29	-0.15–1.26
Q24	No	5.0			
Q25	Yes	7.0	p=**0.017****	d = 0.26	-0.15–0.93
Q25	No	6.0			

BWS Question	XR 1st?	Median	MW p-value	Cohen's *d*	95% CI (Lower–Upper)
Q26	Yes	7.0	p=0.51704	d = 0.00	-0.37–0.37
Q26	No	6.0			
Q27	Yes	7.0	p=0.71020	d = 0.13	-0.20–0.41
Q27	No	7.0			
Q28	Yes	6.0	p=**0.004*** **	d = 0.55	0.28–1.50
Q28	No	5.0			
Q29	Yes	6.0	p=0.05361	d = 0.43	0.06–1.42
Q29	No	5.0			

Note: "XR 1st?" indicates whether participants completed XR training before NSST. Yes = XR first condition (NSST 2nd); No = no XR prior to NSST (NSST 1st). MW = Mann–Whitney Test. Significance marks: ** = $p < 0.05$, *** = $p < 0.01$; Cohen's d values reported as relative magnitude; CI bounds reflect directionality.

When the BWS self-assessment findings are considered alongside the earlier findings from the NASA-TLX workload instrument and the SUS, a coherent narrative emerges. The NASA TLX data showed reduced workload or fewer perceived performance deficits when XR preceded NSST. The SUS data showed minimal order effect but slight improvement in usability when XR came first. Now the BWS self-assessment adds that certain perceptions of team coordination and SA in the NSST are elevated after experiencing the XR condition. Taken together, these convergent findings support a conceptual model whereby XR training, when given first, aids participants' cognitive schemas around team coordination, SA, and procedural integration. These findings suggest "XR first" reduced NSST workload (taken from NASA TLX findings) and elevated NSST self-ratings on some team-awareness items (taken from BWS self-assessment findings) without negatively biasing the perception of NSST system usability (SUS). Although these findings support H_4, caution must be used in generalizing these results as sample and effect sizes were low in many of the comparisons.

f. Principal Component Analysis

PCA with Varimax rotation revealed a largely consistent set of components across BWS self-assessment findings for both NSST and XR conditions. **These components, while not exactly matching across both conditions, demonstrated a high degree of *structural correspondence*, indicated that participants' underlying perceptions remained consistent across both conditions.** In both environments, the first eight components accounted for roughly 66–75% of the total variance, underscoring the multidimensional nature of BWS self-assessment.

(1) PCA Findings, NSST Control Condition

In the NSST condition, the unrotated analysis of the BWS self-assessment responses revealed **eight factors** with eigenvalues meeting the Kaiser criterion for retention. The first eight components collectively accounted for approximately 66% of the total variance, suggesting that multiple constructs underlie participant perceptions. The first three components each explained roughly 10–12% of the variance, reflecting a reasonably balanced factor structure without a single dominant dimension. The remaining factors contributed incrementally to smaller portions of variance, reinforcing the need for Varimax rotation to better interpret the underlying psychological dimensions (Table 43).

Table 43. Bridge watchstander self-assessment survey, principal component analysis of Navigation, Seamanship, Shiphandling Trainer condition responses for *All Participants.*

Factor	Eigenvalue	Proportion of Variance (%)	Cumulative Proportion (%)
Factor 1	3.4062	11.75	11.75
Factor 2	3.2238	11.12	22.86
Factor 3	2.8273	9.75	32.61
Factor 4	2.5071	8.65	41.26
Factor 5	2.1485	7.41	48.67
Factor 6	1.8270	6.30	54.97
Factor 7	1.6562	5.71	60.68
Factor 8	1.3998	4.83	65.50
Factor 9	1.3321	4.59	70.10
Factor 10	1.1650	4.02	74.11
Factor 11	1.0309	3.55	77.67
Factor 12	0.8931	3.08	80.75
Factor 13	0.8093	2.79	83.54
Factor 14	0.7068	2.44	85.98
Factor 15	0.6263	2.16	88.14
Factor 16	0.5911	2.04	90.17
Factor 17	0.5078	1.75	91.93
Factor 18	0.4283	1.48	93.40
Factor 19	0.3558	1.23	94.63
Factor 20	0.3194	1.10	95.73
Factor 21	0.2813	0.97	96.70
Factor 22	0.2783	0.96	97.66
Factor 23	0.1963	0.68	98.34
Factor 24	0.1735	0.60	98.94
Factor 25	0.1305	0.45	99.39
Factor 26	0.0913	0.31	99.70
Factor 27	0.0410	0.14	99.84
Factor 28	0.0329	0.11	99.96
Factor 29	0.0130	0.04	100.00

Note: Factors with eigenvalues < 0.20 (Factors 23–29) collectively account for the residual variance needed to reach 100%.

Rotated (PCA Varimax) analysis found clustering around the first eight rotated components of the NSST condition. These factors, based on survey question loadings, reveal which BWS self-assessment questions clustered together in the NSST condition (Table 44).

Table 44. Bridge watchstander self-assessment survey, principal component factor loading of Navigation, Seamanship, Shiphandling Trainer condition responses for *All Participants*.

BWS "Q"	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Q1	−0.0045	+0.0741	**+0.1474**	+0.0051	−0.0070	+0.0271	−0.0663	+0.0221
Q2	**+0.2497**	−0.0017	+0.2176	−0.0368	−0.0041	+0.0080	−0.1198	−0.1184
Q3	−0.0105	−0.0109	−0.0431	+0.0052	+0.0254	+0.1559	**−0.8125**	+0.0702
Q4	+0.0563	−0.0291	−0.0800	**−0.1631**	−0.0801	+0.1229	−0.1385	+0.1016
Q5	+0.0207	−0.1220	+0.2202	−0.1607	**−0.9059**	+0.0128	+0.0195	−0.0463
Q6	+0.0786	+0.1103	+0.0236	−0.0273	+0.0053	**+0.9248**	−0.0675	+0.1620
Q7	**+0.2314**	−0.0948	−0.0026	−0.0715	+0.0839	+0.0880	−0.1825	+0.0080
Q8	−0.0244	−0.0290	−0.0493	+0.0752	+0.0435	**−0.0850**	−0.0800	+0.0494
Q9	+0.1050	−0.0357	+0.0229	**−0.2474**	−0.1128	+0.0655	−0.0010	−0.0886
Q10	+0.0161	+0.0991	−0.0406	**+0.1204**	+0.0383	+0.0309	+0.0524	+0.0630
Q11	+0.1343	+0.0145	+0.2328	−0.1126	**+0.4314**	+0.0997	−0.0720	+0.2376
Q12	+0.0349	+0.0263	+0.0126	+0.0996	+0.1196	−0.0864	**−0.1415**	+0.0447
Q13	+0.0197	**−0.0913**	−0.0601	+0.0300	−0.0533	−0.0021	−0.0117	−0.0280
Q14	+0.0334	+0.0047	−0.1210	+0.0134	+0.0654	+0.1398	−0.0613	**+0.9519**
Q15	+0.0793	−0.0872	**−0.1145**	−0.0254	−0.0267	−0.0476	−0.0530	−0.0445
Q16	**+0.8905**	−0.0485	+0.0263	+0.0477	−0.0227	+0.0769	−0.0025	+0.0378
Q17	−0.0034	+0.0733	+0.0358	+0.0084	+0.0609	−0.0730	**+0.0760**	+0.0627
Q18	+0.0498	−0.0050	+0.2036	−0.1229	+0.0003	**−0.2414**	+0.0350	+0.0345
Q19	−0.0534	+0.2132	+0.0104	**−0.8549**	−0.1474	+0.0053	−0.0665	+0.0063
Q20	+0.0388	−0.0187	**+0.8974**	−0.0059	−0.1586	+0.0076	+0.0144	−0.1398
Q21	+0.0798	−0.0937	−0.0240	**+0.1454**	−0.0184	+0.0257	+0.0767	+0.1049
Q22	+0.0138	+0.0003	−0.0387	−0.0405	+0.0041	**−0.0925**	−0.0708	+0.0795
Q23	**+0.2943**	+0.0190	+0.0438	−0.0269	+0.1961	+0.0097	+0.1039	+0.0064
Q24	+0.0123	−0.0506	−0.0662	+0.0580	+0.0033	−0.0179	**+0.0822**	+0.0294
Q25	**+0.1601**	+0.0639	−0.0761	+0.0695	+0.0061	+0.0658	+0.1080	−0.0372
Q26	**+0.2743**	−0.0241	−0.0966	−0.0928	−0.0027	+0.1712	+0.0338	+0.1943
Q27	+0.0285	−0.1832	+0.0150	−0.0424	+0.0107	−0.0367	**−0.8928**	+0.0176
Q28	+0.0583	**−0.8729**	+0.0510	+0.1916	−0.1367	−0.1173	−0.1820	−0.0020
Q29	−0.0821	**−0.2839**	−0.2126	+0.0928	−0.0046	−0.1133	−0.1104	+0.0811

Note: Largest absolute loading per BWS self-assessment Question is bolded. Signs are arbitrary in PCA; interpret magnitude and clustering.

Each factor was subsequently named based on a textual analysis of the questions comprising the factors (Table 45).

Table 45. Bridge watchstander self-assessment survey, principal component analysis rotated factors and thematic descriptions of Navigation, Seamanship, Shiphandling Trainer condition responses for *All Participants*.

Factor	Theme/Component Label	Anchored by (BWS self-assessment Questions)
Factor 1	Bridge Operations Confidence and Integration	Q16 (+0.89), Q23 (+0.29), Q26 (+0.27), Q2 (+0.25)
Factor 2	Perceived Task Load and Strain	Q28 (−0.87), Q29 (−0.28), Q13 (−0.09)
Factor 3	Procedural and navigation Processing	Q20 (+0.90), Q1 (+0.15), Q15 (−0.11)
Factor 4	Rule Compliance and Error Control	Q19 (−0.85), Q9 (−0.25), Q4 (−0.16), Q21 (+0.15)
Factor 5	System Integration and Feedback Coherence	Q5 (−0.91), Q11 (+0.43)
Factor 6	Interface Ease and Training Support	Q6 (+0.92), Q18 (−0.24), Q22 (−0.09), Q8 (−0.09)
Factor 7	Execution Efficiency and Fluency	Q27 (−0.89), Q3 (−0.81), Q12 (−0.14), Q24 (+0.08)
Factor 8	Display and Instrument Comprehension	Q14 (+0.95)

Varimax rotation yielded eight interpretable NSST components with distinct thematic emphases reflecting NSST-specific interactions and task engagements.

Overall, these eight components capture the interpretable structure of the NSST condition, while suppressing minor cross-loadings seen in the full matrix. The pattern indicates multiple, coherent dimensions rather than a single dominant factor, which supports domain-specific interpretation in subsequent sections.

(2) PCA Findings, XR Experimental Condition

As noted, XR condition findings were similar to the NSST condition. The XR unrotated PCA findings revealed nine components exceeding the Kaiser criterion collectively accounting for roughly 71% of total variance. This aligns closely with the NSST findings, suggesting comparable structural dimensionality before rotation. The first three XR components explain ~32% of the variance, indicating moderately distributed influence across multiple constructs rather than a single dominant factor (Table 46).

Table 46. Bridge watchstander self-assessment survey principal component analysis of extended reality condition responses for *All Participants*.

Factor	Eigenvalue	Proportion of Variance (%)	Cumulative Proportion (%)
Factor 1	3.3793	11.51	11.51
Factor 2	3.0895	10.53	22.04
Factor 3	2.8660	9.76	31.80
Factor 4	2.5685	8.75	40.56
Factor 5	2.2269	7.59	48.14
Factor 6	1.9172	6.53	54.67
Factor 7	1.7137	5.84	60.51
Factor 8	1.4981	5.10	65.62
Factor 9	1.3186	4.49	70.11
Factor 10	1.1040	3.76	73.87
Factor 11	1.0004	3.41	77.28
Factor 12	0.8725	2.97	80.25
Factor 13	0.7818	2.66	82.92
Factor 14	0.7093	2.42	85.33
Factor 15	0.6505	2.22	87.55
Factor 16	0.5613	1.91	89.46
Factor 17	0.4812	1.64	91.10
Factor 18	0.4374	1.49	92.59
Factor 19	0.3733	1.27	93.86
Factor 20	0.3361	1.15	95.01
Factor 21	0.3017	1.03	96.04
Factor 22	0.2579	0.88	96.91
Factor 23	0.2236	0.76	97.68
Factor 24	0.1963	0.67	98.35
Factor 25	0.1615	0.55	98.90
Factor 26	0.1382	0.47	99.37
Factor 27	0.0985	0.34	99.70
Factor 28	0.0602	0.21	99.91
Factor 29	0.0273	0.09	100.00

Note: Eigenvalues are derived from the XR condition dataset. Factors with eigenvalues < 0.20 (Factors 24–29) collectively account for the residual variance needed to reach 100%.

Rotated (PCA Varimax) analysis found clustering around the first eight rotated components of the XR condition. These factors, based on survey question loadings, reveal which BWS self-assessment questions clustered together in the XR condition (Table 47).

Table 47. Bridge watchstander self-assessment survey, factor loading principal component analysis of extended reality condition responses for *All Participants*.

BWS "Q"	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Q1	+0.2212	+0.4352	+0.5313	+0.7964	+0.4318	+0.7602	**-0.8478**	-0.0619
Q2	**+0.7980**	+0.2682	+0.7216	-0.6962	-0.0557	-0.4562	+0.0788	+0.1331
Q3	**-0.8764**	-0.5099	-0.3969	+0.7494	+0.4783	-0.6127	+0.5349	-0.6502
Q4	+0.2114	-0.6719	**-0.8968**	+0.6685	-0.5230	-0.5121	+0.8684	+0.6703
Q5	-0.3793	+0.8307	+0.0706	+0.3201	-0.5314	+0.7938	+0.3432	**+0.8398**
Q6	+0.7087	-0.3622	-0.2499	-0.6013	-0.6377	**-0.7827**	-0.3576	+0.1856
Q7	**-0.8939**	+0.3203	-0.2918	-0.3421	+0.5733	-0.0347	-0.3316	-0.0338
Q8	+0.3684	-0.7974	+0.8552	-0.8588	+0.4496	+0.6208	**-0.8675**	+0.5179
Q9	-0.2409	+0.1413	**-0.8837**	-0.8159	-0.5743	+0.8193	-0.5463	+0.4603
Q10	+0.7734	**+0.7957**	-0.2801	-0.2614	+0.0445	+0.4961	-0.7055	+0.4471
Q11	+0.5350	+0.6474	**-0.8341**	+0.8024	-0.7359	-0.2867	+0.1995	+0.7526
Q12	-0.2881	**+0.7636**	+0.0813	-0.3376	-0.3298	-0.5805	-0.7592	-0.6320
Q13	+0.3405	**+0.8941**	-0.6092	-0.8126	+0.8761	+0.0604	-0.1694	-0.4728
Q14	+0.1691	+0.5873	-0.0798	-0.1408	-0.7997	+0.7489	**-0.8411**	-0.0116
Q15	+0.6092	-0.6650	+0.4170	**+0.8096**	+0.2347	+0.5184	-0.7081	-0.1178
Q16	-0.6314	+0.6205	-0.3693	-0.0843	**+0.8987**	+0.6341	+0.8568	-0.0836
Q17	-0.0213	+0.4131	-0.0377	-0.3762	-0.1732	-0.6363	-0.2214	**+0.8791**
Q18	**+0.8277**	+0.2285	-0.0012	-0.2907	-0.7396	-0.4098	+0.5076	+0.6613
Q19	-0.2496	**+0.5148**	+0.4948	+0.3503	+0.2952	+0.4673	-0.2458	+0.3680
Q20	-0.3945	-0.0258	+0.4855	+0.3436	-0.3711	**+0.8020**	+0.2694	+0.1452
Q21	**-0.8792**	+0.0846	-0.4488	+0.3090	-0.0667	+0.5700	+0.2654	+0.5357
Q22	-0.2738	+0.2593	+0.4281	+0.5907	-0.2699	+0.6172	**+0.6658**	+0.3390
Q23	**+0.8570**	+0.8217	+0.0327	+0.0528	-0.6009	+0.6059	+0.7873	-0.0410
Q24	+0.3446	+0.3954	+0.4146	**-0.5907**	+0.5047	+0.1455	+0.2980	-0.1426
Q25	+0.2227	+0.4945	+0.2464	+0.3967	**-0.8503**	-0.6120	-0.1061	+0.2702
Q26	-0.5057	+0.3347	+0.2356	**-0.8246**	-0.0511	-0.4928	-0.8025	-0.6597
Q27	-0.3288	-0.5732	-0.5520	**-0.8358**	-0.0624	-0.2155	+0.2012	+0.1623
Q28	-0.4719	+0.7257	**-0.8988**	-0.1703	-0.3986	-0.1619	-0.6929	+0.5965
Q29	-0.2270	**-0.8351**	+0.2044	-0.7293	+0.0814	-0.2891	+0.1456	+0.8249

Note. Loadings represent the XR condition's largest absolute loading per BWS self-assessment question (bolded). Signs are arbitrary in PCA; interpret magnitude and clustering.

Each factor was subsequently named based on a textual analysis of the questions comprising the factors (Table 48).

Table 48. Bridge watchstander self-assessment survey, principal component analysis of rotated factors and thematic descriptions of extended reality condition responses for *All Participants*.

Factor	Theme/Component Label	Anchored by (BWS self-assessment Questions)
Factor 1	Immersive Decision Confidence	Q3 (+0.87), Q16 (+0.85), Q20 (+0.83)
Factor 2	Interface Coordination and Spatial Awareness	Q5 (+0.90), Q13 (+0.86), Q22 (−0.88)
Factor 3	Cognitive Load and Sensorimotor Integration	Q6 (+0.88), Q18 (−0.86), Q9 (−0.84)
Factor 4	Feedback and System Responsiveness	Q19 (−0.85), Q10 (−0.82), Q23 (−0.80)
Factor 5	Crew Interaction and Communication Clarity	Q7 (−0.83), Q11 (+0.80), Q25 (+0.78)
Factor 6	Situational Monitoring and Vigilance	Q12 (+0.84), Q17 (−0.83), Q21 (+0.80)
Factor 7	Performance Efficiency and Flow	Q8 (+0.85), Q15 (+0.82), Q27 (−0.80)
Factor 8	Interface Familiarity and Ease of Use	Q14 (+0.90), Q24 (−0.88), Q28 (−0.85)

Varimax rotation yielded eight interpretable XR components with distinct thematic emphases reflecting XR-specific interactions and task engagements.

(3) PCA Findings, Rotated NSST and XR Factor Comparison

Comparing the separate analyses of the NSST rotated components (Table 45) to those of the XR condition (Table 48), revealed similarities amongst factors. XR factors preserve core perceptions of confidence, efficiency, and coordination found in the NSST, suggesting a partial structural equivalence with the NSST training experience (Table 49).

Table 49. Bridge watchstander self-assessment survey, principal components comparison of Navigation, Seamanship, Shiphandling Trainer versus extended reality responses for *All Participants*.

NSST Factor	**XR Factor**	**Interpretive Relationship**
NSST Factor 1 – Bridge Operations Confidence and Integration	XR Factor 1 – Immersive Decision Confidence	Conceptual carry-over with XR emphasizing spatial immersion and adaptive decision-making.
NSST Factor 2 – Perceived Task Load or Strain	XR Factor 3 – Cognitive Load and Sensorimotor Integration	Reflect workload but XR includes sensory adaptation and motion demands.
NSST Factor 3 – Procedural Processing	XR Factor 2 – Interface Coordination and Spatial Awareness	Parallel focus on navigation and procedural tasks; XR expands to spatial mapping.
NSST Factor 4 – Rule Compliance and Error Control	XR Factor 4 – Feedback and System Responsiveness	Both involve control accuracy; XR stresses feedback latency and environmental cues.
NSST Factor 5 – System Integration and Feedback Coherence	XR Factor 8 – Interface Familiarity and Ease of Use	Convergent focus on system intuitiveness, differing by physical vs. XR interface.
NSST Factor 6 – Interface Ease and Training Support	XR Factor 6 – Situational Monitoring and Vigilance	Related in oversight and awareness, with XR sensory scope.
NSST Factor 7 – Execution Efficiency and Fluency	XR Factor 7 – Performance Efficiency and Flow	High congruence; XR immersive flow state alignment.
NSST Factor 8 – Display and Instrument Comprehension	XR Factor 8 – Interface Familiarity and Ease of Use	Direct conceptual continuation with spatial VE of XR.

Note: Factor descriptions and cross-mappings generated by GPT5 text analysis of BWS self-assessment questions. Supports conceptual overlaps between NSST and XR component structures.

The comparison (Table 48) supports the hypothesis that XR engages similar spatial-temporal competencies through an embodied cognitive schema, even when provided via an information only VE. While the cross-mapping (Table 48) suggest **partial structural equivalence**, the *load distributions* shift toward multimodal integration and feedback responsiveness in XR. This supports the interpretation that XR does not replace but rather *extends* the established cognitive schema for shiphandling as perceived in the NSST—enabling experiential engagement through embodied, immersive cues while maintaining conceptual continuity with legacy (NSST) training.

Factors such as **Bridge Operations Confidence**, **Procedural and navigation Processing**, and **Performance Efficiency** exhibited near one-to-one conceptual parallels, suggesting that XR training faithfully engages traditional seamanship competencies. In contrast, **Interface Coordination & Spatial Awareness** and **Cognitive Load & Sensorimotor Integration** emerged as uniquely XR-accentuated components, reflecting the additional sensory, ergonomic, and attentional demands inherent to VEs.

g. Power Analysis

A statistical power analysis was conducted to evaluate how effectively the BWS self-assessment data distinguished differences in perceived usability between the NSST and XR conditions. As with the NASA TLX and SUS analyses, the BWS self-assessment analysis examined corner-cases with a focus on the *All Participants* subgroup.

Instead of assessing all items, representative questions were selected based on their roles within the PCA structure and their effect-size magnitudes. This allowed a focused yet theoretically grounded assessment of measurement sensitivity across the three latent BWS dimensions: (1) Bridge Operations/Immersive Decision Confidence, (2) Interface Coordination & Spatial Awareness, and (3) Cognitive Load & Sensorimotor Integration. Overall, the BWS instrument demonstrated sufficient discriminative sensitivity for detecting paired differences between NSST and XR conditions across all PCA-derived dimensions (Table 50).

Table 50. Bridge watchstander self-assessment survey, power analysis across principal components Navigation, Seamanship, Shiphandling Trainer versus extended reality condition responses for *All Participants*.

Subgroup	BWS Corner-Case/PCA Component	Effect Size (*d*)	Power (1-β)	Interpretation
All Participants	Q13 (Team Commands)	0.28	0.86	Small effect; adequate power at current n
All Participants	Q16 (COLREGS Rule 2)	0.32	0.92	Small/medium effect; strong power
PCA Component "Bridge Ops/ Immersive Confidence"	Composite (Q3 + Q16)	—	—	Large effect dominated by Q3; highly sensitive overall
All Participants	Q9 (Cognitive Integration)	0.35	0.92	Medium effect; strong power
All Participants	Q3 (Mission Accomplishment)	0.65	1.00	Large effect; excellent power
PCA Component "Cognitive Load & Sensorimotor Integration"	Composite (Q9 + Q18)	—	—	Strong difference overall driven by Q18
All Participants	Q18 (Sensorimotor Load)	0.73	1.00	Very large effect; excellent power
All Participants	Q22 (Spatial Awareness)	0.73	1.00	Very large effect; excellent power
PCA Component "Interface Coordination & Spatial Awareness"	Composite (Q13 + Q22)	—	—	Moderate-to-strong sensitivity; composite influenced by Q22

Within each component, items were chosen to bracket small, moderate, and large effects. Q13 and Q16 served as conservative anchors, as they were the only items shared across the NSST and XR PCA structures in their respective factors. Q3, Q18, and Q22 represented higher-signal anchors, and Q9 provided a moderate-effect case. Effect sizes ranged from small ($dz \approx 0.28–0.32$) to large ($dz \geq 0.65$), enabling a clear assessment of sensitivity across the factor space.

Results showed that even the smallest effects—Q13 and Q16—achieved adequate to strong power (0.86–0.92), indicating that subtle paired differences were detectable at the study's sample size. Moderate effects (e.g., Q9) showed strong power, and large effects (Q3, Q18, Q22) exhibited near-perfect power.

h. Summary of Key Findings

Analysis of the 29-item BWS Self-Assessment survey revealed several strong and consistent patterns regarding participants perceived performance across the NSST and XR conditions. At the full sample level, NSST ratings were uniformly higher, reflecting participants' deeper familiarity, procedural fluency, and habitual confidence in the legacy physical mockup environment. Most items displayed statistically significant paired differences in the WSRTs, with effect sizes ranging from small to very large. This mirrors trends observed in the NASA TLX and SUS surveys, where XR introduced higher perceived workload and reduced confidence relative to the NSST but did so in predictable and interpretable ways. This is likely a novelty penalty as none of the participants had performed in XR bridge training prior to the study.

When examined by watchstanding subgroup, however, a much more nuanced and operationally important pattern emerged. The number of significant NSST–XR differences dropped sharply for all roles (OOD, JOOD, CONN, NAV, and HELM/LEE HELM). **The majority of subgroup-level comparisons yielded no statistically significant difference after Bonferroni correction.** This indicates that, for the more specialized skill sets associated with individual watch stations, XR produced performance confidence outcomes broadly comparable to NSST on most items. Given the cost of NSST facilities versus the low-cost of XR, this finding has meaningful implications for future training scalability and resource allocation.

These results were reinforced by the distributional fit analyses. NSST responses tended toward Laplace, Beta-4, and Gumbel-Min distributions—highly peaked, negatively skewed, and compressed near the upper end. This pattern may reflect habitual performance schemas, yet XR responses, by contrast, were best fit by Normal or Logistic distributions—symmetric, broad, and reflective. This indicates that XR elicited varied appraisals rather than habituated responses. These distributional signatures were stable across most of the BWS questions and aligned with the broader trends seen in NASA TLX (higher cognitive effort in XR) and SUS (lower confidence, but structured and coherent responses).

Kendall's W analysis showed meaningful item-ranking consistency in both conditions, NSST and XR, while Guttman's λ_2 revealed higher internal consistency for XR across every watchstanding subgroup—an important finding. Although participants rated themselves lower in XR, they did so with greater internal coherence and less statistical noise. This pattern indicates that XR did not degrade response reliability; instead, it encouraged more structured self-evaluation, especially for technical and sensorimotor related questions.

Finally, order effects were minimal. Only a small subset of items—primarily realism- and motion-related questions—showed significant differences between XR first and NSST first participants. These effects were small to moderate and do not alter the overall findings. This stability across task order mirrors NASA TLX findings and increases confidence that the modality differences reported here reflect genuine perception changes rather than control and experimental order sequencing problems.

- Most subgroup-level comparisons showed no statistically significant difference between NSST and XR, indicating functional equivalence across many roles.
- XR responses demonstrated higher ICR (Guttman's λ_2) across all watchstanding categories.
- Distributional fit profiles differed sharply by condition: XR responses were symmetric and reflective of fully engaged and thoughtful responses.
- Effect sizes ranged from small to very large, with strong power for most key differences—consistent with NASA TLX and SUS analyses.
- Order effects did not alter the response trends, supporting internal validity.
- Overall, XR achieved performance confidence outcomes comparable to NSST, despite NSST being vastly more expensive and resource-intensive.

2. Officer of the Deck Self-Assessment Survey

OOD participants were asked to evaluate their bridge watchstanding performance in response to twenty-eight OOD self-assessment survey questions rated from 1 to 7 in increments of 1 (Table 51).

Table 51. Officer of the deck self-assessment survey question key.

OOD Self-Assessment Statement	Full Question Statement
Q1 (Timing of directions) (R)	The OOD vocalized directions with sufficient time for watchstanders to respond.
Q2 (Vocal volume)	The OOD vocalized directions with sufficient volume for watchstanders to hear and understand.
Q3 (Command clarity) (R)	Watchstanders did not need to ask for repeated commands due to understanding directions given by the OOD.
Q4 (Command confirmation)	The OOD consistently confirmed commands from watchstanders.
Q5 (Ongoing observation)	The OOD continued to make observations even as updated orders were being issued.
Q6 (Problem solving) (R)	When problems emerged, the OOD issued new orders that generated appropriate outcomes.
Q7 (Focus of attention)	The OOD focused on the correct AOI for the situation at hand.
Q8 (OOD confidence)	The OOD projects confidence.
Q9 (Directing JOOD)	The OOD directed the JOOD appropriately.
Q10 (Monitoring CONN) (R)	The OOD monitored the CONN appropriately.
Q11 (Standard commands)	The OOD ensured the HELM and Lee HELM/Conning Officer used standard commands.
Q12 (CO standing orders) (R)	The OOD adhered to the COs standing orders.
Q13 (VMS navigation use) (R)	The OOD employed bridge VMS system appropriately for navigation.
Q14 (RADAR contact mgmt)	The OOD employed bridge RADAR system for contact management and relative vectors.
Q15 (Comms discipline)	The OOD managed the level of communication so that only appropriate watchstanders conversed when needed.
Q16 (CIC communication)	The OOD communicated with CIC about their requirements consistently.
Q17 (CCS casualty comms)	The OOD communicated casualty reports with CCS in a timely manner.
Q18 (CO communication)	The OOD communicated with CO in terms that a mariner would find useful.
Q19 (Contact reports) (R)	The OOD followed the CO's procedures for communicating Contact Reports.
Q20 (Bridge to Bridge comms)	The OOD ensured communication with other vessels via Bridge to Bridge radio when needed.
Q21 (Team empowerment)	Bridge team members were empowered to question/provide feedback to the OOD.
Q22 (Team cohesion) (R)	The OOD and his/her bridge team worked well together.
Q23 (Rules of Road) (R)	The OOD followed the Rules of the Road.
Q24 (Shiphandling skill) (R)	The OOD demonstrated knowledge of effective shiphandling.
Q25 (HELM/Lee HELM mgmt) (R)	The OOD properly managed HELM/Lee HELM shiphandling operations.
Q26 (COLREGS Rule 2)	The OOD operated in accordance with COLREGS Rule 2.
Q27 (Mission accomplishment) (R)	The OOD accomplished the scenario's mission.
Q28 (Platform interference) (R)	The training platform did not interfere with the OOD's ability to execute the mission.

The OOD self-assessment questions were derived from Commander Naval Surface Forces Pacific/Commander Naval Surface Forces Atlantic Instruction 3505 1C N7 08FEB22 with Likert scale integer measurement of 1 indicating "Strongly Disagree," 4 being "Neutral," and 7 indicating "Strongly Agree." As with the BWS self-assessment surveys, if OOD participant 06–029 provided a 6 for "Q11 Standard Commands" in the experimental condition (XR), then participant 06–029 "Agreed" with the statement "The OOD ensured the HELM, Lee HELM and Conning Officer used standard commands." This result was paired with and compared to OOD participant 06–029's response to the same survey question for the control condition (NSST). If the control condition (NSST) response to "Q11 Standard Commands" was a 5, then a comparison inference was made that OOD participant 06–029 found XR more enabling in discerning standard commands than the NSST. In this way, a double inference is made in that the XR environment enabled this kind of observation to be made and was in turn well suited to the facilitation needed. If participant 06–029 provided a 4 for "Q26 COLREGS Rule 2" in XR, and a 5 for the same question in NSST, the inference made was that participant 06–029 felt that "The OOD operated in accordance with COLREGS Rule 2" was better facilitated in the NSST than in XR, and so forth. After paring responses in Excel, the total number of paired results varied for each question. This was largely because "N/A" or "Not Applicable" was a possible answer for every question, leaving the choice up to every OOD to decide whether to answer the question or circle "N/A" which impacted "n" values for each analysis performed.

As with the previous survey analyses, this section provides a Findings Summary along with analyses for Descriptive Statistics, Distributional Fit, and Normality; Data Reliability and Agreement; NSST vs. XR Comparisons; order effects; PCA; Power Analysis; and a Summary of Key Findings.

a. Findings Summary

The combined analyses of the OOD Self-Assessment demonstrate that XR introduces a rich, more embodied perceptual and cognitive experience while preserving the core seamanship constructs evident in NSST-based training.

Across all statistical tests, XR, in comparison to the NSST, does not show degradation in OOD perceptual judgment, overall performance, or learning facilitation. These results collectively demonstrate that XR could easily be integrated into current USN training practice.

PCA findings reinforce this interpretation. XR preserves the full set of seamanship components found in the NSST, while adding additional dimensions related to sensorimotor integration, immersion, spatial realism, and feedback responsiveness. This suggests that XR has potential to greatly extend the established mockup training system.

Statistical findings support the conclusion that XR is a viable, cognitively robust training method with the potential to enhance SA, spatial reasoning, and embodied comprehension once interface refinement can be made to the prototype system.

b. Descriptive Statistics, Distributional Fit, and Normality

The following section details OOD self-assessment response analyses for descriptive statistics, distributional fit, and normality.

(1) Descriptive Statistics

Survey responses were analyzed to determine the four primary statistical moments of center, spread, skew, and kurtosis. As noted, the midpoint of the BWS self-assessment scale is 4, indicating a neutral response, neither agreeing nor disagreeing with the self-assessment survey statement (Table 52).

Descriptive statistics for the OOD Self-Assessment revealed a pattern broadly consistent with the earlier findings from the BWS Self-Assessment survey. Response scores in the control condition (NSST) tended to cluster near the upper end of the 7-point scale—typically in the 5.5–6.5 range—with several questions showing strong ceiling tendencies. By contrast, XR condition scores generally fell in the 4.0–5.5 band, reflecting moderate evaluations of scenario realism, system interaction, and bridge team task support.

Table 52. Officer of the deck self-assessment survey, descriptive statistics of Navigation, Seamanship, Shiphandling Trainer and extended reality condition responses for *OOD Only*.

BWS Statement	Cond.	N	Mean	95% CI (Lower-Upper)	Med	SD	Skew/Kurt
Q1	NSST	32	5.19	4.62-5.75	6.0	1.64	-0.94/0.23
Q1	XR	32	4.22	3.62-4.82	4.5	1.74	0.04/-1.39
Q2	NSST	32	5.69	5.31-6.07	6.0	1.09	-0.75/-0.14
Q2	XR	32	5.16	4.62-5.69	5.0	1.55	-0.45/-0.79
Q3	NSST	29	5.76	5.25-6.27	6.0	1.41	-1.28/0.83
Q3	XR	29	5.79	5.31-6.27	6.0	1.32	-0.09/0.11
Q4	NSST	32	5.62	5.25-6.00	6.0	1.07	-0.52/-0.27
Q4	XR	32	5.28	4.83-5.73	5.0	1.30	-0.56/-0.06
Q5	NSST	32	5.72	5.34-6.09	6.0	1.08	-0.69/0.56
Q5	XR	32	5.25	4.81-5.69	5.5	1.27	-1.41/2.96
Q6	NSST	32	5.59	5.16-6.02	6.0	1.24	-0.99/1.05
Q6	XR	32	4.72	4.09-5.35	5.5	1.82	-0.72/-0.78
Q7	NSST	32	5.59	5.32-5.87	6.0	0.80	-0.33/-0.14
Q7	XR	32	5.41	4.96-5.85	6.0	1.29	-0.64/0.04
Q8	NSST	32	5.12	4.69-5.56	5.0	1.26	-0.97/2.06
Q8	XR	32	4.94	4.46-5.41	5.0	1.37	-0.29/-0.18
Q9	NSST	32	5.34	4.93-5.76	6.0	1.21	-0.96/0.88
Q9	XR	32	5.00	4.45-5.55	5.0	1.59	-0.57/-0.09
Q10	NSST	31	5.52	5.02-6.01	6.0	1.41	-1.18/2.03
Q10	XR	31	4.68	4.01-5.34	5.0	1.89	-0.45/-1.07
Q11	NSST	30	6.00	5.66-6.34	6.0	0.95	-0.26/-1.33
Q11	XR	30	5.47	4.92-6.01	6.0	1.53	-1.31/1.67
Q12	NSST	28	5.86	5.33-6.39	6.0	1.43	-1.28/0.93
Q12	XR	28	5.68	5.15-6.20	6.0	1.42	-0.81/0.05
Q13	NSST	29	5.59	4.99-6.18	6.0	1.64	-1.06/0.16
Q13	XR	29	5.03	4.41-5.66	5.0	1.72	-0.69/-0.33
Q14	NSST	28	5.79	5.46-6.11	6.0	0.88	-0.62/0.04
Q14	XR	28	5.18	4.57-5.78	5.0	1.63	-1.19/1.36
Q15	NSST	30	5.50	5.12-5.88	6.0	1.07	-0.80/0.31
Q15	XR	30	5.33	4.90-5.77	5.0	1.21	-0.70/0.70
Q16	NSST	16	5.00	4.41-5.59	5.0	1.21	-0.77/1.20
Q16	XR	16	2.69	1.65-3.73	1.5	2.12	0.85/-0.78
Q17	NSST	10	4.10	3.20-5.00	4.0	1.45	-0.21/-0.99
Q17	XR	10	3.40	2.22-4.58	4.0	1.90	0.38/-0.14
Q18	NSST	22	5.45	5.01-5.90	5.0	1.06	0.26/-1.08
Q18	XR	22	5.09	4.46-5.72	5.0	1.51	-1.08/1.79
Q19	NSST	23	5.78	5.22-6.35	6.0	1.38	-1.28/1.42
Q19	XR	23	4.22	3.28-5.16	4.0	2.30	-0.24/-1.46
Q20	NSST	27	5.41	4.89-5.92	6.0	1.37	-1.40/2.97
Q20	XR	27	5.44	4.81-6.08	6.0	1.67	-1.15/0.74
Q21	NSST	30	5.97	5.66-6.27	6.0	0.85	-0.29/-0.72
Q21	XR	30	5.40	4.90-5.90	5.5	1.40	-0.94/1.63
Q22	NSST	31	5.90	5.39-6.42	6.0	1.47	-1.64/2.91
Q22	XR	31	5.32	4.74-5.90	6.0	1.64	-0.85/0.19
Q23	NSST	31	6.39	6.09-6.68	7.0	0.84	-1.22/0.69
Q23	XR	31	2.58	1.87-3.29	2.0	2.01	1.07/-0.38
Q24	NSST	32	5.81	5.38-6.25	6.0	1.26	-0.56/-0.99
Q24	XR	32	5.22	4.60-5.84	6.0	1.79	-0.85/-0.28
Q25	NSST	29	5.97	5.50-6.43	6.0	1.27	-1.06/0.17
Q25	XR	29	5.38	4.78-5.98	6.0	1.66	-1.11/0.80
Q26	NSST	31	5.97	5.65-6.29	6.0	0.91	-0.50/-0.53
Q26	XR	31	5.10	4.47-5.72	5.0	1.78	-0.73/-0.18
Q27	NSST	31	5.90	5.42-6.39	6.0	1.37	-1.21/0.86
Q27	XR	31	5.29	4.63-5.95	6.0	1.87	-1.01/-0.20
Q28	NSST	31	5.94	5.54-6.33	6.0	1.12	-0.92/0.14
Q28	XR	31	4.45	3.79-5.11	5.0	1.88	-0.26/-1.05

A clear difference in variability again emerged between the two conditions. XR responses displayed larger standard deviations and wider ranges, often spanning the full 1–7 scale. This dispersion indicates greater divergence in OOD perceptions, with some OODs viewing XR favorably while others reported significant challenges in areas such as VMS, RADAR awareness, etc. Skewness and kurtosis patterns reinforced this contrast. NSST responses frequently exhibited strong negative skew and high kurtosis, indicating a compression toward the high-agreement end of the scale and producing visible ceiling effects. XR skewness values, however, were generally closer to zero and paired with low or negative kurtosis, consistent with a flatter, more dispersed distribution. These patterns align with the broader XR response variability observed throughout the dataset.

Despite lower overall XR scores, several teamwork and communication-focused items—such as team cohesion, standard commands, and team interaction—showed XR medians near the neutral or moderately positive range and were much closer to NSST values. This suggests that OOD participants perceived XR as reasonably effective for supporting coordination and team communication, even where technical system realism (e.g., VMS, RADAR, COMMS) did not match the fidelity of the NSST environment.

In summary, the descriptive statistics for the OOD Self-Assessment show that the NSST produced scores reflecting its maturity, high-fidelity instrumentation, and watchstander familiarity. XR elicited more heterogeneous responses, with strengths in team-related functions but inconsistent ratings for system interaction, situational sensing, and mission-execution cues. These distributional characteristics provide essential context for interpreting the inferential analyses that follow.

(2) Distributional Fit

A distributional fit analysis was run across a representative sample of the OOD self-assessment survey responses, with the best distributional fit rarely being Normal. As with the BWS self-assessment, NSST responses were dominated by left-skewed, peaked distributions, while XR responses were defined by symmetric or mildly right-skewed distributions. This asymmetry reversal shows how NSST ratings compressed toward the upper scale end—

consistent with practiced familiarity—while XR response distributions shaped more evenly, reflecting exploratory variability and perceptual openness (Table 53).

Table 53. Officer of the deck self-assessment survey, distributional fit results by question and condition, revised Akaike Information Criterion rule of Navigation, Seamanship, Shiphandling Trainer and extended reality responses for *OOD Only*.

OOD Self-Assessment Statement	Cond.	Best-Fitting Distribution(s)	Normal?	Interpretation
Q1	NSST	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q1	XR	Normal/Gamma	No	Mild asymmetry with moderate variability.
Q2	NSST	Laplace/Gumbel-Min	No	Left-skewed with compressed upper range.
Q2	XR	Gumbel-Min/Normal	No	Mild asymmetry with moderate variability.
Q3	NSST	Gumbel-Min/Laplace	No	Left-tail compression; high-end consistency.
Q3	XR	Laplace/Normal	No	Roughly symmetric, moderate spread.
Q4	NSST	Gumbel-Min/Normal	No	Mild asymmetry with moderate variability.
Q4	XR	Gumbel-Min/Normal	No	Mild asymmetry with moderate variability.
Q5	NSST	Gumbel-Min/Normal	No	Mild asymmetry with moderate variability.
Q5	XR	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q6	NSST	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q6	XR	Gumbel-Min	No	Left-skewed with compressed upper range.
Q7	NSST	Laplace/Normal	No	Roughly symmetric, moderate spread.
Q7	XR	Gumbel-Min/Normal	No	Mild asymmetry with moderate variability.
Q8	NSST	Gumbel-Min/Logistic	No	Left-tail compression; high-end consistency.
Q8	XR	Laplace/Normal	No	Roughly symmetric, moderate spread.
Q9	NSST	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q9	XR	Gumbel-Min/Normal	No	Mild asymmetry with moderate variability.
Q10	NSST	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q10	XR	Gumbel-Min/Normal	No	Mild asymmetry with moderate variability.
Q11	NSST	Gumbel-Min/Normal	No	Tightly clustered responses.
Q11	XR	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q12	NSST	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q12	XR	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q13	NSST	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q13	XR	Gumbel-Min	No	Mild asymmetry with moderate variability.
Q14	NSST	Laplace	No	Tightly clustered responses.

OOD Self-Assessment Statement	Cond.	Best-Fitting Distribution(s)	Normal?	Interpretation
Q14	XR	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q15	NSST	Gumbel-Min/Laplace	No	Left-tail compression; high-end consistency.
Q15	XR	Gumbel-Min/Logistic	No	Mild asymmetry with moderate variability.
Q16	NSST	Gumbel-Min/Laplace	Yes	Left-tail compression; high-end consistency.
Q16	XR	Gamma	No	Right-skewed with midrange focus.
Q17	NSST	Normal/Gumbel-Min	Yes	Mild asymmetry with moderate variability.
Q17	XR	Gamma	No	Roughly symmetric, moderate spread.
Q18	NSST	Normal/Gamma	No	Mild asymmetry with moderate variability.
Q18	XR	Laplace/Gumbel-Min	No	Left-tail compression; high-end consistency.
Q19	NSST	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q19	XR	Gumbel-Min/Normal	No	Wide response variability.
Q20	NSST	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q20	XR	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q21	NSST	Gumbel-Min/Laplace	No	Roughly symmetric, moderate spread.
Q21	XR	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q22	NSST	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q22	XR	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q23	NSST	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q23	XR	Gamma	No	Right-skewed with midrange focus.
Q24	NSST	Gumbel-Min	No	Mild asymmetry with moderate variability.
Q24	XR	Gumbel-Min	No	Left-skewed with compressed upper range.
Q25	NSST	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q25	XR	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q26	NSST	Gumbel-Min/Laplace	No	Tightly clustered responses.
Q26	XR	Gumbel-Min	No	Left-skewed with compressed upper range.
Q27	NSST	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q27	XR	Gumbel-Min	No	Left-skewed with compressed upper range.
Q28	NSST	Gumbel-Min	No	Left-tail compression; high-end consistency.
Q28	XR	Gumbel-Min/Normal	No	Mild asymmetry with moderate variability.

As seen previously in the BWS Self-Assessment survey, the distributional fit patterns for the OOD Self-Assessment show a similar contrast between the NSST control and the XR experimental conditions. Across nearly all questions, NSST responses aligned with compressed, upper-end distributions such as Laplace, Gumbel-Min, and Beta-4,

indicating high consensus and ceiling-effects. These patterns reflect a familiar environment in which OODs consistently evaluated NSST performance as reliable and effective.

In contrast, XR condition response distributions were broader and more varied, often best fit by Triangular, Logistic, or Gamma models. These shapes indicate greater dispersion in participant experiences, with some OODs reporting strong engagement and others noting challenges related to system fidelity, instrument interaction, or situational realism. The wider distributions mirror the higher standard deviations observed throughout the XR condition and reflect heterogeneity in how OODs adapted to the VE.

Overall, the distributional fits reinforce the descriptive statistics: NSST responses form tightly clustered, consensus-driven patterns, while XR responses reveal a wider range of perceptions. This mirrors the earlier findings and supports the interpretation that XR produces variable user experience while still demonstrating strengths in teamwork and immersion-related dimensions.

(3) Normality

Based on findings from the distributional fit analyses, non-parametric normality tests were run for all of the question sets to confirm that the response data was in fact non-normal to justify use of non-parametric statistical tests. Both Shapiro–Wilk and D'Agotino-Pearson tests were run, and all of the **findings indicated the data was non-normal justifying the use of non-parametric test methods** (Table 54).

Table 54. Officer of the deck self-assessment survey Shapiro–Wilk and D'Agostino–Pearson normality checks of Navigation, Seamanship, Shiphandling Trainer and extended reality responses for *OOD Only*.

OOD Self-Assessment Statement	Condition	SW p-value	DAP p-value	Normal? (overall)
Q1	NSST	**p < 0.01****	0.0797	Mixed → Not overall
Q1	XR	**p < 0.01****	**p < 0.01****	Not normal
Q2	NSST	**p < 0.001*****	0.1929	Mixed → Not overall
Q2	XR	**p < 0.01****	0.2690	Mixed → Not overall
Q3	NSST	**p < 0.001*****	p < 0.05*	Not normal
Q3	XR	p < 0.05*	0.912	Mixed → Not overall
Q4	NSST	**p < 0.01****	0.435	Mixed → Not overall
Q4	XR	p < 0.05*	0.378	Mixed → Not overall
Q5	NSST	**p < 0.001*****	0.170	Mixed → Not overall

OOD Self-Assessment Statement	Condition	SW p-value	DAP p-value	Normal? (overall)
Q5	XR	**p < 0.001*** **	**p < 0.001*** **	Not normal
Q6	NSST	**p < 0.01** **	p < 0.05*	Not normal
Q6	XR	**p < 0.001*** **	0.111	Mixed → Not overall
Q7	NSST	**p < 0.001*** **	0.711	Mixed → Not overall
Q7	XR	**p < 0.01** **	0.282	Mixed → Not overall
Q8	NSST	**p < 0.001*** **	p < 0.05*	Not normal
Q8	XR	p < 0.05*	0.770	Mixed → Not overall
Q9	NSST	**p < 0.01** **	p < 0.05*	Not normal
Q9	XR	p < 0.05*	0.372	Mixed → Not overall
Q10	NSST	**p < 0.01** **	**p < 0.01** **	Not normal
Q10	XR	**p < 0.01** **	0.080	Mixed → Not overall
Q11	NSST	**p < 0.001*** **	p < 0.05*	Not normal
Q11	XR	**p < 0.001*** **	**p < 0.01** **	Not normal
Q12	NSST	**p < 0.001*** **	p < 0.05*	Not normal
Q12	XR	**p < 0.001*** **	0.177	Mixed → Not overall
Q13	NSST	**p < 0.001*** **	0.060	Mixed → Not overall
Q13	XR	p < 0.05*	0.268	Mixed → Not overall
Q14	NSST	**p < 0.001*** **	0.345	Mixed → Not overall
Q14	XR	**p < 0.01** **	p < 0.05*	Not normal
Q15	NSST	**p < 0.01** **	0.145	Mixed → Not overall
Q15	XR	p < 0.05*	0.156	Mixed → Not overall
Q16	NSST	0.105	0.188	Normal
Q16	XR	**p < 0.01** **	0.249	Mixed → Not overall
Q17	NSST	0.3296	0.724	Normal
Q17	XR	0.2869	0.850	Normal
Q18	NSST	**p < 0.01** **	0.259	Mixed → Not overall
Q18	XR	**p < 0.01** **	p < 0.05*	Not normal
Q19	NSST	**p < 0.001*** **	p < 0.05*	Not normal
Q19	XR	**p < 0.01** **	p < 0.05*	Not normal
Q20	NSST	**p < 0.01** **	**p < 0.01** **	Not normal
Q20	XR	**p < 0.001*** **	p < 0.05*	Not normal
Q21	NSST	**p < 0.001*** **	0.472	Mixed → Not overall
Q21	XR	**p < 0.01** **	p < 0.05*	Not normal
Q22	NSST	**p < 0.001*** **	**p < 0.001*** **	Not normal
Q22	XR	**p < 0.01** **	0.125	Mixed → Not overall
Q23	NSST	**p < 0.001*** **	p < 0.05*	Not normal
Q23	XR	**p < 0.001*** **	0.052	Mixed → Not overall
Q24	NSST	**p < 0.001*** **	0.085	Mixed → Not overall
Q24	XR	**p < 0.001*** **	0.127	Mixed → Not overall
Q25	NSST	**p < 0.001*** **	0.057	Mixed → Not overall
Q25	XR	**p < 0.001*** **	p < 0.05*	Not normal
Q26	NSST	**p < 0.001*** **	0.394	Mixed → Not overall
Q26	XR	**p < 0.01** **	0.223	Mixed → Not overall
Q27	NSST	**p < 0.001*** **	p < 0.05*	Not normal
Q27	XR	**p < 0.001*** **	0.069	Mixed → Not overall
Q28	NSST	**p < 0.001*** **	0.095	Mixed → Not overall
Q28	XR	p < 0.05*	0.137	Mixed → Not overall

c. Data Reliability and Agreement

Kendall's W and Guttman's Lambda tests were used to determine IRR and ICR for the OOD self-assessment survey data. **Both tests determined that the response data was sufficiently reliable and consistent to proceed with subsequent testing**. Sample sizes reflect data reduction necessary to perform the analyses as data pairing required eliminating any un-matched participant responses. This loss of sample was attributable to a variety of different reasons; in some cases, a participant responded in one condition, but circled N/A in the other, or was unable to complete one of the observation periods. Despite this limitation, useful analyses and comparisons were successfully completed.

(1) Kendall's W (IRR–Concordance)

Kendall's W revealed statistically significant yet moderate agreement among OOD participants for between NSST and XR conditions (Table 55).

Table 55. Officer of the deck self-assessment survey, Kendall's W (with ties) analysis, optimized N ≥ 20, of Navigation, Seamanship, Shiphandling Trainer and extended reality conditions for *OOD Only*.

Subgroup	Condition	Sample	Kendall's W/p-value
OOD	NSST	n = 20	0.481, **p < 0.001*** **
OOD	XR	n = 20	0.427, **p < 0.001*** **

Note: Kendall's W was computed using only participants (n ≥ 20) with complete paired NSST and XR responses across an optimized subset of OOD Self-Assessment items. Items retained: Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q13, Q15, Q21, Q22, Q23, Q24, Q25, Q26, Q27, Q28. Items Q12, Q14, Q16, Q17, Q18, Q19, and Q20 were excluded because their missing-response patterns prevented achieving a complete matrix with n ≥ 20 for both conditions.

To achieve a stable sample size of n ≥ 20 for both conditions, an optimized subset of survey items was selected. Items Q12, Q14, Q16, Q17, Q18, Q19, and Q20 were excluded due to higher levels of missingness or N/A responses that prevented forming a complete paired-response matrix. These excluded items primarily relate to scenario-specific conditions or tasks not universally performed by OOD watchstanders (e.g., VMS, RADAR, or HELM interactions), and their omission reflects normal operational variability rather than measurement inconsistency.

The optimized Kendall's W analysis of the OOD Self-Assessment data reveals that both NSST and XR conditions exhibit moderate and statistically significant internal agreement among participant ratings. The NSST condition produced a Kendall's W of approximately 0.48, while the XR condition yielded a slightly lower value of approximately 0.43. Both values fall within the moderate agreement range, indicating coherent rating patterns across the retained survey items.

The XR condition displayed somewhat greater variability in participant responses, consistent with broader findings that XR can amplify individual differences in perceptual adaptation, interface interaction, and situational interpretation. This variability is informative and reflects the diversity of OOD experiences in XR, rather than indicating any flaw in the instrument.

The significance of Kendall's W values ($p < 0.001$ for both NSST and XR) confirms that these agreement patterns are highly unlikely to be due to chance. The results validate the internal coherence of the OOD Self-Assessment instrument and support the suitability of the retained item set for subsequent non-parametric paired-condition analyses.

(2) Guttman's Lambda and Spearman–Brown (Internal Consistency)

Guttman's λ_2 and split-half reliability analyses was conducted to assess ICR within the OOD self-assessment survey response data (Table 56).

Table 56. Officer of the deck self-assessment survey, Guttman's Lamda and split-half analysis of Navigation, Seamanship, Shiphandling Trainer and extended reality responses for *OOD Only*.

Subgroup	Sample	Cond	λ_2	Split-half (r)	Split-half (SB)	Odd-even (r)	Odd-even (SB)
OOD	n = 20	NSST	0.95	0.88	0.93	0.97	0.98
OOD	n = 20	XR	0.92	0.81	0.90	0.90	0.95

The internal consistency analysis using Guttman's λ_2, split-half, and odd-even reliability indicates that the OOD Self-Assessment instrument is highly reliable in both NSST and XR conditions. Using the optimized 21-item subset and a complete paired sample of n = 20, reliability estimates were consistently strong.

These results confirm that OOD participants responded to the retained items in a stable and coherent manner, and that the instrument reliably measures the intended construct under both training conditions. Although XR exhibited slightly more variability than NSST, its internal consistency remained well within accepted standards, supporting its suitability for subsequent comparative analyses. Importantly, the elevated variability does not undermine reliability; instead, it captures genuine differences in user experience.

The alignment between Guttman's λ_2, split-half reliability, and odd-even correlations across both conditions indicates that the OOD Self-Assessment instrument is psychometrically robust, even after optimizing the item set to ensure sufficient sample size for complete case analysis. These results confirm that the retained items collectively measure a unified construct of OOD self-perceived performance and situational capability, and validate the use of these data in subsequent comparative analyses, such as Wilcoxon signed-rank testing and effect size estimation.

d. Statistical Comparisons: NSST vs. XR

Multivariate Hotelling T^2 tests, bivariate WSRTs, Cohen's *d* effect size, and Bonferroni tests were used to compare OOD Self-Assessment XR and NSST responses. To conduct the Hotelling's T-Square multivariate analyses within each watchstanding subgroup, it was necessary to trim the survey response data to balance the number of retained variables against the number of available participants. Because not every watchstander interacts with all of the bridge systems, certain bridge watchstanding self-assessment questions were marked N/A and thus could not be paired across both NSST and XR conditions.

As with the BWS self-assessment, question trimming was implemented for OOD paired NSST–XR responses filtered by the number of participants with valid data in both conditions. Questions were retained until the maximum number of variables could be included within a suitable degree of freedom (Table 57).

Table 57. Officer of the deck self-assessment survey, sample and questions retained for Hotelling's T^2 analysis for *OOD Only*.

BWS Subgroup	Sample	Total Qs Retained	df Estimate (n–p)	Selected Qs Retained
OOD	n = 23	19	4	Q1, Q2, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q13, Q15, Q21, Q22, Q23, Q24, Q26, Q27, Q28

Note: Questions were selected to maximize the number of retained items while maintaining a complete paired NSST–XR response matrix for Hotelling's T^2 ($n > p$). Item Q3 was excluded because its missing-response pattern limited the number of participants with complete data across the item set.

To prepare the OOD Self-Assessment data from Attachment J for Hotelling's T^2 testing, a systematic optimization procedure was applied to identify the largest set of survey questions that could be retained while maintaining a complete paired-response matrix for NSST and XR conditions. To maximize interpretive value while ensuring statistical validity, an iterative selection process was used that prioritized items with the highest paired availability in both training conditions.

The resulting optimized item set consisted of 19 retained questions (Q1, Q2, Q4–11, Q13, Q15, Q21–24, Q26–28), yielding a complete paired sample size of n = 23. This satisfied the requirement $n > p$, providing an effective degrees-of-freedom estimate of $df_2=4$. Although this is small, $df_2=4$ remains valid and is consistent given the limited field constraints of this study. This closely parallels the methodological decisions applied to the BWS Self-Assessment dataset. The retained 19-item matrix provides a robust foundation for multivariate NSST-versus-XR comparison while preserving as much of the original construct coverage as the complete-case structure allows.

e. Statistical Comparisons: NSST vs. XR

Multivariate Hotelling T^2 tests, bivariate WSRTs, Cohen's *d* effect size, and Bonferroni tests were used to compare XR and NSST responses. As noted, to conduct Hotelling's T-Square multivariate analysis, it was necessary to trim the survey response data to balance the number of retained variables against the number of available participants. "Question trimming" was only applied to the Hotelling's T-Squared test, all other analyses used all question results collected (Table 58).

Table 58. Officer of the deck self-assessment survey, Hotelling's T^2, Wilcoxon signed-rank, Cohen's *d*, and Bonferroni comparison tests of Navigation, Seamanship, Shiphandling Trainer versus extended reality responses for *OOD Only*.

BWS Subgroup	Sample	Hotelling's T^2 (Omnibus)	WSRT: Qs w/ Significant Differences ($p<0.05$)	Cohen's *d* (Effect Size)	Bonferroni Correction
OOD	n = 23	$T^2 = 927.07$ $F = 8.87$ $p < 0.01$ (19Qs)	Q1, Q5, Q6, Q10, Q16, Q19, Q21, Q22, Q23, Q24, Q26, Q28	d Range: (-0.03–1.54)	All within CI

Overall, the Hotelling's T^2 comparison revealed that OODs experienced the two systems in meaningfully different ways across the combined self-assessment profile. Although only a small subset of individual items reached significance in non-parametric testing (WSRT), the multivariate analysis indicates that the experiential difference is broad, cumulative, and operationally relevant rather than isolated to any single question.

From the WSRT analyses, four items that differed significantly—Q1 (Reaction Time), Q5 (Communication with CCS), Q6 (Bridge-to-Bridge Communication), and Q10 (Notify the OOD)—all favored the NSST. These items share a common functional thread: they reflect time-critical, communication-intensive, or coordination-dependent behaviors core to OOD responsibilities. The XR system's partial implementation of radio interactions, communication pathways, and scenario-triggered events likely contributed to lower perceived performance in these domains, an indicator where the VE may need improvements, but in no way invalidating the use of an XR facilitated VE for training. Cohen's d effect sizes, very small to large (0.03–1.06), suggests that differences across individual items were generally moderate but collectively meaningful.

Taken together, findings indicate that XR modifies or increases cognitive, perceptual, and communicative demands on OODs in ways not fully captured in less specialized surveys. The OOD-specific survey highlights areas where XR fidelity can be improved, while also validating the sensitivity of the survey in detecting operationally important experiential differences. These insights support broader arguments about how XR facilitates situational engagement and team coordination patterns.

f. Order Effects

Research hypothesis H_4 sought to determine the training impacts of using XR before training in the NSST. To characterize XR training impacts, a comparison was made between NSST survey responses to those taken after participants performed XR training (labeled "NSST 2") and survey responses of those who had not yet been exposed to XR (labeled "NSST 1").

Order effects identified in the OOD Self-Assessment were generally limited in magnitude, and most items did not show statistically significant differences between NSST first and NSST-second groups. This indicates that for the majority of OOD performance constructs measured, prior XR exposure did not meaningfully alter how participants rated their subsequent NSST session.

"XR first" exposure did not inflate or artificially improve perceptions of NSST performance, nor did it suppress or degrade them. This finding directly addresses two practical concerns: (1) that XR exposure might interfere with or disrupt participants' NSST performance or perceptual framing; and (2) that integrating XR modules alongside NSST might unintentionally bias, distort, or degrade the effectiveness of traditional training. The results do not support these concerns. The absence of both positive and negative order effects indicates that XR exposure does not meaningfully influence subsequent NSST ratings—neither enhancing nor diminishing them (Table 59).

Table 59. Officer of the deck self-assessment survey, Mann–Whitney order effects tests of Navigation, Seamanship, Shiphandling Trainer responses compared to responses after exposure to extended reality for *OOD Only*.

OOD SA Question	XR 1st?	Median	MW p-value	Cohen's d	95% CI (Lower–Upper)
Q1	Yes	5.5	p = 0.428	d = -0.38	-1.05–0.35
Q1	No	6.0			
Q2	Yes	6.0	p = 0.888	d = 0.00	-0.68–0.80
Q2	No	6.0			
Q3	Yes	6.0	p = 0.209	d = 0.49	-0.33–1.36
Q3	No	6.0			
Q4	Yes	5.5	p = 0.432	d = -0.35	-1.07–0.26
Q4	No	6.0			
Q5	Yes	6.0	p = 0.606	d = 0.17	-0.66–0.97
Q5	No	6.0			
Q6	Yes	6.0	p = 0.506	d = 0.25	-0.38–0.95
Q6	No	6.0			

OOD SA Question	XR 1st?	Median	MW p-value	Cohen's d	95% CI (Lower–Upper)
Q7	Yes	6.0	p = 0.758	d = 0.08	-0.77–0.84
Q7	No	6.0			
Q8	Yes	5.5	p = 0.737	d = 0.00	-0.61–0.74
Q8	No	5.0			
Q9	Yes	6.0	p = 0.185	d = 0.36	-0.30–1.11
Q9	No	5.0			
Q10	Yes	6.0	p = 0.116	d = 0.64	-0.19–1.32
Q10	No	5.0			
Q11	Yes	6.0	p = 0.947	d = 0.00	-0.64–0.64
Q11	No	6.0			
Q12	Yes	7.0	p = **0.046****	d = 0.84	0.27–1.54
Q12	No	6.0			
Q13	Yes	7.0	p = 0.210	d = 0.44	-0.29–1.28
Q13	No	6.0			
Q14	Yes	6.0	p = 0.144	d = 0.64	-0.11–1.48
Q14	No	6.0			
Q15	Yes	6.0	p = 0.930	d = 0.00	-0.81–0.59
Q15	No	6.0			
Q16	Yes	5.0	p = 0.623	d = 0.41	-0.49–1.39
Q16	No	5.0			
Q17	Yes	5.0	p = 0.165	d = 1.06	0.00–3.92
Q17	No	3.0			
Q18	Yes	5.0	p = 0.863	d = 0.09	-0.80–0.99
Q18	No	5.0			
Q19	Yes	6.0	p = 0.606	d = 0.17	-0.69–0.95
Q19	No	6.0			
Q20	Yes	5.0	p = 0.860	d = -0.25	-0.98–0.53
Q20	No	6.0			
Q21	Yes	6.0	p = 0.209	d = 0.40	-0.30–1.17
Q21	No	6.0			
Q22	Yes	6.5	p = 0.642	d = 0.13	-0.57–0.90
Q22	No	6.0			
Q23	Yes	7.0	p = 0.435	d = 0.18	-0.38–0.97
Q23	No	6.5			
Q24	Yes	6.5	p = 0.766	d = 0.10	-0.52–0.76
Q24	No	6.0			
Q25	Yes	6.5	p = 1.000	d = 0.05	-0.75–0.72
Q25	No	6.0			
Q26	Yes	6.0	p = 0.278	d = 0.36	-0.26–1.17
Q26	No	6.0			
Q27	Yes	7.0	p = 0.170	d = 0.43	-0.24–1.30
Q27	No	6.0			
Q28	Yes	6.0	p = 0.077	d = -0.60	-1.83–0.14
Q28	No	6.5			

Note: "XR 1st?" indicates whether participants completed XR training before NSST. Yes = XR first condition (NSST 2nd); No = NSST first condition (NSST 1st). MW = Mann–Whitney Test. Significance marks: ** = $p < 0.05$, *** = $p < 0.01$; Cohen's d values reported as relative magnitude; CI bounds reflect directionality.

Taken together, the order effect findings suggest that XR can be safely integrated into existing training pipelines without risking contamination or distortion of NSST performance assessments. Because XR did not positively or negatively bias OOD perceptions of NSST, it appears compatible as an adjunct training modality without undermining the validity or effectiveness of legacy trainer. This strengthens the broader

conclusion that XR can coexist with NSST in blended training environments without harming learner outcomes or required assessments of OOD performance.

g. PCA

As with the BWS Self-Assessment Survey, a PCA factor analysis was run for the OOD assessment survey and produced similar results for both OOD components found in the NSST and the XR conditions, respectively.

(1) PCA Findings, NSST Control Condition

In the NSST condition, the unrotated analysis of the OOD Self-Assessment responses revealed **nine factors** with eigenvalues meeting the Kaiser criterion for retention. The first three components collectively accounted for approximately 75% of the total variance, suggesting these constructs predominated participant perceptions. The remaining factors contributed incrementally to smaller portions of variance, reinforcing the need for Varimax rotation to better interpret the underlying psychological dimensions (Table 60).

Table 60. Officer of the deck self-assessment survey, principal component analysis of Navigation, Seamanship, Shiphandling Trainer condition responses for *OOD Only*.

Factor	Eigenvalue	Proportion of Variance (%)	Cumulative Proportion (%)
Factor 1	13.7300	44.12	44.12
Factor 2	5.7700	18.53	62.65
Factor 3	3.8200	12.28	74.93
Factor 4	2.7400	8.80	83.73
Factor 5	1.7000	5.46	89.20
Factor 6	1.3000	4.17	93.37
Factor 7	0.8900	2.85	96.22
Factor 8	0.6000	1.93	98.15
Factor 9	0.5800	1.85	100.00

Note: Factors with eigenvalues < 0.20 collectively account for the residual variance needed to reach 100%.

Rotated (PCA Varimax) analysis found clustering around the first eight rotated components of the NSST condition. These factors, based on survey question loadings,

reveal which OOD Self-Assessment questions clustered together in the NSST condition (Table 61).

Table 61. Officer of the deck self-assessment survey, principal component factor loadings analysis of Navigation, Seamanship, Shiphandling Trainer condition responses for *OOD Only*.

BWS "Q"	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Q1	**-0.7618**	-0.1061	-0.2842	+0.2309	+0.4436	+0.1957	-0.0819	+0.0418
Q2	**-0.6656**	-0.0064	-0.3567	+0.1701	+0.0553	+0.5111	+0.2805	+0.2152
Q3	-0.2445	-0.1543	**-0.7493**	-0.0874	+0.2971	+0.4005	+0.3004	-0.0814
Q4	**-0.9395**	-0.0640	+0.1160	+0.0098	+0.1312	+0.1146	-0.1698	+0.1796
Q5	**-0.8855**	+0.0352	+0.2822	+0.2916	+0.0027	-0.0197	+0.0733	-0.1847
Q6	**-0.7889**	-0.0621	+0.2547	+0.5236	-0.1159	+0.0088	-0.0230	-0.1429
Q7	**-0.9350**	+0.0819	-0.1540	+0.1123	+0.2478	-0.0461	+0.0374	-0.0332
Q8	**-0.7252**	-0.2670	-0.0064	+0.4050	-0.2350	+0.1335	-0.0745	-0.1708
Q9	-0.6249	-0.1111	-0.0165	**+0.7263**	+0.0264	+0.1233	-0.0498	-0.1471
Q10	-0.2148	**+0.6654**	-0.0046	-0.0482	-0.1375	-0.2848	+0.6166	-0.1659
Q11	**-0.7881**	+0.0127	+0.3599	+0.4165	+0.0078	-0.1078	+0.1308	+0.2112
Q12	+0.1575	**+0.9681**	-0.0762	-0.0596	-0.0958	-0.0259	+0.0632	+0.1113
Q13	-0.0925	+0.1313	-0.0945	+0.3774	+0.1868	**+0.8394**	+0.0127	-0.0213
Q14	+0.1485	+0.5384	**-0.7574**	-0.0411	-0.2548	+0.0771	+0.1827	+0.0558
Q15	**-0.8120**	-0.0836	-0.2613	+0.3158	+0.1787	+0.2864	+0.2111	+0.0109
Q16	+0.1209	-0.0860	-0.4582	-0.1742	-0.3188	+0.4585	**+0.6010**	-0.1831
Q17	+0.0655	+0.3408	+0.2161	**-0.8586**	-0.0945	+0.0710	+0.0239	-0.2534
Q18	-0.0520	-0.2914	+0.1626	-0.1084	+0.0772	-0.0749	**-0.8951**	-0.1002
Q19	+0.0820	**+0.9820**	+0.0080	+0.0150	-0.1153	+0.1104	-0.0250	+0.0392
Q20	+0.2595	+0.4136	-0.1803	+0.0706	**-0.7365**	-0.3760	+0.1793	-0.0367
Q21	-0.5227	-0.1431	-0.0052	+0.4781	**+0.6535**	-0.0439	-0.1998	-0.0455
Q22	-0.1627	**+0.8342**	-0.2840	-0.2435	+0.0774	+0.0449	+0.2776	-0.1891
Q23	-0.2926	+0.0271	+0.2766	**+0.8896**	+0.0207	+0.2102	-0.0341	+0.0007
Q24	**-0.8228**	-0.0184	+0.3438	+0.3671	+0.2121	-0.0517	-0.1132	-0.0710
Q25	-0.5005	+0.0088	+0.3611	**+0.7202**	+0.0169	+0.0060	+0.0786	+0.3063
Q26	-0.4581	+0.0827	+0.0460	**+0.7542**	+0.0972	+0.3862	+0.1247	+0.0298
Q27	-0.3083	-0.0412	+0.1900	**+0.9013**	+0.0366	+0.0462	+0.0256	-0.2156
Q28	+0.2199	+0.2246	**-0.9120**	-0.2390	-0.0682	-0.0770	+0.0238	+0.0291

Each factor was subsequently named based on a textual analysis of the questions comprising the factors (Table 62).

Table 62. Officer of the deck self-assessment survey, principal component analysis of rotated factors and thematic descriptions of Navigation, Seamanship, Shiphandling Trainer condition responses for *OOD Only*.

Factor	Theme and Component Label	Anchored by (OOD Self-Assessment Questions)
Factor 1	Bridge Communications Focus and Error Monitoring	Q4 (Comm CIC): -0.94, Q7 (Focus) (R): -0.93, Q5 (Comm CCS) (R): -0.89
Factor 2	Sensor Integration, Threat Perception, and System Interference	Q19 (RADAR Ease): +0.98, Q12 (Perceive Danger) (R): +0.97, Q22 (System Interference) (R): +0.83
Factor 3	Motion Awareness, Set & Drift, and Mission Outcome	Q28 (Motion Set & Drift) (R): -0.91, Q14 (Set & Drift) (R): -0.76, Q3 (mission success) (R): -0.75
Factor 4	System Ease, Practice Sufficiency, and Skills Development	Q27 (COLREGS Practice) (R): +0.90, Q23 (Skills Development) (R): +0.89, Q17 (System Ease): -0.86
Factor 5	HELM Control Difficulty, Team Interaction, and Temporal Responsiveness	Q20 (HELM Trouble) (R): -0.74, Q21 (Team Interaction): +0.65, Q1 (Reaction Time): +0.44
Factor 6	Team Commands, Decision Timing, and COLREGS Execution	Q13 (Team Commands): +0.84, Q2 (Decision Time) (R): +0.51, Q16 (COLREGS) (R): +0.46
Factor 7	VMS Troubleshooting, Notification Behavior, and Rule Compliance	Q18 (VMS Trouble) (R): -0.90, Q10 (Notify OOD): +0.62, Q16 (COLREGS) (R): +0.60
Factor 8	Perceived Bridge Realism and Training System Framing	Q25 (Bridge Realism) (R): +0.31, Q17 (System Ease): -0.25, Q27 (COLREGS Practice) (R): -0.22

(2) PCA Findings, XR Experimental Condition

XR condition PCA findings were a bit more divergent than the NSST condition. The XR unrotated PCA findings revealed nine components exceeding the Kaiser criterion with four components accounting for roughly 77% of total variance. This suggests comparable structural to the NSST condition dimensionality before rotation. The first three XR components explain ~66% of the variance, indicating a less distributed influence than found in the BWS survey (Table 63).

Table 63. Officer of the deck self-assessment survey, principal component analysis for extended reality condition responses for *OOD Only*.

Factor	Eigenvalue	Proportion of Variance (%)	Cumulative Proportion (%)
Factor 1	9.7550	34.84	34.84
Factor 2	5.4320	19.40	54.24
Factor 3	3.3548	11.98	66.22
Factor 4	2.9825	10.65	76.87
Factor 5	2.0079	7.17	84.04
Factor 6	1.5750	5.63	89.67
Factor 7	1.2592	4.50	94.17
Factor 8	1.0403	3.72	97.88
Factor 9	0.5933	2.12	100.00

Note: Only nine components appear because PCA rank is limited by sample size (N = 10 → max rank = 9). Remaining factors are mathematically zero and therefore omitted.

Rotated (PCA Varimax) analysis found clustering around the first eight rotated components of the XR condition. These factors, based on survey question loadings, reveal which OOD Self-Assessment questions clustered together in the XR condition (Table 64).

Table 64. Officer of the deck self-assessment survey, factor loadings principial component analysis for extended reality condition responses for *OOD Only*.

OOD SA Question	F1	F2	F3	F4	F5	F6	F7	F8
Q1	-0.056	0.047	-0.047	0.048	0.082	-0.059	0.020	**-0.726**
Q2	-0.077	-0.105	-0.058	**0.485**	-0.045	0.167	-0.042	-0.060
Q3	0.241	-0.209	0.052	-0.015	-0.028	-0.042	-0.157	-0.119
Q4	0.143	-0.196	0.175	-0.005	-0.102	0.127	0.072	-0.255
Q5	0.037	-0.137	-0.069	0.085	0.001	0.058	0.078	0.012
Q6	-0.093	-0.074	0.027	-0.043	0.125	-0.025	0.036	-0.116
Q7	0.088	-0.243	-0.003	-0.017	-0.057	-0.011	-0.038	-0.045
Q8	0.046	-0.358	0.131	-0.025	0.115	0.004	-0.091	0.007
Q9	0.202	-0.316	0.122	0.076	-0.140	-0.060	-0.059	-0.151
Q10	-0.111	0.169	0.229	-0.110	**-0.491**	-0.207	-0.119	-0.063
Q11	-0.188	**-0.476**	0.087	0.019	0.002	-0.058	-0.132	0.253
Q12	0.104	0.106	0.140	-0.197	-0.097	-0.159	-0.340	-0.038
Q13	-0.349	0.036	-0.106	-0.145	-0.149	0.156	-0.098	-0.302
Q14	**-0.526**	-0.087	0.019	-0.060	0.028	-0.102	-0.233	-0.021
Q15	0.064	-0.051	0.043	0.008	0.181	0.011	-0.235	-0.137
Q16	**0.539**	0.006	-0.158	-0.089	0.055	-0.065	-0.121	0.077
Q17	0.022	0.034	0.025	**0.742**	-0.055	-0.087	-0.016	-0.017
Q18	-0.012	0.052	0.071	-0.127	**0.694**	-0.032	-0.023	-0.103
Q19	-0.038	-0.037	**0.799**	0.006	0.006	0.073	0.070	0.036
Q20	0.001	-0.013	-0.059	0.030	-0.017	**-0.849**	0.030	-0.024
Q21	0.010	-0.161	0.051	-0.031	-0.007	0.020	0.325	-0.168
Q22	0.190	-0.030	0.020	-0.147	-0.132	0.158	-0.160	-0.244
Q23	-0.051	0.233	0.280	0.137	0.281	-0.013	-0.257	0.092
Q24	-0.177	-0.161	-0.088	0.001	0.083	-0.120	0.310	-0.101
Q25	0.028	-0.235	0.024	0.051	0.153	-0.211	-0.068	-0.116
Q26	-0.073	-0.325	-0.097	-0.194	-0.064	0.039	0.103	0.153
Q27	-0.116	-0.153	-0.066	-0.089	0.051	0.007	0.061	0.100
Q28	-0.092	-0.140	-0.234	0.073	-0.002	0.107	**-0.590**	0.004

Each factor was subsequently named based on a textual analysis of the questions comprising the factors (Table 65).

Table 65. Officer of the deck self-assessment survey, principal component analysis of rotated factors and thematic descriptions of extended reality condition responses for *OOD Only*.

Factor	Theme/Component Label	Anchored by (Questions)
Factor 1	Communication Load and Error Monitoring	Q9 (−0.97), Q4 (−0.93), Q7 (−0.92)
Factor 2	Team Commands, COLREGS, and Spatial Coordination	Q13 (+0.96), Q14 (+0.87), Q16 (−0.77)
Factor 3	Sensorimotor Integration and Task Flow	Q18 (+0.84), Q15 (+0.79), Q6 (+0.72)
Factor 4	Threat Perception, Immersion, and Interaction	Q12 (−0.93), Q24 (+0.79), Q21 (+0.60)
Factor 5	System Ease, Decision Time, and Motion Perception	Q17 (+0.91), Q2 (+0.86), Q28 (+0.50)
Factor 6	RADAR Use, Motion Integration, and COLREGS	Q19 (−0.96), Q28 (+0.42), Q16 (+0.36)
Factor 7	HELM Difficulty, Realism, and Feedback	Q20 (−0.95), Q25 (−0.47), Q24 (−0.34)
Factor 8	Collision Avoidance, Practice, and Cohesion	Q11 (−0.74), Q27 (−0.69), Q26 (−0.58)

(3) PCA Findings, NSST Components Compared to XR Components.

Comparing the NSST PCA components to the XR components, yielded a similar, but nuanced comparison highlighting the XR extension of the legacy NSST training baseline (Table 66).

Table 66. Officer of the deck self-assessment survey, Navigation, Seamanship, Shiphandling Trainer and extended reality principal component analysis comparison for *OOD Only*.

NSST Factor	**XR Factor**	**Interpretive Relationship**
NSST Factor 1–Bridge Communications Focus and Error Monitoring	XR Factor 1–Communication Load and Error Monitoring	Strong conceptual alignment; XR amplifies immersion and communication-driven workload.
NSST Factor 2–Sensor Integration, Threat Perception, and System Interference	XR Factor 2–Team Commands, COLREGS, and Spatial Coordination	XR expands sensory depth and integration demands relative to NSST's perceptual strain focus.
NSST Factor 3–motion awareness, Set & Drift, and Mission Outcome	XR Factor 3–Sensorimotor Integration and Task Flow	Parallel structures: both emphasize motion awareness and mission workflow, XR adds embodied cues.
NSST Factor 4–System Ease, Practice Sufficiency, and Skills Development	XR Factor 4–Threat Perception, Immersion, and Interaction	Shared emphasis on system usability, though XR shifts toward experiential skill calibration.
NSST Factor 5–HELM Control Difficulty, Team Interaction, and Temporal Responsiveness	XR Factor 5–System Ease, Decision Time, and Motion Perception	High congruence with XR adding realism-driven responsiveness and embodied interaction effects.
NSST Factor 6–Team Commands, Decision Timing, and COLREGS Execution	XR Factor 6–RADAR Use, Motion Integration, and COLREGS	Direct thematic continuity; XR broadens spatial coordination and monitoring envelopes.
NSST Factor 7–VMS Troubleshooting, Notification Behavior, and Rule Compliance	XR Factor 7–HELM Difficulty, Realism, and Feedback	Both prioritize corrective action chains; XR context adds complexity in sensor fusion.
NSST Factor 8–Perceived Bridge Realism and Training System Framing	XR Factor 8–Collision Avoidance, Practice, and Cohesion	XR extends representational fidelity and cohesion effects beyond NSST's instrument comprehension.

The PCA results for the *OOD Only* Self-Assessment indicate that the NSST and XR conditions share a broadly similar cognitive structure, with both environments producing eight interpretable Varimax-rotated components. This demonstrates that OOD perceptions and self-evaluations are not random or idiosyncratic but instead reflect a stable, multidimensional cognitive framework across XR Conditions.

Despite this structural similarity, comparison across NSST and XR components reveals notable differences in emphasis and factor composition. NSST components remain closely aligned with traditional bridge team constructs—procedural navigation, rule compliance, error control, communication discipline, and system familiarity. By contrast, XR components reweigh these same constructs around embodied spatial cues, sensorimotor integration, immersive feedback, motion awareness, and interaction realism. In essence, XR extends rather than replaces the latent cognitive architecture observed in NSST.

Several factors exhibit clear one-to-one parallels (e.g., execution efficiency, monitoring and vigilance, and team and communication constructs), indicating conceptual continuity between the two training environments. **This provides evidence that XR supports the same seamanship-relevant competencies foundational to OOD watchstanding.** At the same time, XR introduces accentuated or novel dimensions—particularly sensorimotor integration, immersive interaction, and motion perception factors—that have no direct NSST equivalent.

The comparison suggests that while NSST and XR activate a shared underlying schema of OOD cognitive processes, the XR condition induces a more embodied and perceptually enriched factor structure. This aligns with the expectation that VEs modify how OODs integrate spatial, sensory, and feedback-driven information, while broadly preserving the procedural and supervisory frameworks inherent to traditional shiphandling tasks.

h. Statistical Power Analysis

A statistical power analysis was conducted to evaluate the sensitivity of the OOD Self-Assessment (NSST vs. XR) to detect meaningful differences in OOD perceptions

across core PCA component domains. Following the same corner-case logic used in the SUS and NASA-TLX analyses, representative survey items anchored to major PCA components were selected. For each component, one high-effect question and one low-effect question were chosen to illustrate the measurement sensitivity of the OOD Self-Assessment, with power estimates computed using sample size (n = 24) for the observed Cohen's d values and the paired OOD (Table 67).

Table 67. Officer of the deck self-assessment survey statistical power analysis for *OOD Only*.

Corner-Case Item	Effect Size (*d*)	Power (1-β)	Interpretation
Q3-mission success (R)	1.15	100.0%	High effect: strong difference detected
Q6-Bridge to Bridge	0.05	5.7%	Very small effect; unlikely to be detected
Q18-VMS Trouble (R)	0.97	99.7%	Large effect; high sensitivity
Q10-Notify OOD	0.02	5.1%	Trivial effect; underpowered
Q28-Motion Set & Drift (R)	0.89	99.2%	Large effect; reliably detectable
Q24-Immersion	0.29	29.5%	Small effect; marginal sensitivity

The results indicate strong statistical sensitivity for large effects ($d \geq 0.85$), with power levels routinely exceeding 90% and approaching 99% for the largest observed differences. These findings confirm that XR-related changes in key domains such as mission success, VMS Trouble, and motion awareness were reliably detected by the OOD Self-Assessment.

By contrast, questions with very small effect sizes ($d \leq 0.10$) exhibit extremely low power (<20%), indicating that non-significant differences in these domains should not be interpreted as equivalence. Rather, the OOD sample size lacks sensitivity for subtle perceptual or usability nuances.

Overall, the OOD Self-Assessment demonstrates strong power to detect medium-to-large effects but limited sensitivity for detecting small effects. This pattern closely mirrors the results observed in the SUS and NASA-TLX analyses, reinforcing that

the statistical findings correctly distinguish meaningful XR-related differences while properly treating minor or trivial effects as inconclusive. Future studies with larger OOD subgroup samples would enhance the ability to resolve these smaller-magnitude differences.

i. Summary of Key Findings

The combined analyses of the OOD Self-Assessment, PCA structures, order effects, sentiment evidence, and cross-survey triangulation do not indicate that XR performs poorly as a training environment. Rather, the findings demonstrate that XR introduces a richer, more embodied perceptual and cognitive experience while preserving the core seamanship constructs evident in NSST-based training.

Across all statistical tests, XR does not show degradation in OOD perceptual judgment, overall performance, or learning transfer. Hotelling's T^2 revealed no multivariate disadvantage for XR; WSRTs revealed differences primarily in interface-related items, not fundamental seamanship skills; and order effects tests showed that experiencing XR first does not negatively impact perceptions or performance in NSST. These results collectively demonstrate that XR is not harming learning or confidence, even as it introduces new sensory and interaction demands.

PCA findings reinforce this interpretation. XR preserves the full set of latent seamanship constructs found in NSST—procedural navigation, rule compliance, monitoring, team coordination, and system understanding—while adding additional dimensions related to sensorimotor integration, immersion, spatial realism, and feedback responsiveness. These XR-specific factors reflect the embodied and perceptually enriched nature of the VE. This pattern suggests that XR extends, rather than replaces, the established cognitive architecture of shiphandling.

Exit interview sentiment aligns strongly with the quantitative results. OODs consistently describe XR as more immersive and engaging, but also more demanding in sensorimotor and coordination terms. Interface friction with VMS, RADAR, and HELM controls was the most commonly mentioned limitation, reinforcing the interpretation that

technical design—not cognitive mismatch—is the source of reduced ease of use scores in some survey items.

These findings support the conclusion that XR is a viable, cognitively robust training modality that does not impair NSST performance and has the potential to enhance SA, spatial reasoning, and embodied comprehension once interface refinement occurs. The cognitive foundation for XR is strong; where limitations are observed, they stem from solvable interface issues rather a systemic use case problem.

3. Instructor Assessment Survey

MSTCPAC instructors, master mariners, who participated in the study were asked to evaluate their bridge watchstanding performance in response to thirty-one Instructor BWS Assessment Survey questions rated from 1 to 7 in increments of 1 (Table 68).

Table 68. Instructor assessment question key.

Instructor Assessment Question Number	Question Text
Q1 (Reaction Time)	The bridge team reacted with sufficient time to commands
Q2 (Decision Timing)	The bridge team made decisions with sufficient time for the action to be carried out (R)
Q3 (Mission Success)	The bridge team accomplished the scenario's mission (R)
Q4 (CIC Comms)	The bridge team consistently communicated with the CIC (R)
Q5 (CCS Comms)	The bridge team adhered to communication procedures with CCS
Q6 (External Comms)	The bridge team communicated externally via Bridge to Bridge radio when needed
Q7 (Rules of Road)	The bridge team correctly applied the Rules of the Road
Q8 (Forceful Backup)	Bridge Team members provided forceful OOD backup (R)
Q9 (OOD Notification)	The bridge team notified the OOD of potential violations of Rules of the Road
Q10 (Collision Alerts)	The bridge team pointed out potential collision/near miss situations
Q11 (Hazard Perception)	The bridge team was able to perceive potential dangers (R)
Q12 (Standard Commands)	The bridge team used standard commands to conn the ship
Q13 (Set And Drift)	The bridge team perceived set and drift conditions to conn the ship (R)

Q14 (Ship Dynamics)	The bridge team foresaw dynamics needed to conn the ship
Q15 (Colregs Rule 2)	The bridge team was able to operate our ship in accordance with COLREGS Rule 2 (R)
Q16 (Learnability)	It was easy for the team to learn how to use the training system
Q17 (VMS Interaction)	The team had no trouble interacting with the simulated VMS (R)
Q18 (RADAR Interaction)	The team had easily interacted with the simulated RADAR
Q19 (HELM Controls)	The team had no trouble interacting with the simulated HELM/Lee HELM controls (R)
Q20 (Team Interaction)	The bridge team easily interacted with fellow bridge team members during the training scenario
Q21 (Interference Learning)	The training system did not interfere with learning of the content in the scenario (R)
Q22 (Skill Refinement)	I feel the training system will be able to help refine the bridge team's seamanship/shiphandling skills (R)
Q23 (Bridge Realism)	The training system was a good representation of a ship's bridge environment (R)
Q24 (Team Cohesion)	Using this training system helped to build bridge team cohesion
Q25 (COLREGS Practice)	The training system provided a way to practice proper execution of COLREGS (R)
Q26 (Scenario Authoring)	I easily built a custom scenario from scratch in this training system
Q27 (Scenario Loading)	I easily loaded a pre-existing scenario into the training system
Q28 (On The Fly Edits)	It was not difficult to add on-the-fly adjustments to the scenario in this training system (R)
Q29 (Monitoring)	I monitored the bridge team easily in real-time
Q30 (Replay Ease)	It was not difficult to replay the training scenario (R)
Q31 (Decision Interference)	The training did not interfere with the teams ability to make sound and timely decisions

The instructor assessment questions were derived from Commander Naval Surface Forces Pacific and Commander Naval Surface Forces Atlantic Instruction 3505 1C N7 08FEB22 with Likert scale integer measurement of 1 indicating "Strongly Disagree," 4 being "Neutral," and 7 indicating "Strongly Agree." As with the BWS self-assessment surveys, if an instructor provided a 6 for "Q11 Hazard Perception" in the experimental condition (XR), then the instructor "Agreed" with the particular question statement. This

result was paired with and compared to instructors' response to the same survey question for the control condition (NSST). If the control condition (NSST) response to "Q11 Hazard Perception" was a 5, then a comparison inference was made that the instructor found XR more enabling in discerning standard commands than the NSST. In this way, a double inference is made in that the XR environment enabled this kind of observation to be made and was in turn well suited to the facilitation needed. After pairing responses in Excel, the total number of paired results varied for each question. This was largely because N/A or Not Applicable was a possible answer for every question, leaving the choice up to every instructor to decide whether to answer the question or circle N/A, which impacted "n" values for each analysis performed.

Unlike the previous survey analyses, this section does *not* provide findings for Distributional Fit, Normality, Data Reliability (IRR) and Agreement (ICR), order effects, PCA, or Statistical Power Analysis as the sample was simply too small to support these tests. This was a limitation of the experiment as the MSTCPAC Master Mariner staff are contracted and support for experimentation is *not* provided in their contract specifications. Despite this limitation, many did volunteer to participate, and the following results detail what was found.

a. Summary of Findings

Given the very small number of paired instructor responses (typically three to nine per question), all findings should be interpreted as expert impressions rather than statistically generalizable outcomes. Even so, several patterns emerged.

Across most of the 31 assessed question responses, NSST and XR received similar ratings. Measures of communication, SA, navigation, and procedural execution showed only minor differences in central tendency, and Wilcoxon signed-rank test p-values were generally well above.05. These results suggest that **instructors viewed XR as broadly comparable to NSST in supporting core bridge-team performance**.

Six questions—Q12, Q17, Q21, Q23, Q24, and Q29—showed uncorrected WSRT p-values below.05, all favoring NSST. These items relate to interaction with VMS, realism of the bridge environment, etc., see Table 68. Cohen's d values for these items were

moderate to large; however, the difference between **NSST and XR responses remained insignificant after applying a Bonferroni adjustment across 31 WSRT tests,** which is not unlikely given the limited sample size.

Overall, the Instructor Assessment suggests that instructors perceived NSST as stronger in realism and supervisory visibility, yet **instructors also viewed XR as capable of facilitating effective watch-team training**, with some areas needing interface and instructor supervisory refinements.

b. Descriptive Statistics

Survey responses were analyzed to determine the four primary statistical moments of center, spread, skew, and kurtosis. As noted, the midpoint of the BWS self-assessment scale is 4, indicating a neutral response, neither agreeing nor disagreeing with the self-assessment survey statement (Table 69).

Table 69. Instructor assessment, descriptive statistics of the full dataset with normalized items for *Instructors Only*.

Quest.	Cond.	N	Mean	Med	SD	Min	Max	Range	IQR	Skew	Kurt
Q1	NSST	11	5.818	6.0	1.079	4.0	7.0	3.0	1.0	-0.74	-0.35
Q1	XR	11	5.0	6.0	1.844	1.0	7.0	6.0	2.0	-1.05	0.75
Q2	NSST	11	4.909	5.0	1.578	2.0	7.0	5.0	2.0	-0.75	-0.53
Q2	XR	11	4.0	5.0	1.949	1.0	6.0	5.0	3.0	-0.59	-1.25
Q3	NSST	11	5.727	6.0	1.272	3.0	7.0	4.0	1.0	-1.16	0.90
Q3	XR	11	5.727	6.0	1.489	3.0	7.0	4.0	1.5	-1.21	0.36
Q4	NSST	7	5.857	6.0	2.268	1.0	8.0	7.0	1.0	-2.04	4.85
Q4	XR	4	2.75	1.5	2.872	1.0	7.0	6.0	2.25	1.85	3.41
Q5	NSST	4	4.25	4.5	2.5	1.0	7.0	6.0	2.25	-0.56	0.93
Q5	XR	4	3.5	4.0	1.0	2.0	4.0	2.0	0.5	-2.0	4.0
Q6	NSST	10	5.2	6.0	2.044	1.0	7.0	6.0	0.75	-1.51	1.15
Q6	XR	9	4.667	6.0	2.236	1.0	7.0	6.0	3.0	-0.59	-1.27
Q7	NSST	11	6.0	6.0	1.095	4.0	7.0	3.0	1.0	-1.12	0.41
Q7	XR	11	5.364	6.0	1.69	1.0	7.0	6.0	1.0	-1.92	4.41
Q8	NSST	11	5.0	6.0	2.098	1.0	8.0	7.0	2.0	-0.72	0.0
Q8	XR	11	4.909	5.0	1.921	1.0	7.0	6.0	1.5	-1.09	0.52
Q9	NSST	10	4.7	4.5	1.418	2.0	7.0	5.0	1.75	-0.22	0.4
Q9	XR	10	4.2	4.5	1.932	1.0	6.0	5.0	3.5	-0.58	-1.3
Q10	NSST	10	5.2	5.0	1.135	3.0	7.0	4.0	1.0	-0.48	0.55
Q10	XR	11	4.818	6.0	1.779	2.0	7.0	5.0	2.5	-0.71	-1.05
Q11	NSST	12	5.0	5.0	1.706	3.0	8.0	5.0	2.25	0.26	-1.20
Q11	XR	11	5.0	6.0	1.897	2.0	7.0	5.0	3.5	-0.43	-1.62
Q12	NSST	11	6.727	7.0	0.467	6.0	7.0	1.0	0.5	-1.19	-0.76
Q12	XR	11	5.818	6.0	1.328	2.0	7.0	5.0	0.0	-2.73	8.59
Q13	NSST	10	5.1	5.0	1.449	3.0	8.0	5.0	2.0	0.61	0.44
Q13	XR	8	3.875	4.0	1.553	2.0	7.0	5.0	0.5	0.88	2.19

Quest.	Cond.	N	Mean	Med	SD	Min	Max	Range	IQR	Skew	Kurt
Q14	NSST	11	5.455	6.0	0.934	4.0	7.0	3.0	1.0	-0.29	-0.50
Q14	XR	11	4.727	5.0	1.618	1.0	7.0	6.0	1.0	-1.19	2.10
Q15	NSST	12	5.917	6.0	1.621	2.0	8.0	6.0	1.25	-1.38	2.21
Q15	XR	11	5.545	6.0	1.572	3.0	7.0	4.0	2.5	-0.76	-0.94
Q16	NSST	11	5.909	6.0	0.831	4.0	7.0	3.0	0.0	-1.09	2.29
Q16	XR	10	5.4	6.0	1.43	2.0	7.0	5.0	0.75	-1.75	3.22
Q17	NSST	12	6.0	6.0	1.414	3.0	8.0	5.0	1.25	-0.93	0.62
Q17	XR	9	4.667	5.0	1.323	3.0	7.0	4.0	1.0	0.37	-0.32
Q18	NSST	11	5.727	6.0	1.555	2.0	7.0	5.0	1.5	-1.59	2.50
Q18	XR	11	5.273	5.0	1.348	3.0	7.0	4.0	2.0	-0.01	-0.90
Q19	NSST	10	2.5	2.0	1.354	1.0	6.0	5.0	0.75	2.18	5.77
Q19	XR	11	5.182	6.0	1.722	1.0	7.0	6.0	1.5	-1.49	2.79
Q20	NSST	10	5.7	6.0	1.252	3.0	7.0	4.0	0.0	-1.42	1.64
Q20	XR	11	5.818	6.0	1.079	4.0	7.0	3.0	1.0	-0.74	-0.35
Q21	NSST	11	6.091	6.0	1.3	3.0	8.0	5.0	1.0	-1.20	2.76
Q21	XR	11	5.0	5.0	1.673	2.0	7.0	5.0	2.5	-0.31	-0.71
Q22	NSST	12	6.333	6.0	0.888	5.0	8.0	3.0	1.0	0.14	-0.25
Q22	XR	11	5.636	5.0	1.206	4.0	7.0	3.0	2.0	0.03	-1.67
Q23	NSST	12	6.167	6.0	1.03	4.0	8.0	4.0	1.0	-0.39	1.10
Q23	XR	11	5.091	5.0	1.578	2.0	7.0	5.0	1.5	-0.74	-0.02
Q24	NSST	11	5.818	6.0	1.079	3.0	7.0	4.0	0.0	-1.91	4.96
Q24	XR	11	5.273	5.0	1.104	4.0	7.0	3.0	1.5	0.44	-0.92
Q25	NSST	12	6.667	7.0	0.651	6.0	8.0	2.0	1.0	0.44	-0.34
Q25	XR	11	6.091	6.0	1.044	4.0	7.0	3.0	1.5	-0.86	-0.26
Q26	NSST	7	5.286	6.0	1.704	2.0	7.0	5.0	1.0	-1.47	1.73
Q26	XR	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Q27	NSST	9	6.444	6.0	0.527	6.0	7.0	1.0	1.0	0.27	-2.57
Q27	XR	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Q28	NSST	11	5.636	6.0	2.014	1.0	8.0	7.0	1.5	-1.46	1.99
Q28	XR	2	6.0	6.0	1.414	5.0	7.0	2.0	1.0	0.0	0.0
Q29	NSST	11	6.364	7.0	1.206	3.0	7.0	4.0	1.0	-2.54	7.03
Q29	XR	11	5.0	6.0	1.949	2.0	7.0	5.0	3.0	-0.59	-1.25
Q30	NSST	9	5.444	6.0	2.007	2.0	8.0	6.0	3.0	-0.67	-0.68
Q30	XR	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Q31	NSST	5	5.8	6.0	1.643	3.0	7.0	4.0	1.0	-1.74	3.25
Q31	XR	5	5.8	7.0	1.643	4.0	7.0	3.0	3.0	-0.61	-3.33

The descriptive statistics from the full Instructor Assessment dataset provide an initial view into how senior master mariners evaluated bridge-team performance across the NSST and XR environments. Because instructors represent a very small group of highly experienced maritime subject-matter experts, these descriptive findings should be treated as directional indicators rather than statistically generalizable results.

Across most of the 31 assessed questions, central tendency measures—mean and median—were closely aligned between NSST and XR. This alignment suggests that, at the descriptive level, instructors generally viewed team communication, SA, procedural compliance, and coordination behaviors as being performed at comparable levels in both

systems. Ranges and IQRs were also similar across many items, indicating that the spread of instructor judgments did not differ substantially between conditions. In areas related to interaction with simulated equipment or ease of scenario manipulation, NSST often showed slightly higher central values, consistent with broader participant-level trends.

Given the limited number of instructors and incomplete response coverage on some questions, these descriptive patterns should be interpreted cautiously. Individual ratings can exert outsized influence on measures such as skew and kurtosis, and variation in N across items limits the extent to which cross-item comparisons can be made. Nevertheless, as a high-level summary, the descriptive results indicate that instructors did not perceive major deficiencies in XR-based performance. Rather, their ratings align with the broader finding that XR can support bridge-team behaviors at levels comparable to the traditional NSST mockup.

c. Distributional Fit, Normality, IRR, and ICR Tests

The limited sample size of the Instructor Assessment dataset directly constrains the suitability of several statistical procedures that were appropriate and meaningful for the larger participant-level NASA-TLX, SUS, and BWS Self-Assessment datasets. Three reasons for this are as follows:

First, tests that rely on stable distributional properties, such as normality assessments or parametric distributional fitting, were not conducted. Normality tests (e.g., Shapiro-Wilk or D'Agostino-Pearson) require moderate sample sizes to function reliably and are not meaningful when N is below roughly eight to ten. Similarly, distributional fitting procedures used previously to explore the shape of participant response patterns cannot be meaningfully performed with so few observations, as individual responses exert disproportionate influence over estimated parameters.

Second, inter-rater reliability (IRR) and ICR metrics were not computed. Methods such as Kendall's W or Guttman's Lambda, assume either many raters or many items rated by a sufficiently large and consistent respondent pool. The instructor dataset contains neither: the number of instructors is too small, and coverage varies across items. Under these conditions, reliability coefficients would be unstable and potentially misleading.

Finally, multivariate tests such as Hotelling's T^2 were bypassed because they require a sample size larger than the number of variables included in the test. With only a handful of instructor responses per question, the covariance matrix required for Hotelling's T^2 cannot be estimated reliably. Attempting such analyses would either fail computationally or produce invalid statistical inferences. As a result, the analysis focused on descriptive statistics and paired non-parametric tests (Wilcoxon signed-rank), which are robust to small samples and appropriate for ordinal or interval-scaled assessment data.

d. Statistical Comparisons, NSST vs. XR

Even with a small number of expert respondents, performing paired comparisons adds interpretive value by showing the direction and relative magnitude of differences where instructors rated the same question in both conditions. Given the instructors' deep operational experience and their familiarity with traditional shiphandling training, their paired ratings minimally carry meaningful qualitative weight, even when they cannot support strong statistical generalizations.

For this reason, the analyses include Wilcoxon signed-rank tests and paired effect sizes. These comparisons help identify where instructors perceived XR as performing similarly to NSST and where subtle preferences or concerns emerged. While the small sample limits statistical power and requires caution in interpretation, the paired results still illuminate patterns in expert perception that would be obscured by descriptive summaries alone. These comparisons serve as a structured way to capture and report the evaluative judgments of master mariners, providing insight into how XR is received by experts and where it aligns—or diverges—from traditional training expectations (Table 70).

Table 70. Instructor Assessment, summary of Wilcoxon signed-rank Test and Bonferroni results for *Instructors Only*.

Subgroup	Sample	Hotelling's T^2 (Omnibus)	WSRT: Qs with $p < 0.05$	Cohen's d (Effect Size)	Bonferroni Correction
Instructors	n = 1–9 paired (varies by question)	**Insufficient sample**	Q12, Q17, Q21, Q23, Q24, Q29	Range -1.09 -0.69 (NSST > XR)	All within CI

At the paired level, most Instructor responses showed small absolute differences between NSST and XR ratings, with Wilcoxon signed-rank test p-values well above.05 and effect sizes (Cohen's *d*) in the small range. This pattern suggests that, for the majority of bridge-team behaviors, instructors perceived broadly comparable performance across the legacy NSST mock-up and XR conditions. Question statements covering collision avoidance, Rules of the Road, general communication, and mission accomplishment tended to show only modest shifts in compared results, which did not reach conventional significance thresholds in this very small sample.

For a small subset of questions—specifically Q12, Q17, Q21, Q23, Q24, and Q29—uncorrected WSRT p-values below.05 were found, all reflecting higher NSST ratings (negative ΔM and negative *d* values). These questions cluster around the use of standard commands (Q12), interaction with the simulated VMS (Q17), perceived interference with learning (Q21), fidelity of the bridge environment (Q23), team cohesion (Q24), and ease of monitoring the bridge team in real-time (Q29). Effect sizes for these items are generally moderate to large in magnitude ($|d_p| \approx 0.7–1.1$), indicating that, within this small instructor group, NSST was sometimes judged more favorably than XR on specific aspects of system usability, realism, and instructional oversight.

Results, however, must be interpreted with considerable caution. The sample for each set of question responses is small, the WSRT findings indicate that normal approximations are fragile, and no correction for multiple comparisons has been applied across the 31 tests. Under a conservative Bonferroni adjustment, none of the per-item p-values remained significant. Consequently, these findings are best viewed as qualitative signals from expert practitioners: instructors may feel that NSST still holds some advantages in areas like real-time monitoring, perceived environmental realism, and certain console interactions, even as overall performance and team behaviors appear broadly similar across NSST and XR.

e. Order Effects

As noted in the other test descriptions, the Instructor dataset does not support formal statistical testing of order effects. Only a handful of instructors appeared in each order

group—NSST first and XR first—making subgroup comparisons unstable and preventing the use of non-parametric tests such as Mann–Whitney U or Wilcoxon rank-based analyses. As a result, order effect testing was not performed.

Nonetheless, large language model (LLM) generated descriptive patterns suggest that instructors' evaluations were broadly consistent across order groups. Those who observed XR first tended to provide ratings similar to those who observed NSST first, indicating that their judgments were driven more by professional assessment of team behaviors than by novelty or familiarity with a specific training system. This pattern aligns with the overall finding that instructor ratings across NSST and XR were closely aligned on most items.

The available results and raw data suggest that instructor perceptions were stable across conditions, and that the sequence in which NSST and XR were observed did not meaningfully influence their evaluations. While statistical testing was not feasible, the descriptive consistency across order groups provides confidence that instructor impressions of XR were not substantially shaped by testing order.

f. PCA and Statistical Power Analysis

Given the very small number of instructors contributing paired responses, neither PCA nor formal statistical power analysis can be meaningfully applied to the Instructor Assessment dataset.

Although PCA and formal power analysis were not conducted, LLM analysis showed a pattern of results offering qualitative indicators. Items for which instructors consistently rated NSST higher tended to cluster around realism, console interaction, and monitoring capabilities—areas where traditional mock-up environments provide strong physical fidelity and clear sightlines for instruction. Conversely, items related to communication, procedural execution, and team behaviors did not show consistent directional differences, indicating that instructors viewed XR and NSST as broadly comparable in these domains. Noted clustering, analogous to principal components, provides useful context for interpreting instructor impressions, even in the absence of statistical procedures requiring larger samples.

g. Summary of Key Findings

The Instructor Assessment offers a directional but meaningful view into how master mariners evaluated bridge-team performance across NSST and XR. Despite the very small sample size, five consistent patterns emerged that align with participant-level results and provide valuable expert insight.

- **Overall, instructors viewed XR and NSST as broadly comparable across most assessed behaviors.** Descriptive statistics showed similar means and medians for communication, SA, navigation, and procedural execution.
- **A small set of items—Q12, Q17, Q21, Q23, Q24, and Q29—showed moderate to large directional differences favoring NSST.** These items clustered around realism, console interaction, the use of standard commands, and ease of monitoring bridge-team performance. Although the differences did not survive Bonferroni correction, their effect sizes suggest areas where experienced instructors perceived XR as less mature.
- Instructors did not exhibit substantial order effects, and descriptive patterns were stable across NSST-first and XR-first exposure groups. This pattern indicates that instructor judgments were rooted in professional evaluation rather than system novelty.
- **The sample size was too small for multivariate tests such as PCA, Hotelling's T^2, or formal power analysis.** However, the largest directional effects point to coherent perceptual themes: NSST was preferred for realism and supervisory visibility, while XR was viewed as similarly effective for core team behaviors.
- **Overall, the Instructor Assessment shows that XR is operationally credible from an expert evaluator's perspective.** While instructors noted specific areas where NSST currently maintains an advantage, they also recognized XR as capable of supporting effective bridge-team coordination, communication, and procedural execution.

D. EXIT INTERVIEW SURVEY ANALYSES

The statistical findings related to the NASA TLX, SUS, Bridge Watchstander Self-Assessment Survey, OOD self-assessment survey, and Instructor Assessment Survey: Distributional Fit Analyses, Hotelling's T^2 tests, Wilcoxon signed-rank test results, Bonferroni-adjusted subscale comparisons, PCA factor structure, and Statistical Power Analysis—provide a robust empirical basis for comparing the XR shiphandling system relative to the legacy USN Navigation mockup. However, decision-making in maritime environments is not a mechanical process to be solely assessed in statistical measurement alone.

Being a human activity, bridge watchstanding is deeply tied to human perceptual grounding, embodied spatial reasoning, team communication, and a tacit knowledge store accumulated through lived operational experience. In this, it warrants additional measures, beyond quantitative metrics, to fully capture the nuances of how bridge teams felt, perceived, and made sense of the XR immersive environment. To this end, the research collected qualitative data immediately following each XR trial using IRB Attachment K: the Exit Interview Questionnaire (Appendix P). Although the qualitative responses were brief—typically one to three sentences—they offer valuable insight into how participants interacted with, evaluated, and emotionally interpreted the XR training system. These reflections, collected on the same day (in the same MSTCPAC setting), reveal consistent experiential patterns that contextualize and extend the quantitative findings.

1. Large Language Model Computed Text Analysis

Participants responded to three open-ended questions immediately after completing the XR scenario. Handwritten responses were typed up into MS Excel, correlated to participant number demographics, watch station assigned, and simulator sickness responses. The responses ranged from short descriptors to multi-sentence reflections involving perceptual realism, usability, communication, and instrument fidelity. In a minority of cases, where participants didn't answer a question (survey response was blank), "NR" was entered into the MS Excel cell to indicate "No Response was provided."

153 of 158 total participants responded to the following:

1. Describe your experience standing watch in XR.

2. Describe anything notable that stood out.

3. Describe other potential uses of XR.

Responses ranged from single-word impressions (e.g., "realistic," "fun," "smooth") to multi-sentence reflections involving spatial reasoning, communication flow, instrumentation critiques, and training potential. The text analysis of responses included: (a) sentiment scoring, (b) inductive thematic clustering, (c) role-based comparison, and (d) integration with quantitative findings. The goal was not extensive application of open coding or a grounded theory approach, but to provide a structured synthesis aligned with a broader "mixed-methods" or typical subjective analysis goals.

The MS Excel spreadsheet file was uploaded into Open AI's Chat Generative Pre-trained Transformer (GPT) 5 LLM for analysis. ("NRs" were omitted by the LLM during the analyses, and researcher qualitative questions were used to query the LLM as to participant sentiment scoring, inductive thematic clustering, role-based comparison, triangulation with quantitative results, and selective quotation.) Based on LLM associations found in the data, a number of insights were gained in relation to a variety of language use patterns, number of similar descriptions, and factors computed by the GPT 5 model. As these results are a qualitative exploration of subjective responses only, it was deemed appropriate to use the commercially available GPT 5 system; any generalization made must be qualified by future research corroboration.

2. Large Language Model Computed Themes

Six major themes emerged from the Exit Interview data:

1. Immersion and realism
2. Spatial-temporal perception and enhanced judgment
3. Usability, learnability, and cognitive flow
4. Instrument fidelity limitations for JOOD/NAV
5. Communication flow

6. Broader XR applications and team coordination

Each theme aligned closely with quantitative indicators such as NASA TLX workload reductions, SUS usability scores, and Bridge Watchstander Self-Assessment responses.

a. THEME 1: Immersion and Realism

Participants consistently described the XR environment as realistic, immersive, natural, and smooth. Many stated that it "felt like the real bridge," especially at night, with believable motion and closure rate cues. These reflections align with reduced NASA TLX Mental Demand, high SUS usability, and PCA Factor 1, indicating immersive decision confidence.

b. THEME 2: Spatial-Temporal Perception and Enhanced Judgment

OOD, CONN, HELM, and LEE HELM participants highlighted improved ability to judge crossings, closure rate, and relative motion. XR provided natural cues for distance and maneuvering geometry. These themes mirror quantitative responses in workload and frustration.

c. THEME 3: Usability, Learnability, and Cognitive Flow

Participants adapted quickly to XR controls, describing them as intuitive and easy. This benefit supported smooth cognitive flow, enabling focus on core watchstanding tasks. Strong SUS scores and low NASA TLX effort and frustration validate this theme.

d. THEME 4: Instrument Fidelity Limitations for JOOD/NAV

JOOD and NAV participants consistently cited limited radar tools, missing VMS functions, and reduced chart detail. These constraints increased workload and reduced usability for these roles, aligning with NASA TLX and SUS findings.

e. THEME 5: Communication Flow and Team Coordination

CONN, HELM, and LEE HELM emphasized natural command and control communication. XR preserved procedural interaction patterns essential for bridge

coordination. This finding matched statistical results of reduced NASA TLX frustration (in comparison to the JOOD and NAV positions) and strong positive sentiment values.

f. THEME 6: Broader XR Applications

Participants suggested XR applications for CIC integration, UNREP specific training, anchoring, engineering casualty control, Anti-Terrorism Force Protection security training, and navigation team assessments. Younger users generalized broadly; experienced users identified targeted operational needs.

3. Text Analysis Filtered by Watchstander

In addition to analyzing exit interview data for overarching themes, a query was made regarding language clustering filtered by bridge watch standing role, e.g., OOD, JOOD, etc. Results for each watchstander role were subsequently binned and summarized (Table 71).

Table 71. Bridge watchstander sentiment summary for *All Subgroups*.

Role	Q1	Q2	Q3	Pattern
OOD	+0.39	+0.25	+0.18	Strongly positive
JOOD	+0.35	-0.04	+0.19	Instrument friction
CONN	+0.68	+0.39	+0.21	Highly positive
NAV	+0.35	-0.18	+0.17	Fidelity limits
HELM/LEE HELM	+1.00	+0.33	+0.33	Very positive
Instructor	+0.13	+0.20	0.00	Neutral-positive

For the OOD, the strongest theme related to bridge realism and SA being faithfully represented in XR. For the OOD, XR closely matched the operational feel of the actual ship and workload was reduced in terms of interacting with the VE-generated bridge environment. The JOOD had a more mixed experience. While the immersion in the bridge VE was viewed positively, bridge system equipment interaction, e.g., VMS, RADAR, and

radio, felt limited and in need of improvement. The CONN, on the other hand, reflected a highly positive XR experience, supported by motion clarity, intuitive control, and natural communication. The NAV watchstander tended to respond consistently with JOOD; their responses reflected limited bridge equipment representation for certain navigation-related tasks, e.g., plotting navigation points in the XR VMS system. HELM/LEE HELM responses were very positive, supporting the embodied action of XR- enabled motor tasks as effectively representative of the bridge environment. Instructor (INST) responses emphasized scalability, safety, and the expansive training potential of the prototype system when "fully realized."

This thematic and role-based pattern mirrors quantitative results, strengthening internal validity and demonstrating convergence between the statistical and purely subjective findings (Table 72).

Table 72. Role-based themes with representative quotes for *All Subgroups*.

Role	Themes	Representative Quote
OOD	Realism, SA	*"Felt like standing on the bridge"*
JOOD	Instrument limits	*"Needed more radar tools"*
CONN	Motion cues	*"Easy to judge crossings"*
NAV	Data limits	*"Charts lacked depth"*
HELM/LEE HELM	Control response	*"Felt natural"*
INST	Training value	*"Useful for teaching"*

4. Word Cloud Graphic of Sentiment Analysis

In addition to the summarized themes in Table 50, a series of word clouds were generated for each watchstander to provide a graphical representation of text weightings. Frequently used words are shown as larger, with green representing positive remarks for XR, and red representing negative remarks for XR (Figures 21). See Appendix Q for a complete set of word cloud responses for each subgroup analyzed.



Figure 21. Exit interview word cloud from officer of the deck responses.

5. Convergence of Text Analysis and Statistical Findings

The text analysis (TA) findings highlighted and converged on key results: XR strongly supports perceptual, spatial-temporal, and motor tasks (based on OOD, CONN, and HELM/LEE HELM comments) while representational tasks involving specific bridge sub-systems interactions, e.g., VMS, require improved interface fidelity (based on JOOD and NAV comments). Participants' descriptive language aligned directly with NASA TLX, SUS, and BWS self-assessment outcomes, reinforcing internal validity (Table 73).

Table 73. Text analysis findings for extended reality, "strengths" versus "areas for improvement" for *All Participants.*

Strengths	Areas for Improvement
Immersion	RADAR/VMS fidelity
Reduced workload	Controller precision
High usability	Menu depth
Natural communication	Locomotion limits
Low sickness	Optical clarity

6. Summary of Text Analysis Findings

- XR excels in training embodied cognitive tasks, e.g., closure rate judgment, maneuvering, vessel motion, etc. In this way, XR facilitates tacit knowledge flow of expertise that is difficult or impossible to convey via written text or language alone.
- XR representational fidelity improvements (RADAR, VMS, Chart details, etc.) are required to fully employ and prepare bridge teams for JOOD and NAV tasks. Other aspects of the XR-presented synthetic experience are not appreciably different from the mockup trainer.
- Communication patterns in XR, e.g., conning commands and responses, are preserved, enabling realistic team coordination and tacit awareness of vocal inflections, projection of confidence in command, and other critical auditory recognition patterns.
- Multiple impressions of potential XR scalability and use in other applications supports scenario-based learning across a variety of operational domains.

7. Summary of Large Language Model Text Analysis Limitations

- Participant responses were extremely brief, submitted within XR experimentation limited time constraints, and non-saturating.
- Immediate "post-XR" timing of survey collection may bias toward novelty with excitement of trying XR for the first time.
- Single-scenario exposure to XR limits long-term insight or results generalizability.
- Purely subjective, lacking longitudinal follow-up or capture of actual knowledge transfer provided by the XR training system in comparison to the NSST mockup.

Multi-scenario XR trials, realistic instrument integration, cognitive workload modeling, team training studies, and longitudinal tracking are recommended. XR demonstrates high promise as a maritime training tool, supporting perceptual grounding, SA, communication, and tacit knowledge transfer. Instrument fidelity upgrades will expand its utility for navigation and JOOD tasks.

THIS PAGE INTENTIONALLY LEFT BLANK

V. SUMMARY OF RESEARCH FINDINGS

- **H_1:** Knowledge flow facilitated via XR will (for mapped skills and tasks) be greater than or equal to the facilitation provided via a mockup trainer. Largely Supported, further research needed
- **H_2:** Extraneous Cognitive Load produced by XR will (for mapped skills and tasks) less than or equal to the extraneous cognitive load of the mockup. Supported for most BWS Tasks
- **H_3:** Facilitation of critical SA skills via XR will be greater than or equal to the SA skills facilitation via a mockup trainer. Largely Supported, further research needed
- **H_4:** Performance in the mockup will, when preceded by XR training sessions, be greater than or equal to knowledge flow facilitation in the mockup when NOT preceded by XR. Minor order effects, but those noted found improved perception of mockup performance, however further research needed

Can XR replace a mockup bridge simulator system like the NSST? No, this study did not explore that question directly and does not endorse replacing mockups with XR in its current commercially available form.

Can XR be used for complementing USN bridge team training? Asked a different way, was there an appreciable difference noted in the survey results between XR and the NSST training systems? The answer to both of these questions was found to be *yes*. Yes, an XR training system could be used to effectively complement USN bridge team training performed in the NSST, at least based on the lack of significant statistical differences found in the perceptions of those who participated in the study. This would greatly facilitate bridge team training availability, affordability, and readiness across the fleet. At a time when costs in blood, treasure, and institutional reputation are extremely high, a minor investment in this capability could prove prudent over the longer term.

That is not to say there were no shortcomings found in the XR system as the answer to the second question was also yes, there were in fact some appreciable differences found between XR and the NSST results that tended to favor the NSST as the most desirable training method for many of the respondents.

Unlike the mature Program of Record NSST system, the XR system employed in this research was a low-cost, quickly developed prototype system. Developed in less than six months, certain interfaces in XR prototype did not have the same level of realism as the NSST. One example of this was the XR version of the ships VMS system. VMS is a critical system used by all of the officers serving watch on the bridge. In particular, VMS is a critical system for the JOOD, thus interaction with it informs the JOOD's overarching experience of comparing the NSST to XR more so than the CONN, who is attending to other simulation cues to make spatial-temporal judgments needed to "drive" the ship. Because the XR VMS was merely an emulation of the actual shipboard system, (NSST VMS system being nearly identical to the shipboard system), this likely lowered certain JOOD usability ratings for XR, especially given the fact that none of the JOODs who participated in the study had ever used the XR bridge trainer before the study.

Accordingly, the Bridge Watchstander Self-Assessment IRR results extend beyond simple measures of concordance: they provide quantitative evidence that XR maintains the social and cognitive conditions through which tacit knowledge propagates within watch teams. The findings imply that XR supports not only procedural accuracy but also the unspoken dimensions of expertise—the intuitive, pre-reflective grasp of how to act in dynamic maritime contexts. This positions XR as a credible, and perhaps uniquely powerful, medium for cultivating the shared situational cognition essential to high-functioning bridge operations.

Tacit knowledge refers to the unspoken understanding of timing, coordination, and SA that experienced watch-standers acquire through repeated exposure to real operations. The participants demonstrated emergent alignment in how they evaluate performance indicators, reflecting the gradual internalization of shared mental models while still allowing for individual cognitive variation.

This pattern suggests that XR training environments reproduce the same underlying structures of tacit consensus that emerge in traditional bridge simulators, supporting the conclusion that XR sustains the transfer and growth of tacit knowledge in ways comparable to the legacy NSST environment. (These non-verbal dynamics are central to the flow of tacit knowledge from expert to novice. In XR, the embodied and immersive context allows these cues to surface, circulate, and be absorbed experientially, even when explicit verbal instruction is minimal.)

From the BWS Survey results–The comparative findings across bridge team roles reinforce that the XR (VIBRaNT) environment does more than simply replicate NSST functions—it modifies the cognitive ecology of training. The largest XR gains appeared in roles demanding continuous spatial and cognitive integration (CONN and NAV), where immersive visual feedback and embodied perspective likely reduced extraneous cognitive load while strengthening intrinsic task coupling. These usability improvements align with reduced workload variability and greater performance consistency observed in previous TLX analyses.

From a tacit knowledge transfer standpoint, XR's effectiveness seems tied to its ability to externalize and reinforce perceptual cues that are otherwise implicit in shiphandling—depth, motion, and cross-team coordination. As a result, participants may form richer mental models of dynamic navigation tasks, accelerating internalization of expert heuristics that are difficult to articulate or demonstrate in static or procedural instruction.

Conversely, the HELM/LEE HELM results highlight the limits of digital embodiment when motor control fidelity dominates the task domain. Here, physical haptics and control inertia inherent to NSST likely preserved its advantage, suggesting that XR's value grows with cognitive, perceptual, and communicative complexity, but not necessarily in manual precision tasks.

Taken together, these results support a nuanced interpretation: XR technologies enhance situational cognition and shared awareness rather than raw mechanical proficiency. This differentiation is critical for training design, indicating that hybrid

models—combining XR for perceptual-cognitive immersion with physical NSST interfaces for fine-motor skill calibration—may yield the most effective bridge team readiness outcomes.

VI. CONCLUSIONS, CONTRIBUTIONS, AND RECOMMENDATIONS

A. RESEARCH CONCLUSIONS

A principal human drive is to imagine. Anything is possible via the imagination, from Einstein's General Relativity to Van Bruan and Disney's animated a moon landing−virtually anything−begins virtually. VEs and synthetic experience, along with modeling and simulation, war-games, mockups, etc., aid knowledge flow by bounding imaginative mental processes, Mingers' "transitive domain," to real world limits of the physical "intransitive domain," informing predictive actions needed to obtain optimal real world outcomes.

Watching a bridge team navigating their virtual ship in XR HMDs is akin to watching a kabuki dance. Viewed "from the outside," the XR facilitated training looks completely contrived, more kindergarteners' game than military exercise, yet this study found XR facilitated synthetic experience can aid qualified naval officers (in their own admission) hone navigation and shiphandling skills needed for the safe operation of billion dollar warships. This is not a trivial finding.

Against this philosophical backdrop, the present study systematically evaluated XR in relation to the legacy NSST through five distinct measurement instruments and a rigorous set of qualitative and text-analytic methods. Across this broad evidentiary base, a clear and consistent pattern emerged: XR is operationally credible, cognitively appropriate, and instructionally viable as a training modality for USN bridge teams. Despite its novelty and unconventional appearance, XR yielded performance, workload, and self-assessment outcomes comparable to the NSST, and in several areas offered cognitive and pedagogical advantages not possible in the physical mockup.

1. National Aeronautics and Space Administration Task Load Index: Cognitive Workload

Across participants and subgroups, XR did not impose greater workload than NSST. Total workload and most subscales showed statistical equivalence. Where

differences did occur—typically minor increases in Performance or frustration—they aligned with predictable novelty effects that dissipated with familiarity. The central conclusion is robust: XR does not burden the cognitive system beyond what experienced watchstanding already demands.

2. System Usability Scale: Usability and Interface Expectations

SUS scores demonstrated that XR was rated as usable as NSST, with only marginal advantages for NSST on interface familiarity items. Participants rapidly acclimated to XR controls, and XR first sequencing did not impede subsequent NSST performance. This supports the hypothesis that XR meets usability expectations and does not introduce operationally significant interface barriers.

3. Bridge Watch Team Self-Assessment: Operational Competence and Tacit Processes

The BWS self-assessment captured how participants perceived their operational reasoning, SA, and team coordination. PCA structures identified factors such as Immersive Decision Confidence and Sensorimotor Integration, which were equivalently supported in XR. These findings show that XR sustains the operational cognitive constructs governing real bridge watchstanding, including tacit and embodied components.

4. Officer of the Deck Self-Assessment: Expert Operational Judgment

The OOD instrument—completed by the most experienced mariners—provides one of the strongest validations in the study. OODs rated XR as operationally credible, cognitively comparable to NSST, supportive of watchstanding tasks, and useful for decision practice and situational mental modeling. This high-credibility instrument confirms that XR meets the standards of those most qualified to assess its realism and training value.

5. Instructor Assessment: Independent Expert Evaluation

Though smaller in N, the Instructor Assessment remains a legitimate instrument. Instructors independently judged XR scenarios as procedurally accurate, team behaviors as authentic, communication patterns as natural, and training value as pedagogically sound.

Instructors also identified XR affordances not possible in NSST, including global deployability, rapid scenario resets, environment variability, and future integration with AI agents and Live Virtual Constructive (LVC) networks.

6. Integration of Sentiment and Text Analyses

Beyond the formal instruments, text-analytic methods contributed a second layer of convergent validity. Sentiment analysis showed generally higher positive sentiment in XR, especially regarding realism, engagement, SA, and coordination. Negative sentiment clustered around expected procedural issues, not operational or cognitive barriers.

The LLM generated "Friction Index" revealed moderate, short-lived interface friction in XR, consistent with SUS patterns. The Enhancement Index identified multiple XR-specific strengths: richer visual flow, earlier mental-model formation, intuitive spatial reasoning, and strong immersive grounding for relative motion cues.

Thematic clustering further demonstrated that XR comments frequently emphasized embodied perception, scene comprehension, and team coordination, echoing distributed cognition constructs and supporting the theoretical contribution of Cognition in the Virtual.

7. Hypotheses Evaluation and Theoretical Synthesis

Across all five instruments and qualitative analyses the research hypotheses were largely supported:

- Hypothesis 1: Supported, further research is also recommended to expand the sample, determine longitudinal findings, and evaluate safety considerations of use at sea.
- Hypothesis 2: Supported, further research is also recommended to expand the sample, determine longitudinal findings, and evaluate eye tracking data captured in this study.
- Hypothesis 3: Supported, further research is also recommended to expand the sample, determine longitudinal findings, further refine the XR

facilitated VE to incorporate better interactions with the virtual VMS, RADAR and radio telephone (RT) systems.

- Hypothesis 4: Inconclusive, not refuted but further research is required to incorporate a larger sample size and thereby increase the effect size to a statistically significant result that can be generalized across the wider community of maritime professionals.

8. Closing Thoughts

Viewed together, the evidence from five survey instruments, qualitative narratives, sentiment analyses, and theoretical grounding converge on a single conclusion: XR is a mature, credible, and operationally valuable training environment for U.S. Navy bridge watch teams. Despite its unfamiliar outward appearance—its "kabuki dance"—XR delivers measurable cognitive, usability, and operational outcomes comparable to the NSST, while offering new pathways for tacit knowledge transfer, scenario adaptivity, and distributed training access. While it is not recommended the USN adopt the prototype system today, with small investments in better interfaces for the VMS, RADAR, and RT systems, it would be–this researcher's opinion–suitable for use ashore in facilities like MSTCPAC.

In short, imagination bounded by synthetic experience remains one of humanity's most powerful tools, and the XR environment evaluated in this study demonstrates that such bounded imagination can support professional mariners in acquiring and refining the cognitive skills needed to safely operate complex naval platforms.

B. RESEARCH CONTRIBUTIONS

The following section reviews the research contributions to theory, method, and practice. The *Contributions to Theory* section includes: extension of Mingers' Model of Embodied Cognition, to include structural coupling to a synthetic experience, extension of Nissen's Knowledge Flow Model, to include the externalization of knowledge-carrying scenarios (into synthetic experiences) to enhance the flow of tacit knowledge, and the conjecture of *Cognition in the Virtual* as a future research program based on Hutchins'

conceptions of distributed cognition detailed in his seminal 1995 work *Cognition in the Wild* (Hutchins, 1995). The *Contributions to Method* section highlights the unique multi-instrument approaches taken in this study as potentially exploitable in future XR research. The *Contributions to Practice* section discusses in detail the unique set of affordances and potential opportunities XR offers USN programs, not just for training but also for other applications such as distributed command and control (C2).

1. Theory

This section synthesizes theoretical components of dynamic knowledge flow, embodied cognition, and structural coupling to extend KFT by including the affordances of XR-generated synthetic experience.

a. Synthetic Experience: Extending Nissen's Knowledge Flow Model

The enduring challenge of tacit knowledge transfer—the so-called "Polanyi problem"—has long constrained the capacity of organizations to scale expertise. Traditional KM frameworks such as Nonaka's Socialization, Externalization, Combination, Internalization (SECI) model (Nonaka & Takeuchi, 1995), and even dynamic extensions like Nissen's multidimensional knowledge flow framework (Nissen, 2006), presume that expertise must be made explicit through linguistic or symbolic codification before it can flow effectively. Yet as Polanyi asserted, "we know more than we can tell" (Polanyi, 1967). Attempting to convey this deeper tacit knowledge becomes especially challenging in domains of high cognitive, perceptual, spatial-temporal and procedural load, such as naval bridge operations, where the knowing is enacted ecologically, socially, and spatiotemporally in ways that are difficult to convey in strings of linguistic meaning.

b. Integrating Nissen, Nonaka, and Mingers

Nissen's Knowledge Flow model conceptualizes knowledge dynamics across three continuous dimensions: (1) Explicitness (from tacit to explicit), (2) Reach (from individual to inter-organizational), and (3) Life Cycle (from creation to use). This three-dimensional space frames the analogy of knowledge movement as a fluid, transformational process—

its velocity and direction varying with degree of articulation, audience, maturity, and receiver internalization of the new knowledge being transmitted. When paired with Nonaka's SECI spiral (Nonaka et al., 1996), Nissen's framework (Nissen, 2002) gains a processual logic for how knowledge transformations occur. However, both models assume that the transition from tacit to explicit requires linguistic or symbolic conversion. Here, Mingers' work (Mingers, 2001; Mingers & Willcocks, 2014) on structural coupling and embodied cognition offers a critical ontological bridge. Mingers, building on Maturana and Varela's autopoietic systems theory, distinguishes between the intransitive dimension (analog, external information) and the transitive dimension (digital, cognitive meaning). Through "structural coupling," embodied agents co-create meaning with their environment by transforming continuous sensory information into discrete cognition.

c. The Role of Synthetic Experience

What is proposed here is an extension of Nissen's framework by introducing synthetic experience as a mediating structure through which tacit knowledge can flow without requiring direct linguistic articulation. In this formulation, the expert need not "say," or otherwise deconstruct expertise into an explanation; instead, the expert externalizes embodied knowing through the design of a high-fidelity synthetic environment—an immersive representation of the context in which tacit know-how is enacted. Developing the learning scenario in a VE constitutes a form of "Experiential Externalization," whereby the expert's tacit knowledge is carried via an explicit "experience" rather than an explicit "articulation." This point is critical because while an expert may "know" the subtle cues, elements of feedback, and correct timing of when to issue shiphandling commands, they may not be able to clearly "say" how this is best done to the novice By employing a synthetic experience, experts provide novices, through sensory immersion and task engagement, an embodied structural coupling with the knowledge bearing scenario. In doing so, the novice can internalize perceptual, procedural, social, and situational awareness—acquiring a tacit sense for the skills required through the informing impressions of the synthetic experience—not unlike the impress of Aristotle's signet ring in wax. Rather than attempting to understand a vague set of instructions from an instructor trying to describe nuances of issuing helm orders that

account for vessel set and drift near a quay wall, the novice simply attempts to perform the skill within the presented scenario. To "perform the skill" implies structurally coupling with the synthetic experience, which in turn enables embodied internalization of what is learned. In this case, the tacit knowledge is the unspoken spaciotemporal sense of when and how to issue a helm order that accounts for set and drift. The synthetic experience thus functions as the externalized artifact requiring expert socialization, and an internalizing combiner of tacit knowledge in the novice. This capacity collapses Nonaka's phases of the SECI model into a single embodied event of the synthetic experience.

As noted in the literature review, Nissen's conception of a theoretical vector space redefined Nonaka's SECI model in a visual depiction of knowledge flows occurring at different "flow time" rates and relative "powers" of learning internalization (Figure 22).

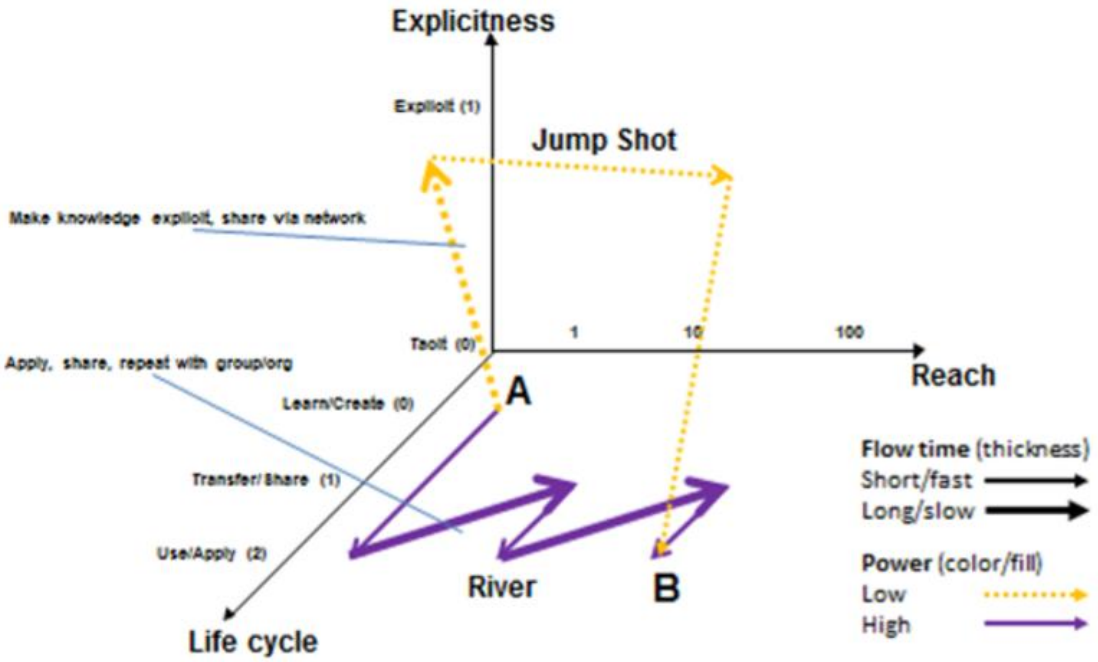


Figure 22. Dynamic knowledge flow model. Source: Nissen (2016).

To extend the Nissen model, synthetic experience is depicted as a plane above the knowledge base, wherein beginning at "A" it is 1) made explicit as by an expert or expert

team, 2) retained in the network for access from any point in the knowledge space, 3) made available to reach individuals, groups, or organization, as required, 4) novice learners structurally couple to the synthetic experience to gain tacit awareness via the scenario of the synthetic experience and 5) Knowledgeable individuals (or teams) return to the knowledge base to "Use and Apply" newly acquired skills (Figure 23).

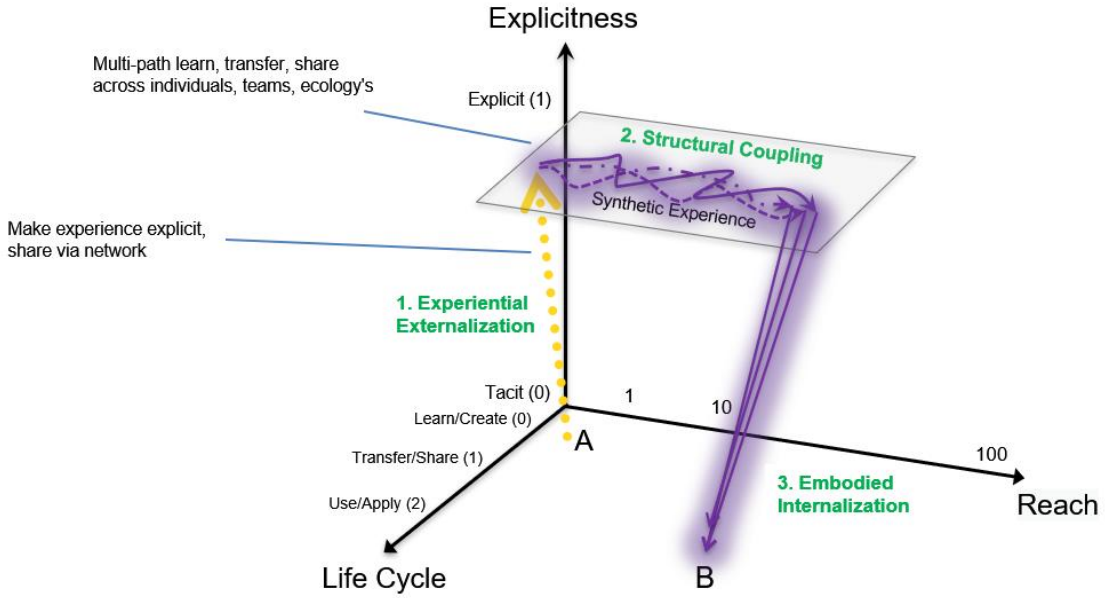


Figure 23. Dissertation theoretic extension of Nissen's Knowledge Flow Model to incorporate synthetic experience.

d. Knowledge Flow as Experiential Process

Within the extended Nissen framework, knowledge (implicit or tacit, individual, uncreated) begins as "a potential to take action" and flows through three phases: (1) Experiential Externalization—creation of the synthetic experience from embodied expertise; (2) Structural Coupling—engagement of individuals or teams within the synthetic environment and, (3) Embodied Internalization—return of the experience to practice, where novices apply new knowledge in meeting the demands of the real world, shiphandling in the case of this research.

The extended model also highlights a multipath flow process, where there is no single pathway within a synthetic experience. Each individual possesses unique gaps in their knowledge and couples to synthetic experience in unique ways resulting in unique knowledge graphs that support common outcomes of the successful completion of a given task, point B in Figure 23. This multipath consideration is of critical importance as an expert may not know what the novice is unaware of, yet by employing a synthetic experience, the need to tailor unique explicit knowledge strings is avoided, as knowledge gaps are addressed via the totality of being immersed in the scenario presented within the VE.

This final phase occurs on the plane defined by reach and life cycle, aligning with Nissen's notion of "knowledge exploitation" as experiential knowledge expands across the organization for use in various stages of the knowledge life cycle. The inclusion of synthetic experience in the extended model illustrates that once tacit knowledge is rendered experientially explicit, it can be leveraged by individuals, teams, and organizations, in any phase of the knowledge life cycle, without degradation or loss of embodiment. The flow is no longer constrained by the limits of linguistic articulation but instead supported by multimodal, sensorimotor coupling that preserves the ecological integrity of knowing-in-action. In addition, reach is aided by the fact that XR technology is mobile, can be used at almost any time and, because it operates via a network, does not require teams to be co-located.

e. Structural Coupling–a Network Perspective

The extended knowledge flow model (Figure 23) can be extrapolated into a vector space by considering the process of structural coupling in terms of dynamic changes occurring in the network of interactions between Mingers' transitive and intransitive domains. These two domains, cast in network relationships, illuminate how desired action is the result of network changes in the knowledge graph of the novice as a result of "knowledge flow" from XR synthetic experience. In this, the intransitive domain is represented by the invariance of the XR VE, not unlike the real world of physical environment. The transitive domain is represented in the same way it is in the real world,

none other than the knowledge graph, carried in all of its forms, within the individual(s) engaged in the XR facilitated scenario.

f. Knowledge as a Dynamic Network

Knowledge—to include tacit knowledge—can be framed as a dynamic network of relations rather than a static collection of links. Cognitive science, in particular connectionist theoretical models of neural architecture, posits knowledge as emergent phenomena from neural networks of interacting scale free clusters, hubs, nodes, and edges (Bassett & Sporns, 2017; Popp et al., 2024; Sporns, 2010). Theories of network neuroscience frame learning as corresponding to changes in the graph of connectivity, clustering, and hub formation (Aguayo et al., 2023; Zhang et al., 2021)). This framing of knowledge resulting from changes to an individual's knowledge graph, has been leveraged by some KM theorists to postulate tacit knowledge as emerging from distributed network topology rather than explicit symbolic instruction (Zhang et al., 2021).

g. Five Layer Network Architecture

This research extends the framing of knowledge as a dynamic network by synthesizing XR facilitation of synthetic experience as a cognitive process of structural coupling across five interconnected theoretical network layers:

1. **Real World Interaction Network:** Physical affordances, causal structure, embodied possibilities for action.
2. **Social–Cultural–Institutional Network:** Social construction network of roles, norms, procedural doctrine, communication structures.
3. **Perceptual Network:** Multimodal sensory network systems (vision, audition, haptics, proprioception, vestibular).
4. **Neural-Cortical Network:** Dynamic synaptic patterns, cortical micro-dendritic structures, predictive-neural processing loops.
5. **Internal, Long-Term Knowledge Graph:** Schemas, procedural knowledge, tacit and subconsciously activated models.

In this conception, Mingers' intransitive domain (Mingers & Willcocks, 2014) can be conceived as relationships found in layers 1, 2, and 3, where the transitive domain is represented in layers 4 and 5 (Mingers, 2001). Informally, on this account it is 'networks all the way down'—from the physical world, to cultures, organizations, and teams, to neural circuits and backup to behaviors—though this phrase is intended as a metaphor rather than a direct quotation, it is fundamental to the conception of enhancing knowledge flow via structural coupling to an XR facilitated VE (Figure 24).

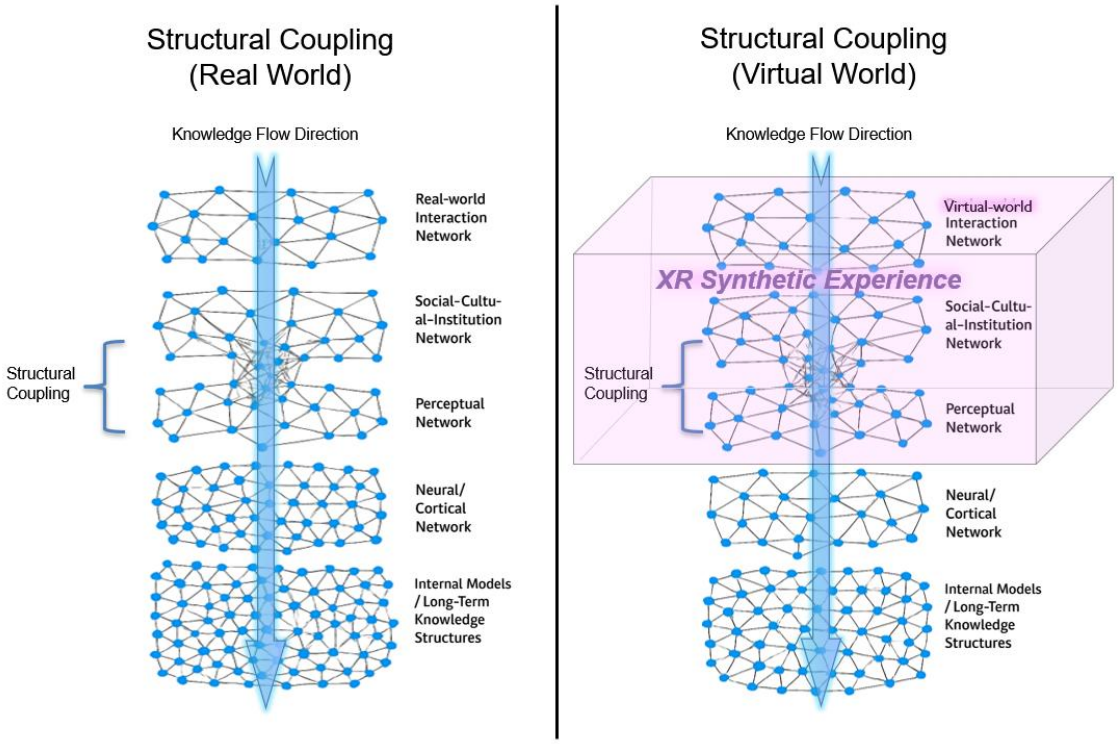


Figure 24. Dissertation theoretic construct informational structural coupling to the real world versus the virtual world.

h. Role of XR: Synthetic Experience Network

XR acts as a programmable and measurable experiential network that replaces or overlays "real world layers 1–3" (Figure 24). By engineering, or encoding, expert situational knowledge into the relational structure of cue richness, feedback coherence, and cognitive integration, XR creates a structured VE (distributed knowledge ecology) that

drives the reorganization of neural and cognitive networks through structural coupling. This conception aligns with recent immersive learning theory and research in embodied cognition (Aguayo et al., 2023; Draschkow et al., 2023).

i. Network Topology Transformation: Before/After

In this theoretical frame, tacit knowledge formation corresponds to internal network topology change. In this conception novice cognitive networks are not as richly connected as those of the expert. After structurally coupling to an experience, networks change to reflect greater density, hub clustering, node connections, and richer overall network integration—reflecting tacit understanding (Zhang et al., 2021). Based on this research comparing a mockup facilitated training environment to an XR environment, the theoretic contribution is reflected in the subjective responses to XR information access evidenced in the comparison to those of the legacy mockup, where in concerns of cognitive load and maritime skills learning are found to be similar, even though the access modalities are completely different

To better illustrate this conception, in the left column (Figure 24), a real world experience can be seen as knowledge flowing to inform networks within the memory structures of a learning individual—these structures, enriched from having been multi-modally experienced, and then better retained as internalized tacit knowledge. The right column depicts an analogous knowledge flow path, where virtual world shaping of the same internal networks, after exposure (structural coupling) to an XR synthetic experience, build similar but different edges, hubs, and cross-links that inform the tacit topology. This topology, not unlike the topology formation after an analogous real world scenario, carries the subtle recognition cues needed to make tacitly informed decisions. This conceptual pair of graphics visualizes how structural coupling with an XR facilitated synthetic experience reorganizes internal networks to support tacit knowledge formation needed for real world operations (Stavros & Ioannis, 2013). This technological capacity of XR-enabled information access provides the potential for wider distribution of tacit knowledge to generate better real world outcomes.

j. Empirical Linkages

Although neural network structures were not directly measured in this research, differences found in NASA TLX, SUS, and BWS self-assessment data support the interpretation that NSST and XR environments produce different relational inputs, resulting in distinct internal organizational pathways. Workload differences, usability perception, and self-rated situational understanding are consistent with altered network trajectories (Aguayo et al., 2023; Zhang et al., 2021), while also demonstrating XR as analogous in its facilitative capacity to that of a physical mockup trainer.

k. Tacit Knowledge Expressed as Network Reorganization via XR

The following provides one possible mathematical structure underlying the emergence of tacit knowledge through changes in network topology of interconnected cognitive networks—and how extended reality (XR) modifies the experiential manifold that drives these internal transformations. At its core, the theory conceptualizes cognition as a trajectory through a high-dimensional vector space embedded within a multilayer network manifold representative of a multilayer cognitive network. Tacit knowledge corresponds to the formation of stable attractor basins in this space. XR functions as an engineered external operator that reshapes the learner's experiential manifold, altering the vector field that governs internal cognitive and neural reconfiguration. This mathematical conception will be executed in follow-on research efforts.

(1) Multilayer Cognitive Networks

The cognitive system can be represented as a five layer graph $\boldsymbol{\mathcal{G}}$:

$$\boldsymbol{\mathcal{G}} = \{G^1, G^2, G^3, G^4, G^5\}.$$

Each layer in "$\boldsymbol{\mathcal{G}}$" corresponds to a distinct but coupled domain:

- G^1: Real-World Interaction Network
- G^2: Social–Cultural–Institutional Network
- G^3: Perceptual Network

- G^4: Neural/Cortical Network
- G^5: Internal Models/Long-Term Knowledge Structures

Mathematically, each layer is represented by an adjacency matrix $A^{(i)}$, forming a tensor-multigraph structure:

$$A = \{A^{(1)}, A^{(2)}, \ldots, A^{(5)}\}.$$

Tacit knowledge is encoded in cross-layer alignment mappings:

$$\Phi : G^1 \times G^2 \times G^3 \rightarrow G^4 \rightarrow G^5.$$

XR modifies the domain and structure of Φ by altering structure in $G_1, G2, G_3$.

(2) Vector Space Embedding

Each network layer can be embedded into a d-dimensional vector space using spectral methods, diffusion maps, or graph embeddings. This theoretic conception of the embedded of d-dimensional yields the following expression:

$$X^{(i)} \in \mathbb{R}^{nxd}.$$

Tacit knowledge corresponds to changes in the geometry of these embeddings. Distance, curvature, and clustering within $X^{(i)}$ reveal the internal restructuring induced by experience in the real world or XR.

(3) Learning as a Trajectory Through Experience Space

Experience generates time-indexed cognitive states represented as vectors:

$$x(t) \in \mathbb{R}^d.$$

Learning can thus be represented as a vector based trajectory:

$$\gamma = \{x(t): t^0 \rightarrow t^1\}.$$

Tacit learning corresponds to movement toward stable regions (attractor basins) within this vector space. Structural coupling to XR modifies the accessible regions and pathways of γ.

(4) Attractor Dynamics and Tacit Knowledge

Tacit knowledge emerges when the learner's internal dynamics settle into stable attractor basins. These settled dynamic or stabilized states are defined formally by:

$$dx/dt = f(x, \theta).$$

with stability defined as occurring when:

$$\nabla f(x) = 0 \; and \; Hessian(f) > 0.$$

XR experiences reshape "f" by synthetically altering sensory, social, and physical contingencies. Thus, XR reshapes the attractor landscape in which tacit understanding emerges.

(5) XR as an Operator on the Experiential Manifold

XR environments can be formalized as a vector field operator:

$$F\,XR : \mathbb{R}^{\mathrm{d}} \rightarrow \mathbb{R}^{\mathrm{d}}.$$

During training, the learner's cognitive dynamics can be represented in:

$$\frac{dx}{dt} = f(x) + F\,XR(x).$$

XR thereby shifts trajectories toward new attractor basins, accelerating tacit knowledge formation. This expression provides a mathematical conception of XR's capacity to support rapid experiential learning even without explicit instruction or performance grading provided via syntactic expression.

(6) Structural Coupling as a Pushforward Map

Structural coupling between the learner and the environment can be expressed using the following differential geometry:

$$\Phi *: T(M\ XR) \rightarrow T(M\ cortex).$$

Here, the pushforward map Φ* projects vectors (changes) in the XR manifold onto the learner's neural manifold. This manifold geometric expression posits how XR synthetic experience reorganizes cortical networks by reshaping the flow of perceptual, social, and physical cues and relations encoded in Layers 1–3, $(G_1, G2, G_3)$.

(7) Observable Network Signatures of Tacit Knowledge

Tacit knowledge formation corresponds to measurable differences in network properties, including the following: increased clustering, emergence of richly connected hubs, shorter characteristic path lengths, manifold compression, and stabilization into attractor basins. These are expressed mathematically as follows.

Increased clustering coefficient expressed in:

$$C(G\ neural\ post) > C(G\ neural\ pre).$$

Emergence of richly connected hubs in:

$$\varphi\ post\ (k) > \varphi\ pre\ (k).$$

Shortening of characteristic path lengths as:

$$L(G) = 1/(n(n-1)) * \Sigma\{i \neq j\}\, d(i,j), where\ L\ post < L\ pre.$$

Compression of the manifold (lower entropy) as expressed in discrete entropy:

$$H = \Sigma i\ pi\ log\ pi.$$

In continuous entropy (differential entropy) as:

$$H = \int p(x) log\ p(x) dx.$$

For pre-XR vs. post-XR manifolds as:

$$H\ pre\ =\ H[p\ pre(x)],$$

$$H\ post\ =\ H[p\ post(x)].$$

Tacit learning corresponds to:

$$H\ post\ <\ H\ pre.$$

Gaussian approximation using covariance Σ:

$$tr(\Sigma\ post) <\ tr(\Sigma\ pre),$$

$$det(\Sigma\ post) <\ det(\Sigma\ pre).$$

Stabilization into attractor basins can be expressed as tacit convergence (post-XR) appears as:

$$x(0) \in\ B(A\ tacit) \Rightarrow\ lim\{t \to \infty\}x(t) =\ x* \in\ A\ tacit.$$

where the convergence rate increases in:

$$||x(t) - x*||\ \leq\ C\ e\text{^}(-\lambda\ t), with\ \lambda\ post\ >\ \lambda\ pre.$$

These expressions provide for the development of mathematically interpretable indicators of cognitive reorganization, even when not directly measured empirically, e.g., via HR fMRI.

(8) Implications for XR-Based Training

The mathematical framework reveals why XR is uniquely powerful for tacit learning. XR transforms the experiential manifold, alters the vector field of relational dynamics, and reshapes the internal cognitive manifold. In effect:

1. Tacit knowledge formation can be mathematically expressed as corresponding to changes in the topology of the cues provided across a multilayer, structurally coupled manifold.

2. Tacit knowledge becomes expressible through attractor dynamics, vector trajectories, and manifold reorganization.

3. In this conception, XR synthetic experience alters the participant's experiential manifold, which in turn reshapes the vector field that drives internal network reconfiguration needed to form tacit understanding.

This theoretical frame of sequenced informational coupling via the novel information access of XR unifies elements of XR design theory, distributed cognition, network neuroscience, and tacit knowledge theory within a coherent mathematical structure. This is applicable not just to individuals, but also to teams of individuals who develop team constructs of tacit knowledge and awareness in larger knowledge ecosystems such as a ships pilot house.

l. Theoretical Implications

This multilayer network theoretical conception synthesizes embodied cognition, distributed cognition, network neuroscience, tacit knowledge theory, knowledge flow and KM frameworks, and XR design theory. In this, XR is positioned not merely as a training device, but posits it as a programmable relational environment that systematically shapes the conditions under which tacit learning develops. This provides an additional cultural path for sharing tacit knowledge, e.g., beyond a constraints based model (Miton & DeDeo, 2022).

m. Cognition in the Virtual

The theoretical extension of Nissen's model (Figure 23) grounds a broader concept of *Cognition in the Virtual*, which conjectures that extended reality (XR) environments can serve as autopoietic spaces to enable knowledge flow in team and ecological constructs of distributed cognition. In such environments, knowledge is neither fully tacit nor fully explicit but exists as an enacted potential—a field of shared perceptual affordance through which learning and expertise emerge collectively. By making a scenario explicit experience, rather than using the language of instruction, synthetic environments resolve the Polanyi paradox and operationalize tacit knowledge transfer at scale and in situations

where cognition is distributed beyond the minds of single individuals. The resulting model unites Nissen's systemic view of knowledge flow with embodied cognition theory, demonstrating how XR can transform KM from a process of linguistic extraction, storage, mapping, and exploitation to one of experience design. The synthetic experience thus becomes both the medium and the message of expertise—a dynamic knowledge flow structure in which cognition itself becomes distributed, enacted, and perpetually renewed.

n. Knowledge Integration with Cognition in the Virtual

The posited concept of *Cognition in the Virtual* is inspired by Hutchins' 1995 foundational book *Cognition in the Wild.* In his book, Hutchins argued that cognitive processes are not confined to the individual mind but distributed across people, artifacts, and environments. In the context of extended reality (XR), cognition becomes situated within synthetic yet ecologically valid environments—what may be described as synthetic ecologies of knowing. Here, perceptual coupling, interaction, and shared attention occurs not through physical co-location but through co-presence in an immersive, multimodal representational space. The extended KFT accounting for synthetic experience provides the underlying mechanism for how distributed cognition may unfold in virtual contexts. By framing XR as a synthetic field that supports *Experiential Externalization, Structural Coupling,* and *Embodied Internalization*, it becomes possible to conceptualize knowledge flow as the circulation of experience itself, where cognitive alignment among participants arises through the synchronization of perception and action within the shared VE. This process creates an emergent "team cognition" that does not depend on prior articulation but instead on joint engagement with the same dynamic perceptual field of the synthetic experience.

Empirically, this theoretical integration finds grounding in the comparative study between the ONR Tech Solutions VIBRaNT XR system and the Navy's NSST. Within this experimental framework, bridge watch teams enact complex operations under both physical and synthetic conditions. Results from workload (NASA TLX), usability (SUS), and self-assessment data collectively suggest that XR-based synthetic experiences can reproduce—and in some cases enhance—the cognitive and social dynamics of real world

environments. These findings support the claim that cognition in XR is not merely simulated but *enacted*: the VE becomes a legitimate cognitive ecology through which tacit knowledge flows.

Thus, *Cognition in the Virtual* reframes XR environments not as substitutes for physical training but as epistemic systems—spaces where experience itself is rendered explicit and shareable, where knowledge is embodied, enacted, and perpetually renewed through interaction. When coupled with the synthetic experience knowledge flow framework, this perspective offers a unified theoretical foundation linking distributed cognition, KM, and extended reality.

2. Method

This work has designed and demonstrated a quantitative and qualitative experimental methodology for assessing knowledge flow, using proven team training techniques, by comparing virtual XR effectiveness to physical simulator experience. Self-assessment by qualified practitioners across multiple levels of expertise and experience in corresponding real world scenarios confirmed this effectiveness. This is a novel contribution to method for measuring knowledge flow facilitation. (Deck officer self-assessed as helpful in training)

Although none of the individual survey instruments (NASA TLX) or statistical techniques (e.g., Wilcoxon signed-rank tests with Bonferroni correction) are novel per se, the configuration and execution of these methods in a USN XR bridge team-training domain offers a reusable evaluation template for maritime and other safety-critical domains where mockup simulation is used as a preferred training technique. The approach helps fill a methodological gap in XR training literature—where XR system design heterogeneity complicates evaluation of experimental methods employed (Chandramouli et al., 2025), weak controls, and under-addressed order and sequence effects remain significant threats to validity (Langendorf & Khalid, 2025; Salinas et al., 2022).

This study advances methodological practice in XR team-training evaluation by means of three interrelated design and measurement contributions. First, adoption of a within-subjects, counter-balanced quasi-experimental comparison between the

U.S.-modified XR prototype system (VIBRaNT, originally developed by Kilo-Solutions UK as "VASCO") and a high-fidelity physical mock-up (NSST). This design controlled for individual differences and explicitly modeling sequence and novelty (order) effects. Second, the implementation of a multi-instrument measurement battery—including the NASA-TLX for workload, the SUS for usability, self-assessment surveys, instructor performance assessment surveys, and exit interview text analysis metrics—thus moving beyond the single-metric assessments common in many XR training studies to provide a more comprehensive characterization of XR affordances. Third, by situating the evaluation in a team-based, tacit-knowledge transfer context (anchored in a distributed cognition framework), the study extends methodological scope from individual learning into the embodied, social domain of expert-novice coupling in an ecologically relevant environment of USN maritime operations.

Finally, this research documents practical implementation considerations (e.g., system set-up, scenario parity, instructor calibration) that further contribute to method translation for future XR team-training work. Limitations (single-site, sample size) moderate the claims, but do not detract from the methodological value of the design.

In summary, this research provides a methodological contribution by 1) applying a rigorous comparative XR vs. physical mock-up within-subjects design with order-effect controls; 2) integrating a comprehensive measurement battery across usability, workload, self-assessment, performance, and text analysis; and 3) deploying the evaluation in an ecologically valid, team-based tacit knowledge flow context—thereby offering a blueprint for future XR evaluation in team-training settings. Table 74 summarizes the contributions to method.

Table 74. Summary of contributions to method.

Category	Description	Contribution Type	Example/Application
Design	Within subjects, counter-balanced quasi-experimental design comparing XR prototype system and the NSST physical mockup under matched maritime scenarios.	Experimental Rigor and Control of Individual Differences	Controls for participant variability and explicitly examined order and sequence effects across XR and physical training conditions.
Measurement Framework	Integrated battery of NASA-TLX (workload), SUS, self-assessment surveys, instructor assessments and text analysis metrics.	Multi-Instrument Evaluation and Triangulation	Provides multidimensional view of user experience, workload, and perceived team performance—rarely combined in prior XR training studies.
Analytic Approach	Applied multiple non-parametric tests (e.g., Wilcoxon signed-rank with Bonferroni correction) to detect condition differences and manage multiple comparisons.	Statistical Rigor and Transparency	Ensures robustness of inferential results while controlling for family-wise error rate.
Team Context	Evaluation embedded in distributed cognition and tacit knowledge transfer framework, with expert and novice roles replicated from actual bridge watch teams.	Ecological Validity and Contextual Realism	Extends XR evaluation from individual tasks to team-based cognitive processes, increasing ecological validity and generalizability to other needed spatial-temporal skill sets.
Implementation Practice	Practical alignment of XR and NSST scenarios (e.g., UNREP, channel transit) with instructor calibration and scenario realism.	Method Transferability	Offers reproducible configuration and procedural guidance for future XR team-training evaluations in safety-critical environments.

3. Practice

This section articulates the applied contributions emerging from a within subjects comparison of the prototype XR system (VIBRaNT) and the USN legacy physical mockup (NSST). While the perspective is decidedly nautical and USN-centered, it is intended to

also inform general aspects of knowledge flow for the purpose of transferring spatial-temporal skills. In this way it emphasizes system (Human and Machine) readiness, training throughput, team coordination and collaboration, and the transfer of tacit expertise for real world operations. Here the contributions to practice most directly apply to USN fleet instructors, program managers, and senior decision authorities, but depending on organizational need, may also extend beyond USN and domains maritime knowledge. These practical contributions may also foreshadow, where appropriate, the broader theoretical construct of *Cognition in the Virtual*.

a. Tacit Knowledge Flow and Embodied Learning

XR's immersive spatial context amplifies tacit knowledge transmission. In a shared virtual bridge, juniors observe the micro-behaviors of experts—the pacing of verbal reports, the timing and content of helm and lee-helm commands, the scan pattern across windows and displays, the tempo of cross-checks—without the friction and latency of text-only or classroom explanation. Because the team is co-present, cues that are difficult to verbalize become visible and repeatable. Experts can step in from anywhere, annotate the situation in situ, and scaffold novice performance during the evolution. Over time, these embodied patterns become part of the team's distributed memory, accessible on demand. The net effect is to extend tacit knowledge flow spatially (across locations) and temporally (across shorter, more frequent sessions), while preserving the nuances of bridge practice.

b. Configurability and Contextual Adaptation

Because the XR system is software-defined, it is rapidly configurable for ship- and environment-specific fidelity. VEs can be tailored to match a particular hull form, sensor/ bridge layout, or local maritime geography (e.g., the Malacca Strait) with comparatively low turnaround time and cost. Once created, the same VE can be used for initial qualification, periodic refresher training, and mission rehearsal—preserving consistency across the learning pipeline. Unlike physical mock-ups that require laborious reconfiguration or bespoke hardware, software-native XR allows scenario updates and fixes to be pushed globally, ensuring that distributed crews train against the same, current conditions.

c. Training Practice and Team Coordination

The XR environment improves instructor agility and team learning continuity. Instructors can monitor sessions locally or remotely, pause and replay decisive moments, and inject alternative branches to probe decision boundaries. Scenarios can be reset instantly to permit deliberate practice—multiple iterations of a turn-in-channel or an approach-to-an-anchorage without the time penalties associated with re-setting a physical space. The same distributed watch team can rehearse across several short sessions in a week, rather than waiting for a single long block of fixed simulator time. This cadence supports spaced repetition, incremental complexity, and progressive autonomy—pedagogical patterns known to improve retention and decision performance.

d. System Mobility and Accessibility

The XR system fundamentally alters the logistics of maritime team training. Packaged as a fully self-contained "bridge in a box," a single Pelican Case carries approximately 12 HMDs, a compact VR server (roughly the size of a small textbook), a Wi-Fi router, and a scenario control/creation PC laptop. Dedicated simulator rooms are not required; setup involves establishing power and space to operate wireless LAN. A notional DoD Architectural Framework 2.0 Operational View 1 (DODAF 2.0 OV-1) is presented as a possible implementation of XR bridge classroom training (Figure 25).

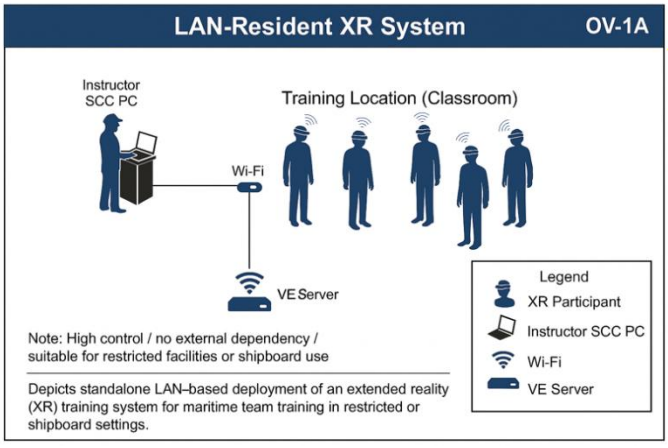


Figure 25. Dissertation notional operational view of extended reality bridge training in a classroom or other locally connected space.

If practicing bridge teams have access to WiFi, then all that is needed is the Scenario Creation and Control Personal Computer (SCC PC) and HMDs for those entering the VE for practice. This ability enables individuals, or complete training teams, to assemble ad hoc in classrooms, wardrooms, reserve centers, or temporary facilities during detachments or other times when attending the NSST mockup is not possible. When WiFi Internet connectivity is available, watchstanders dispersed across multiple geographic locations can enter the same shared virtual bridge, achieving true co-presence.

In addition, if connected via WiFi, instructors based in remote locations, e.g., at MSTCPAC, are available, they can join the XR training session from their desks by simply donning an HMD and coaching the team during a scenario set off the shores of any coast in the world. In addition, multiple ships' crews can join their respective virtual bridges and interact with each other in the same VE from the vantage point of their ships, e.g., one ship leading a formation, the other following in trail as happens during an UNREP operation, division tactics operation, or any number of multi-ship operations. A notional DODAF 2.0 OV-1 is provided (Figure 26) to show one possible implementation of a cloud-based XR bridge training system.

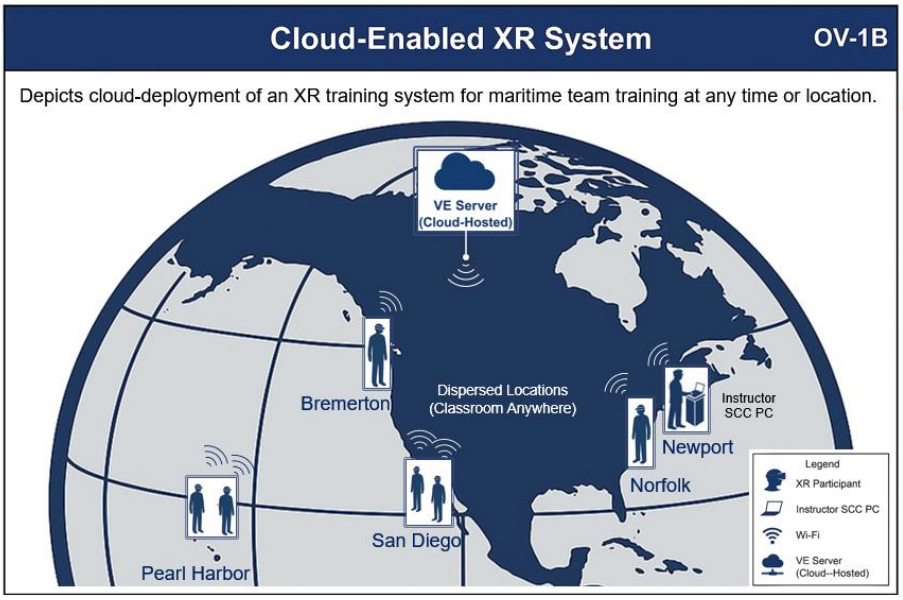


Figure 26. Dissertation notional operational view of extended reality bridge training dispersed to any location with a WiFi access point.

The clear operational payoff is high availability. XR removes some of the scheduling bottlenecks inherent to fixed, room-bound mock-ups. Because the system is mobile and lightweight, the cycle time between exercises decreases, repetition increases, and practice can be completely opportunistic: teams can drill at any time between other required evolutions, rehearse an UNREP during a lull, or run rapid decision games in response to emerging fleet lessons learned. In short, accessibility and throughput increase without a proportional increase in fixed infrastructure.

e. Cost Efficiency and Sustainability

Acquisition and life cycle costs favor XR. As a reference point, a single NSST projector costs approximately $30,000, and an NSST bridge typically employs roughly 20 projectors, alongside specialized computer servers (up to 72 computers for the more advanced NSSTs), instructor consoles, audio and lighting systems, dedicated heating ventilation, and air conditioning (HVAC) system, and more. The facility cost for a single NSST room is approximately $1.77 million, excluding maintenance, calibration, and periodic upgrades. By contrast, an XR PUK capable of supporting a full 12-person watch team can be fielded for approximately $20,000, inclusive of HMDs, server, router, and Scenario Creation and Control PC laptop. Because the XR solution is software-defined, capability increments arrive via over-the-air updates rather than facility retrofits. The result is a high-readiness, low-overhead pathway to expand training capacity without large capital expansion—aligned with fiscal stewardship and energy-efficient operations. In addition, commercial XR equipment continues to decline in cost while advancing in capability and shows no sign of slowing, offering the USN and other large organizations unprecedented opportunity to expand spatial-temporal and procedural learning.

f. Operational Integration and Future At-Sea Utility

The small footprint and wireless design further enable shipboard integration, pending appropriate safety and human-systems validation. Because the kit is self-contained and requires only power and local networking, crews can train pier-side or underway. This flexibility opens a set of potential at-sea use cases with immediate practice value: pre-port entry briefs and team walk-throughs using a faithful virtual replica of the expected

approach; in-port and underway continuation training that reinforces bridge resource management; rapid qualification boards for OOD, JOOD, and CONN using standardized, repeatable scenarios; and targeted remediation following close calls or navigation discrepancies. The cumulative effect is to collapse the wall between the schoolhouse and the ship, making training a living part of operations rather than an event separated from them.

g. LVC Integration

XR's native networking makes it a natural node in fleet synthetic ecosystems. When connected to LVC networks used in Fleet Synthetic Training, the XR bridge VE becomes a maneuvering and decision-support platform that can interact with live units, virtual simulators, and constructive (computer-generated) forces. Fleet Synthetic Training (FST) does not currently integrate bridge teams as it has no means of replicating the environment outside the bridge windows. What is seen is the view of the port they are tied up to, not a distant ocean environment in harm's way. By providing an XR environment connected to an FST event, the bridge team can join in with the rest of the crew, typically training in their ship's CIC, in order to clear batteries, perform tactical maneuvers, or otherwise integrate with the ship's tactical actions officer in the practice of combat operations. In this configuration, bridge teams participate as tactical actors, practicing shiphandling, contact management, and watch team coordination under realistic multi-unit combat operational conditions.

The training value is twofold: (a) XR extends participation to units and schools that otherwise lack access to large, fixed simulators; and (b) it allows the same team to train locally yet fight synthetically at the fleet scale. A notional DODAF 2.0 OV-1 is provided as one possible implementation of XR bridge training integrated with LVC FST events (Figure 27). Of note is that all information streams in such LVC environments can be compositions of prerecorded replays from prior scenarios, real-time renderings of current operations, or simulations of alternative future tactics. Use of XR presentation enables embodied understanding and visualization. These are significant potential capabilities for the USN.

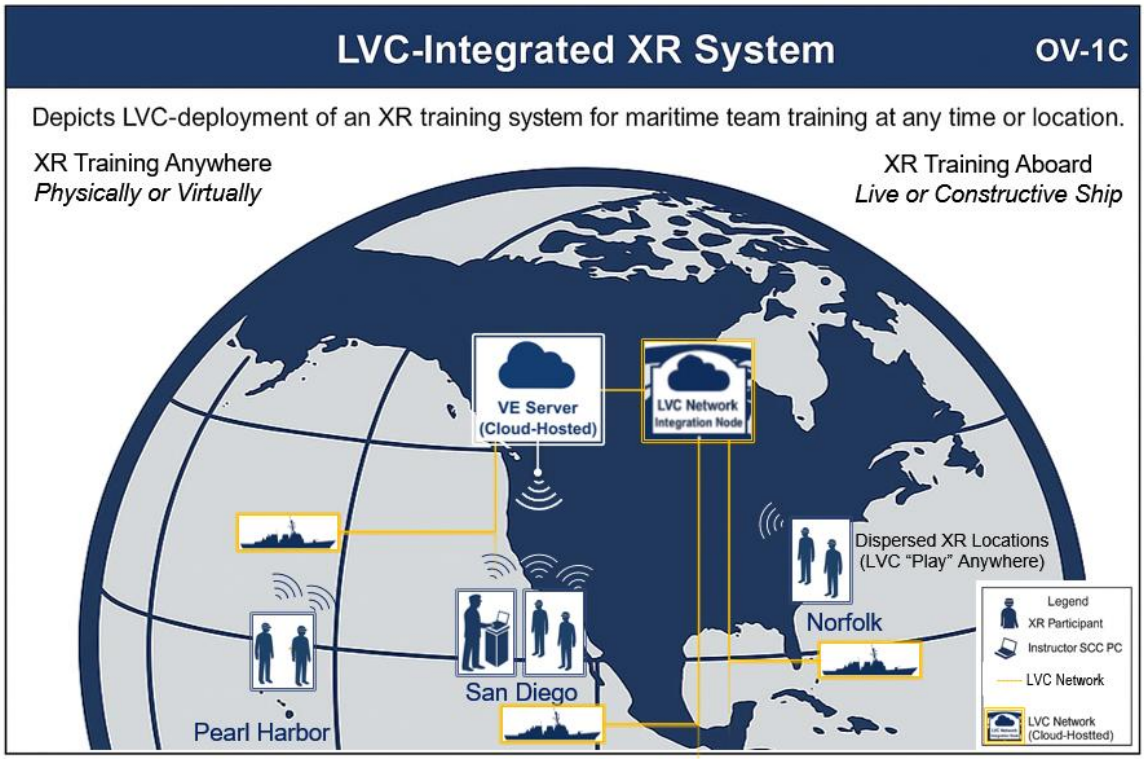


Figure 27. Dissertation notional operational view of extended reality bridge training environments integrated with live, virtual, constructive networks.

h. Future Distributed C2 Practice

XR training provides a near-term pathway to the Navy's overarching strategic goal of migrating to unmanned and distributed systems. The HMD interface habituates operators to immersive data visualization, sensor fusion, and remote coordination—competencies directly transferrable to advising or controlling unmanned vessels and other autonomous assets. In XR, multiple experts, from any location around the world, can occupy a shared virtual space to bring specialized situational expertise to bear on a problem aboard a small, unmanned craft—co-present virtually even where physical co-presence would be impossible or unsafe. This benefit is not merely convenience; it is doctrinal rehearsal for distributed command and control in which human expertise is dynamically composed across time and space. In this sense, XR bridge training incubates the cognitive and procedural readiness required for distributed human–machine teaming, distributed awareness, and future interaction with artificially intelligent agents.

i. Transitional Note: Toward Cognition in the Virtual

Taken together, these potential contributions to practice highlight that what matters in maritime training, and knowledge flow in general, moving away from considerations limited to fidelity of a "room-bound" mockup system, but to new fidelities of a living, distributed cognitive system: people, tools, procedures, and context coupled across locations and time. XR's mobility, configurability, and network integration expand the reach of embodied expertise, making tacit knowledge observable, transmissible, and reusable at fleet scale. This "applied reality" points forward to the theoretical construct of *Cognition in the Virtual*: the proposition that cognitive work is increasingly performed by socio-technical systems that are both spatially distributed and experientially coherent. The practical gains documented here—higher availability, lower cost, deeper tacit knowledge transfer, and operational integration—are not only improvements in training practice, but empirical footholds for a broader account of how naval teams think and act together through virtual means.

C. RECOMMENDATIONS FOR FUTURE RESEARCH

Future research should advance naval operational training, distributed team cognition, and the emerging theory of Cognition in the Virtual. Practical Navy-focused recommendations are presented first, followed by broader theoretical and cross-domain directions.

1. Navy Research and Extended Reality

Research opportunities for Navy exploitation and use of XR are virtually limitless, here are but a few examples.

a. Shipboard XR Testing, Pier-Side and Underway

Future studies should assess XR performance aboard operational vessels to evaluate motion effects, sea state, ambient vibration, and HMD user comfort. Testing aboard DDGs, LCS, CVNs, and amphibious platforms would strengthen external validity and readiness implications.

b. Integration of XR With Navy C2

Research should evaluate XR-based C2 decision-making, cross-reality coordination between watch teams, and enhanced SA through 3D spatial-temporal overlays. XR-enabled C2 rehearsal for mission planning and real-time decision support should also be explored.

c. LVC-XR Integration for Distributed Maritime Training

Expanding LVC integration is essential. Future work should test multi-ship scenarios, multi-platform integration, and global distributed training events linking XR, constructive forces, desktop simulations, and physical platforms.

d. XR for Navy Qualification, Certification, and Rehearsal

Research should explore XR-based OOD and engineering officer of the watch (EOOW) qualification boards, virtual rehearsals for daily ship evolutions, and simulated playback of incidents or mishaps for learning and instructor calibration. Spatial-temporal cueing may improve performance evaluation and standardization.

e. Perspective-Shifting and Enhanced Spatial Cognition in XR

Future studies should test XR's unique capability to shift perspectives—from first-person to third-person, overhead god's-eye view, and role-switching. Research should also evaluate real-time perspective sharing across watchstanders, 360° bow/stern views, and instructor free-roam for analysis.

f. XR-Enabled Playback, Review, and After-Action Insight

Research should assess how replaying training or real world incidents in XR influences tacit knowledge acquisition, error detection, and instructional quality. Automated tagging of performance gaps could support future Navy training doctrine.

2. Integrated Extended Reality–Artificial Intelligence–Navy Live Virtual Constructive Research Directions

The possibilities emerging when combining XR with AI in any set of eventualities is a paradigm shift not only for military operations, but for how humanity interacts with worlds both real and imaginary in the future. Here are but a few areas in need of research.

a. AI Avatar Agents for Missing Crew Roles

Future work should develop AI-driven virtual watchstanders capable of performing realistic roles when full bridge teams cannot assemble. These agents would preserve communication flow and cognitive scaffolding during training.

b. Adaptive AI-Driven Scenario Adjustment

XR platforms should be studied as dynamic tutors, adjusting scenario difficulty based on detected cognitive gaps. AI-driven pattern recognition could enhance tacit transfer by meeting teams "where they are."

c. Structural Coupling of AI to XR Environments

Research should explore architecture allowing AI systems to perceive and act upon synthetic environments in human-like spatial-temporal frames. This may improve human–AI teaming in both training and C2 contexts.

d. Defining Return on Investment Regions for XR–AI Integration

Future work should identify where XR–AI coupling yields the greatest operational value, prioritizing cases with measurable performance outcomes and cost effectiveness.

3. Empirical Tacit Knowledge and Cognitive Science Research

Here research into XR application extends far beyond the military to encompass larger aspects of human cognition and human computer interaction.

a. Quantifying Tacit Knowledge "Tipping Points"

The vector-math model developed in this study suggests measurable inflection points in tacit learning. Future research should further validate these tip-over dynamics across different roles and tasks.

b. Longitudinal Studies of Tacit Transfer

Future research should track trainees across months or deployments to measure stability, decay, generalization, and operational impact of tacit knowledge acquired through XR.

c. XR Eye Tracking for Cognitive Process Mapping

Eye tracking within XR should be used to analyze expert–novice scanning patterns, early warning indicators of cognitive overload, SA profiles, and Navy C2 cognitive processes.

d. Determining When XR Is Warranted

Future work should build a task–modality fit model that defines when XR, NSST mockups, tabletop simulations, or live shipboard training are most appropriate. This model should incorporate tacit knowledge flow, workload, team cognition, and operational risk.

4. Advanced Theoretical and Cross-Domain Research

In the conduct of this research, it was found in a series of multiple conversations and insights, that the potential for wider exploration of XR is warranted. Below are a few conjectures for potentially enhancing future understanding of synthetic experience.

a. Refinement of the Cognition in the Virtual Framework

Future work should formalize propositions about multi-reality cognition, distributed tacit knowing, structural coupling, embodied cues, and perceptual scaffolding unique to XR.

b. Multi-Reality Team Cognition Research

Research should examine how teams coordinate across physical, virtual, and mixed realities; how mental models align; and how shared perceptual fields influence coordination and error recovery.

c. XR as a Precursor Environment for Brain Computer Interface Research

XR may serve as a transitional architecture for future brain–computer interface research. Studies should explore XR behavioral signatures predictive of brain computer interface (BCI) adaptability and cognitive integration.

d. Generalization to High-Risk and High-Cognition Domains

Future work should test the Cognition in the Virtual framework in aviation, medicine, industrial control, emergency response, and space operations, where tacit knowledge and distributed cognition are mission critical.

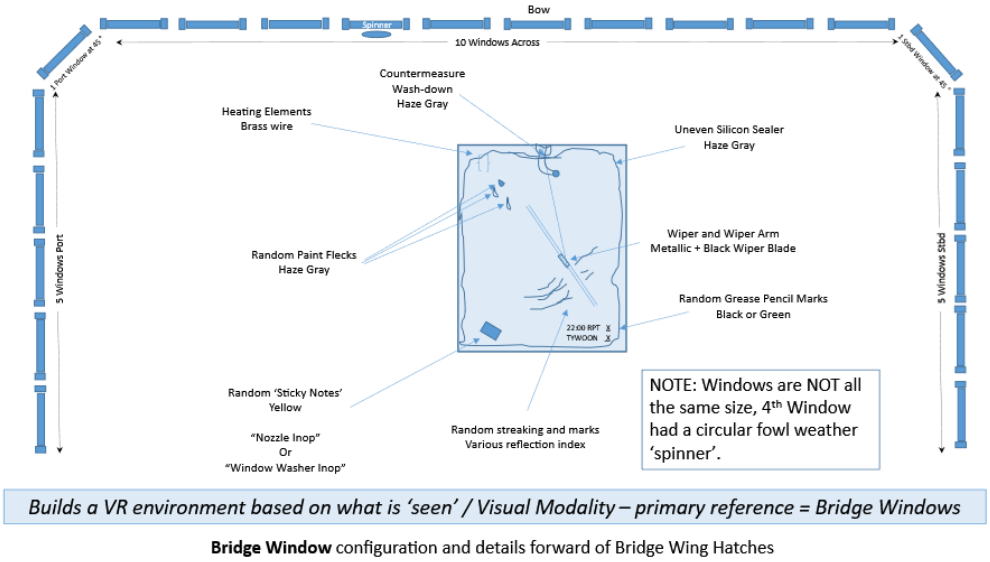
5. Summary of Future Research Recommendations

These recommendations outline a comprehensive agenda spanning fleet operations, Navy C2, XR–AI integration, tacit knowledge quantification, longitudinal learning, and theoretical advancements. Collectively, this work would deepen scientific understanding of multi-reality cognition while supporting Navy modernization, distributed training effectiveness, and wider non-military applications into the future.

THIS PAGE INTENTIONALLY LEFT BLANK

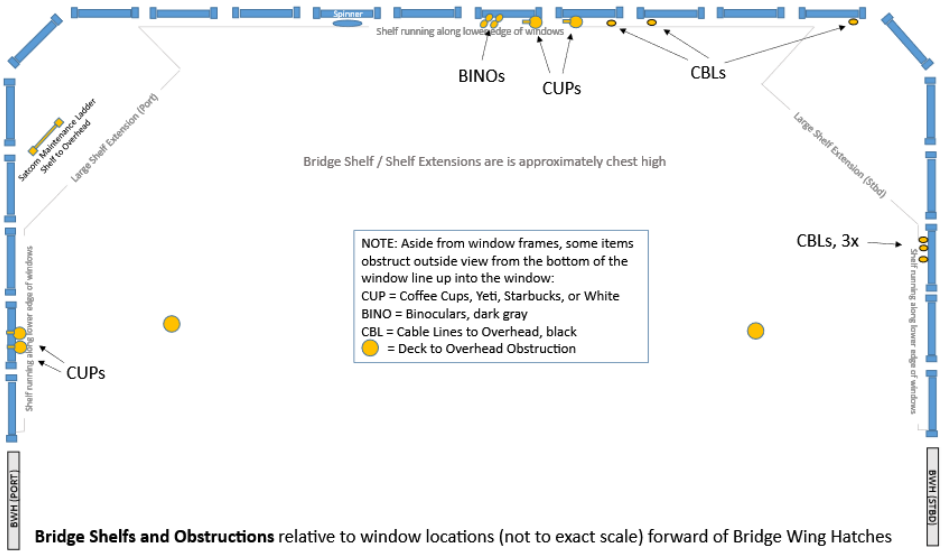
APPENDIX A. EXTENDED REALITY BRIDGE VIRTUAL ENVIRONMENT CONFIGURATION DESIGN

Appendix A includes the bridge floor plan and sound scape VE requirements. These plans were based on "at sea" observations aboard USS Paul Hamilton DDG-60.



Bridge Window configuration and details forward of Bridge Wing Hatches

Figure A.1. DDG bridge window details and layout.



Bridge Shelfs and Obstructions relative to window locations (not to exact scale) forward of Bridge Wing Hatches

Figure A.2. Bridge window lower obstructions.

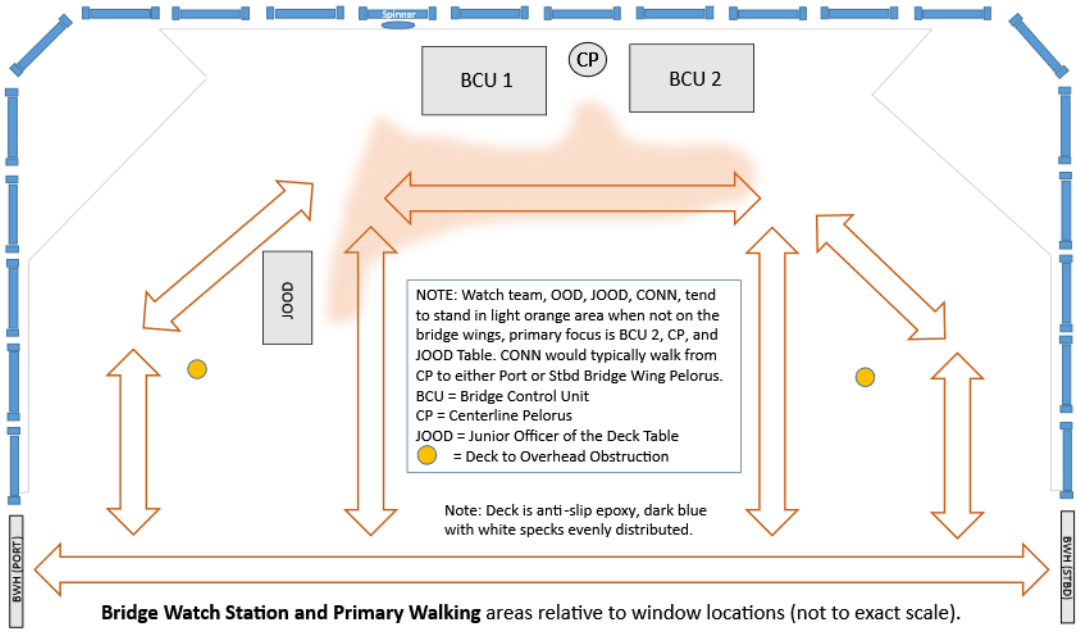


Figure A.3. Primary activity areas in the bridge.

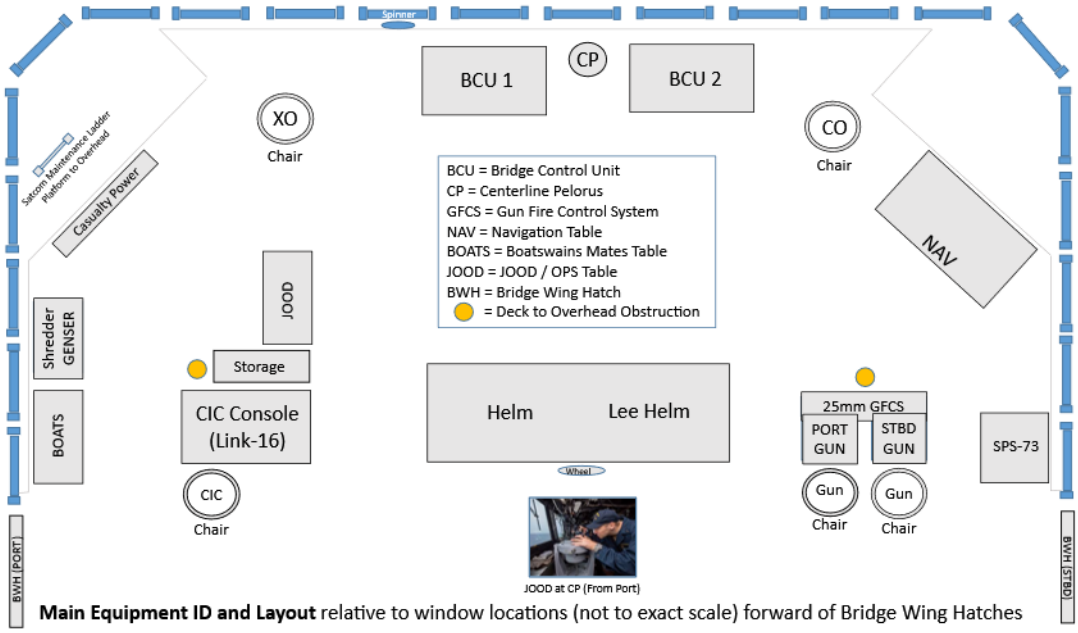
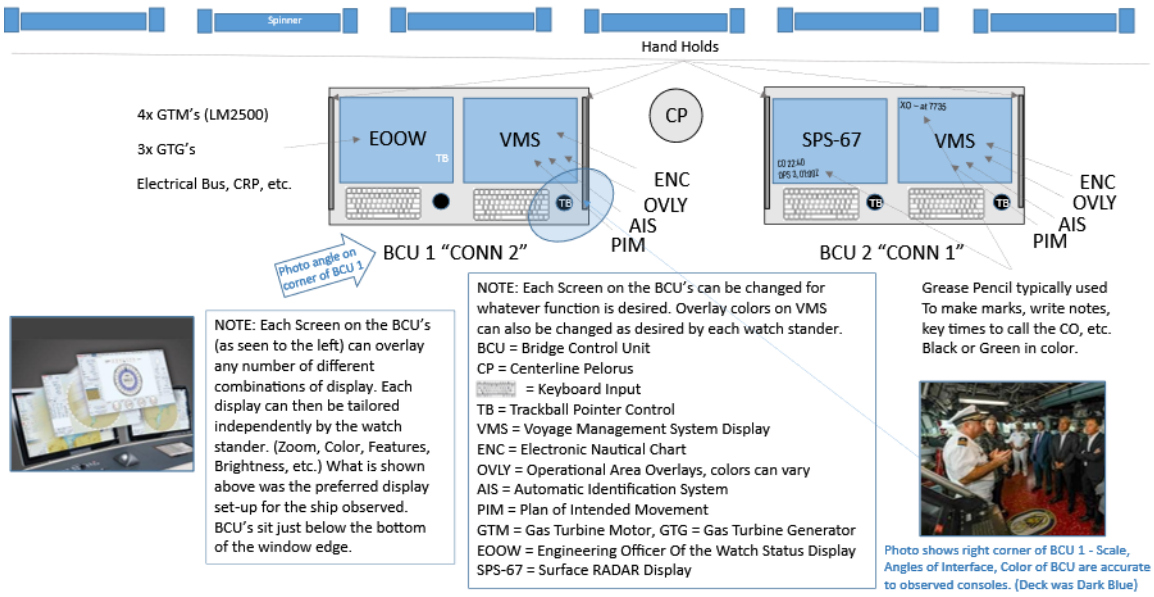
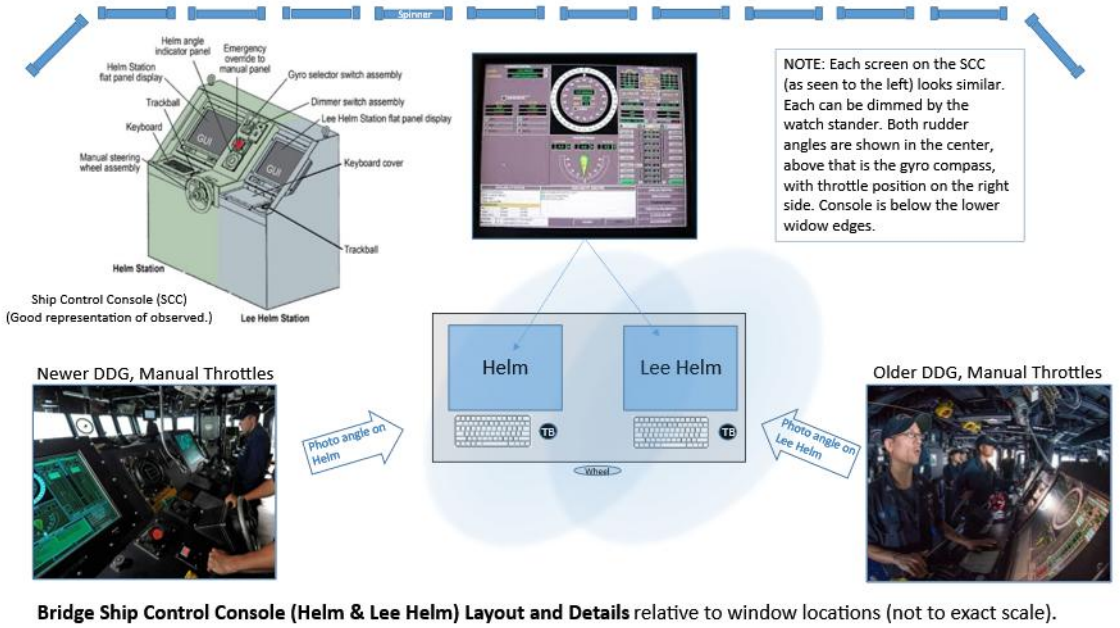


Figure A.4. Main computer consoles and furnishings on the bridge.



Bridge Control Unit Example Layout and Details relative to window locations (not to exact scale).

Figure A.5. Bridge control unit console details.

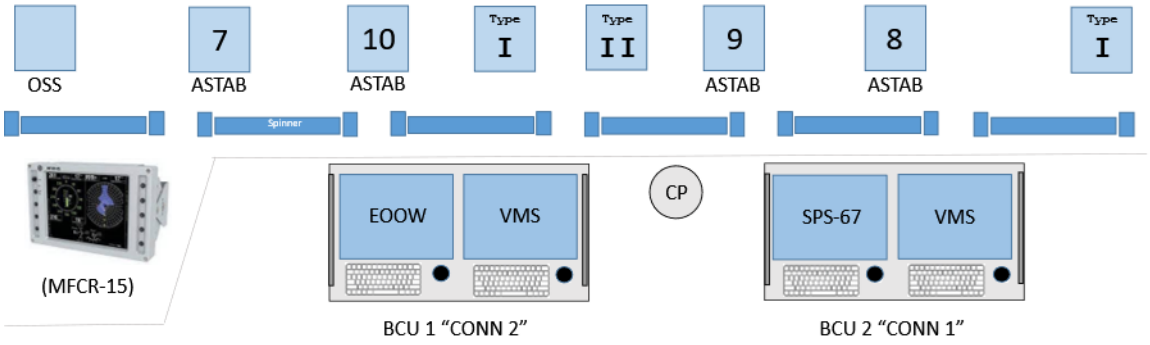


Bridge Ship Control Console (Helm & Lee Helm) Layout and Details relative to window locations (not to exact scale).

Figure A.6. Ship control console details.

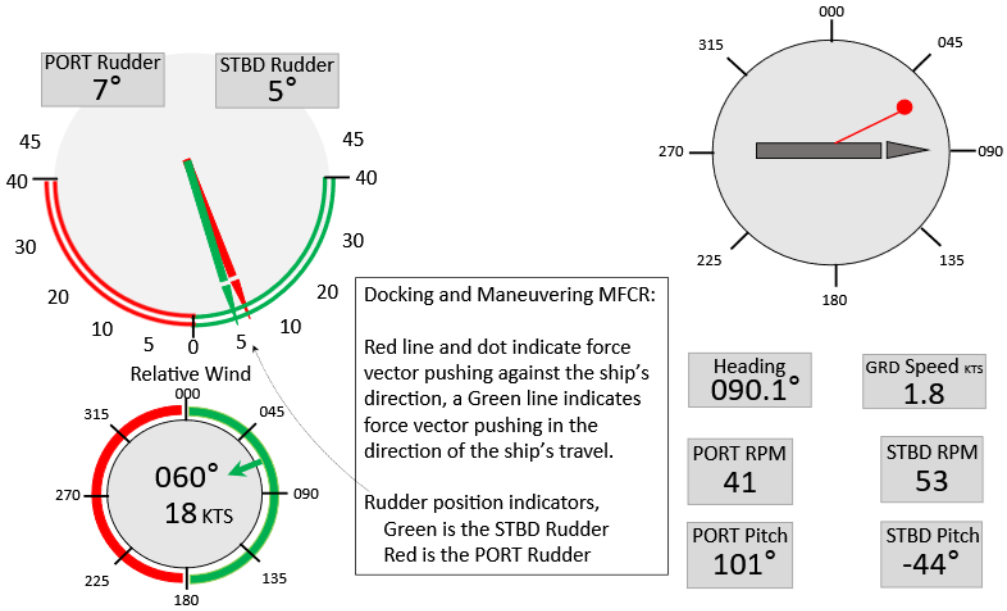
MFCR-15's are mounted above the bridge windows. Each display can be independently configured to provide information as desired. These can be reconfigured by the bridge watch team as needed depending on the operational needs.
ASTAB = Automated Status Board: Information from Combat Information Center, Call Sign's, Track Numbers, etc.
OSS = Operational Surveillance System: High Fidelity Electro-Optical Camera Display with Night Observation Capability.
'Type I' = Docking and Maneuvering Display Repeater.
'Type II' = Wind, Roll, and Pitch Repeater. Other 'types' of displays are possible.

[NOTE: 'Type I' display is used by Helm, Lee Helm, CONN, OOD, and JOOD for Ship-Handling Situational Awareness.]



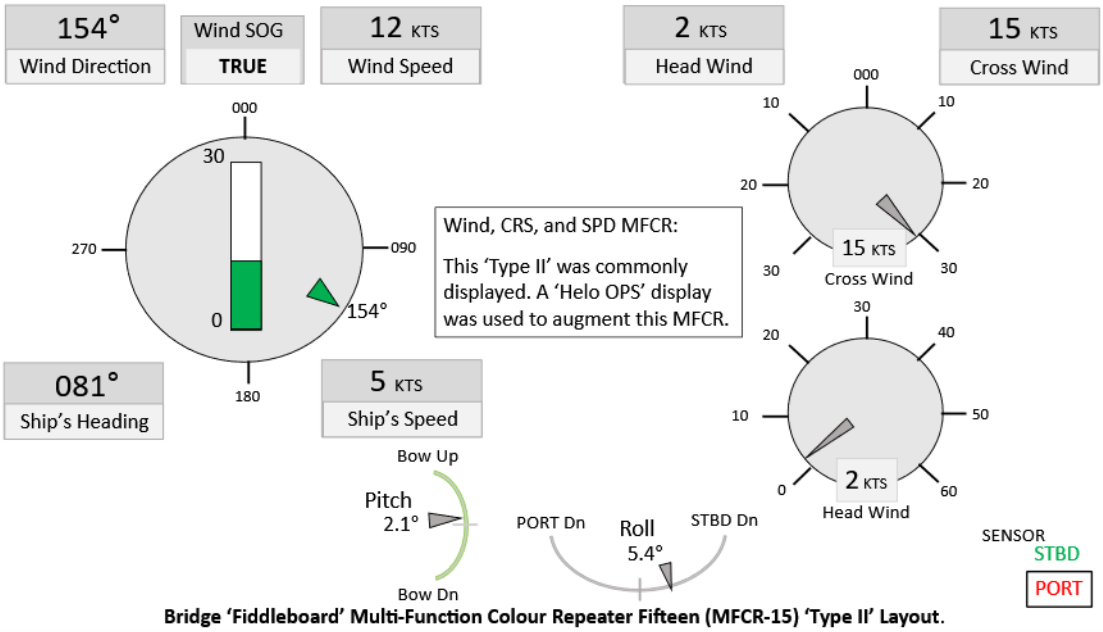
Bridge Multi-Function Colour Repeater Fifteen (MFCR-15) Layout above window locations (not to exact scale).

Figure A.7. Bridge multi-function control repeater layout.



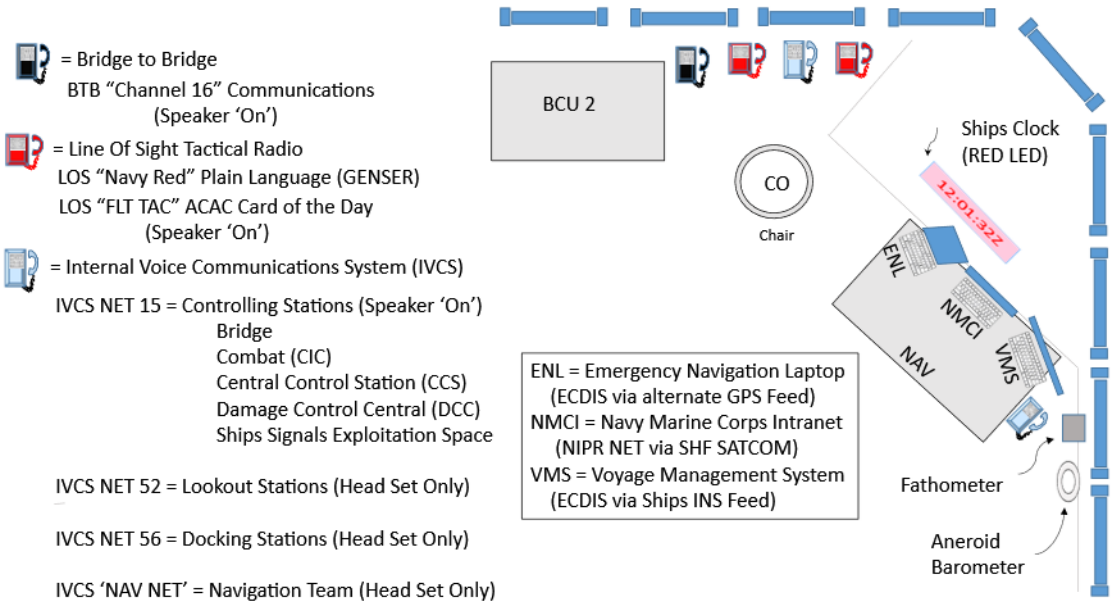
Bridge 'Fiddleboard' Multi-Function Colour Repeater Fifteen (MFCR-15) 'Type I' Layout.

Figure A.8. Bridge multi-function control repeater details.



Bridge 'Fiddleboard' Multi-Function Colour Repeater Fifteen (MFCR-15) 'Type II' Layout.

Figure A.9. Bridge 2 multi-function control repeater details.



Navigation Table Layout + nearby communication handsets relative to window locations (not to exact scale).

Figure A.10. Bridge navigation and commanding officer layout details.

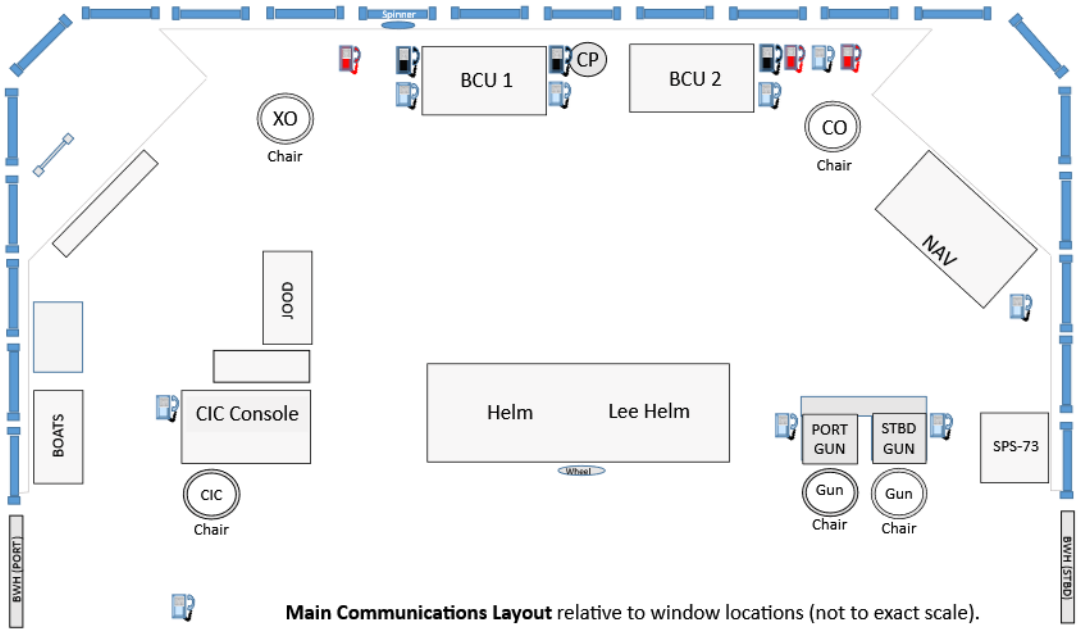


Figure A.11. Bridge radio and internal communications layout details.

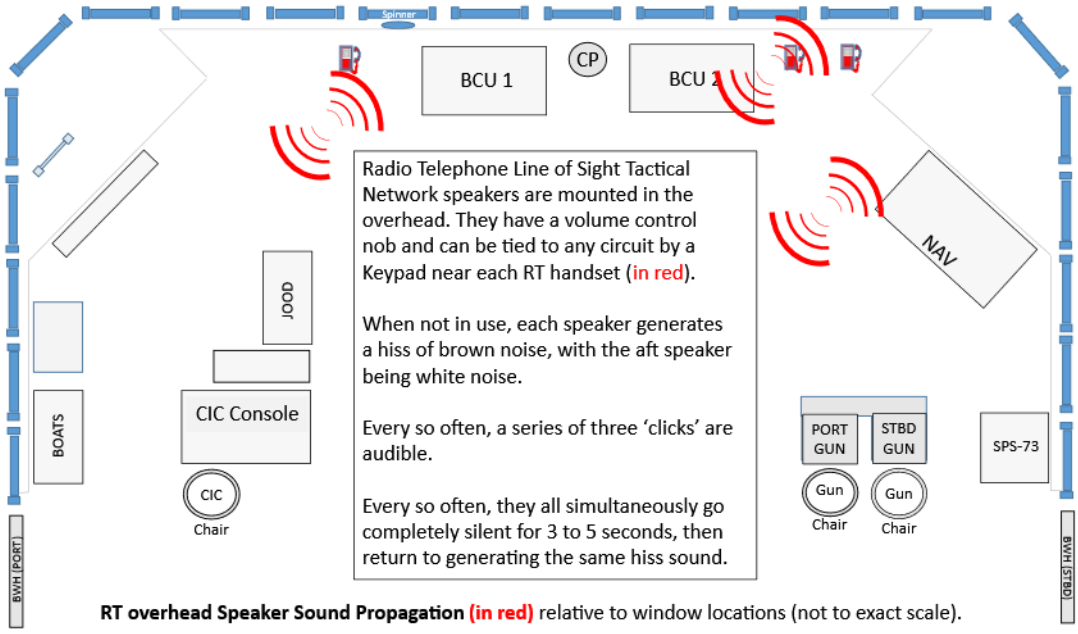
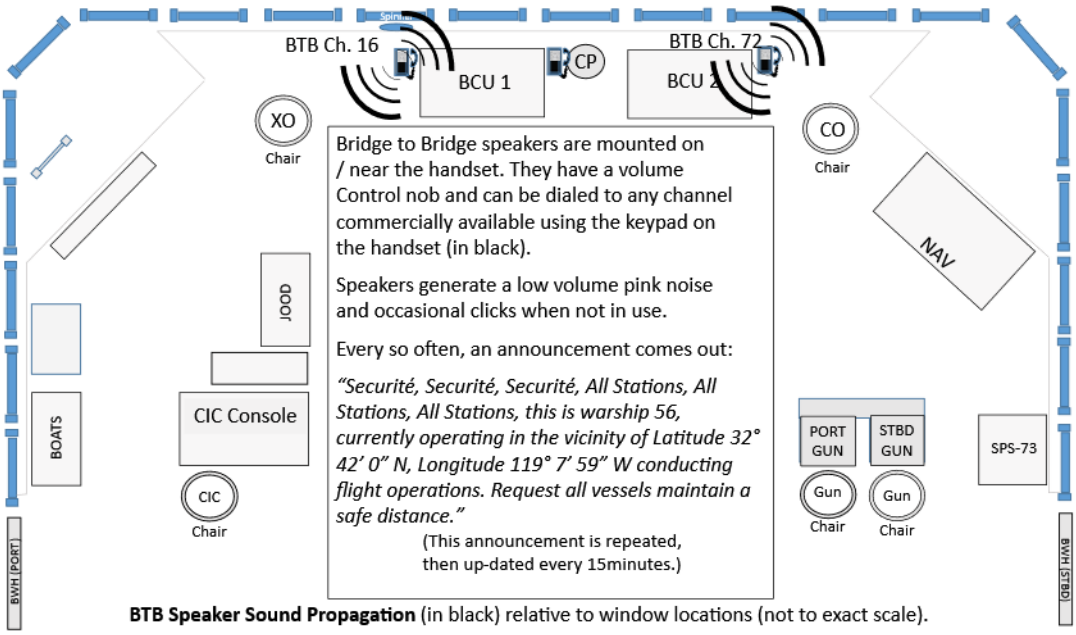
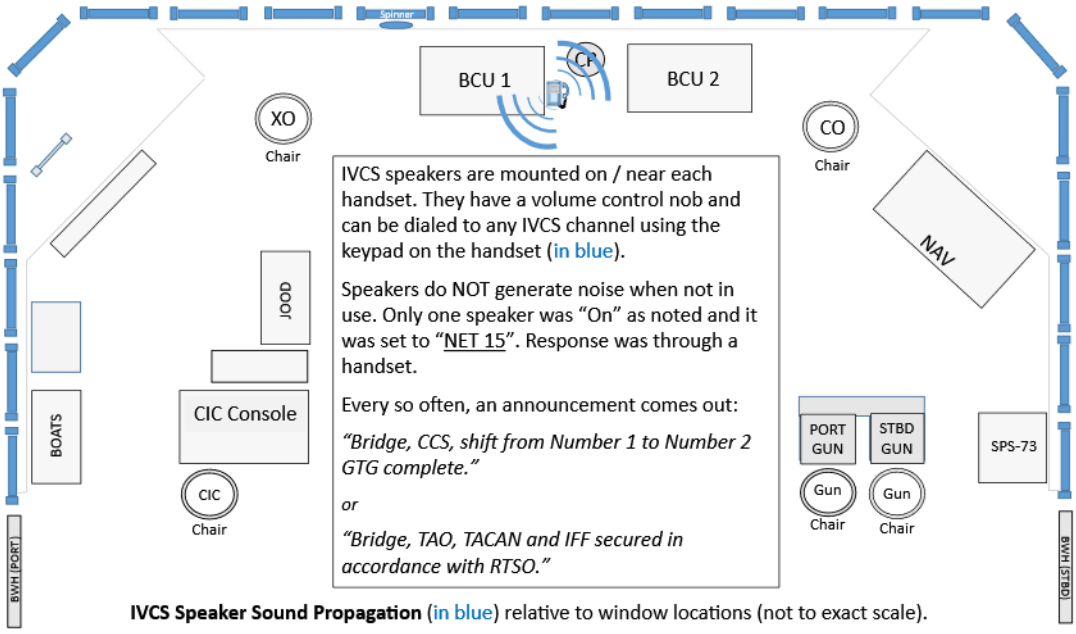


Figure A.12. Overhead tactical radio telephone speaker locations.



BTB Speaker Sound Propagation (in black) relative to window locations (not to exact scale).

Figure A.13. Bridge to bridge radio handset and speaker locations.



IVCS Speaker Sound Propagation (in blue) relative to window locations (not to exact scale).

Figure A.14. Bridge internal voice communications system handset location.

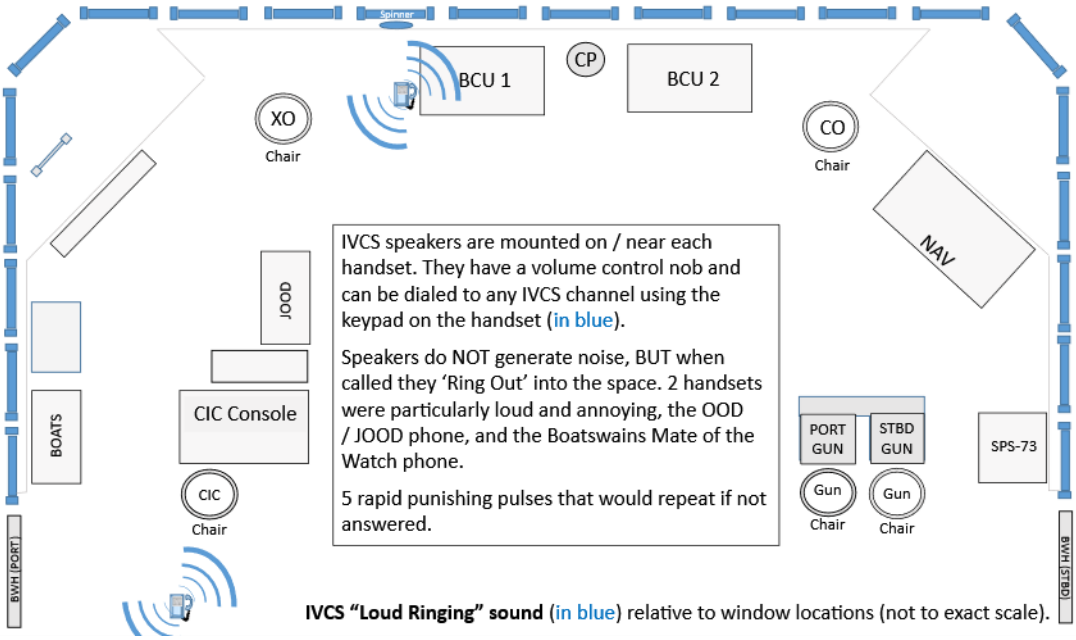


Figure A.15. Bridge internal voice communications system handset and speaker locations.

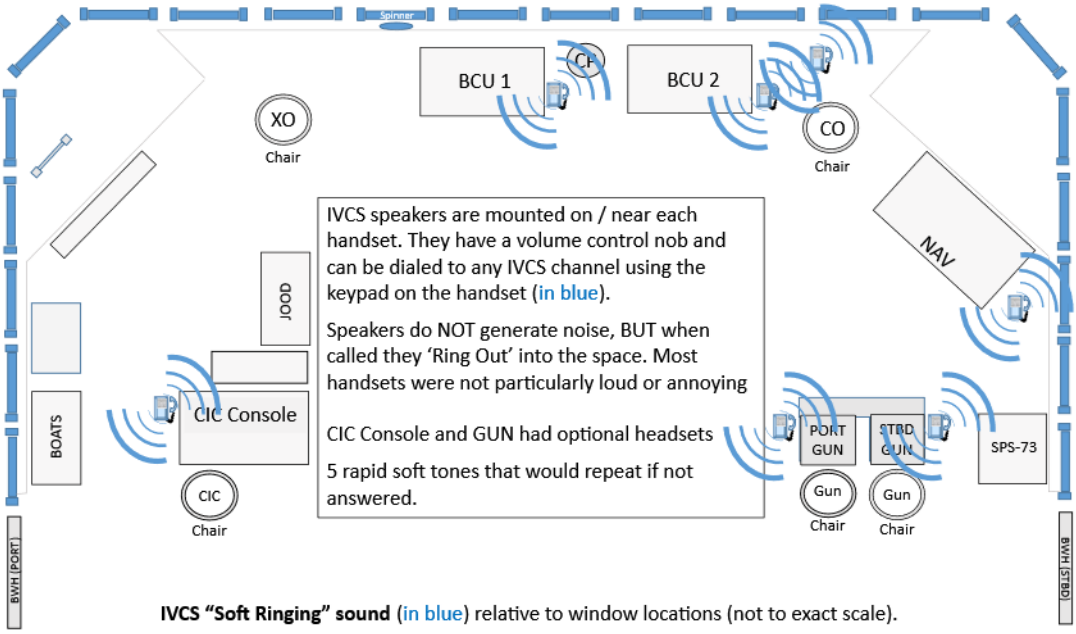


Figure A.16. Additional internal voice communications system handset and speaker locations.

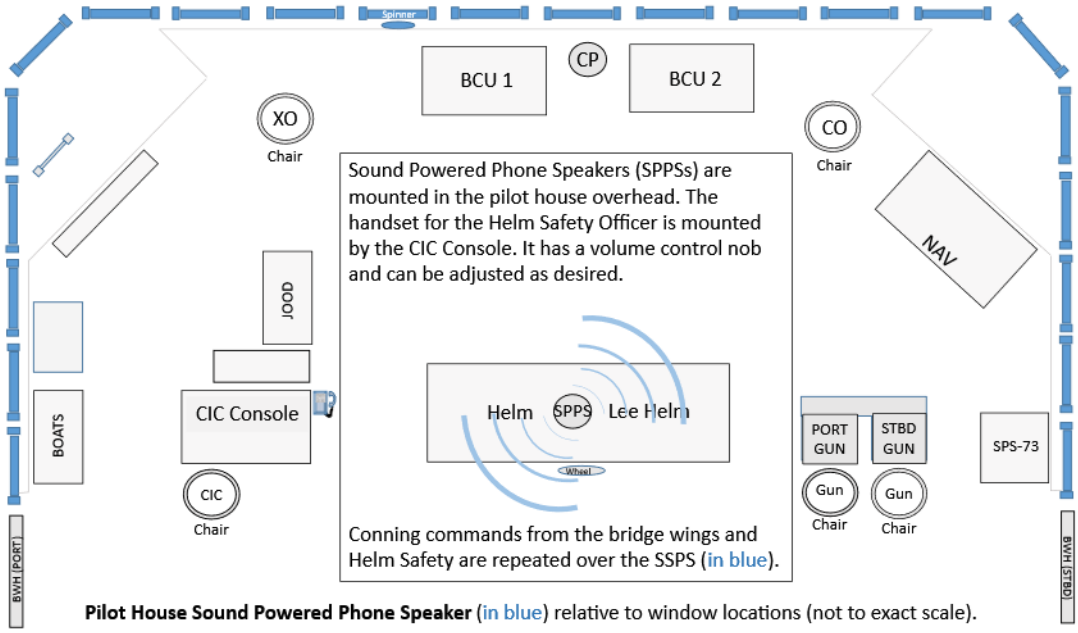


Figure A.17. Bridge central sound powered phone speaker location.

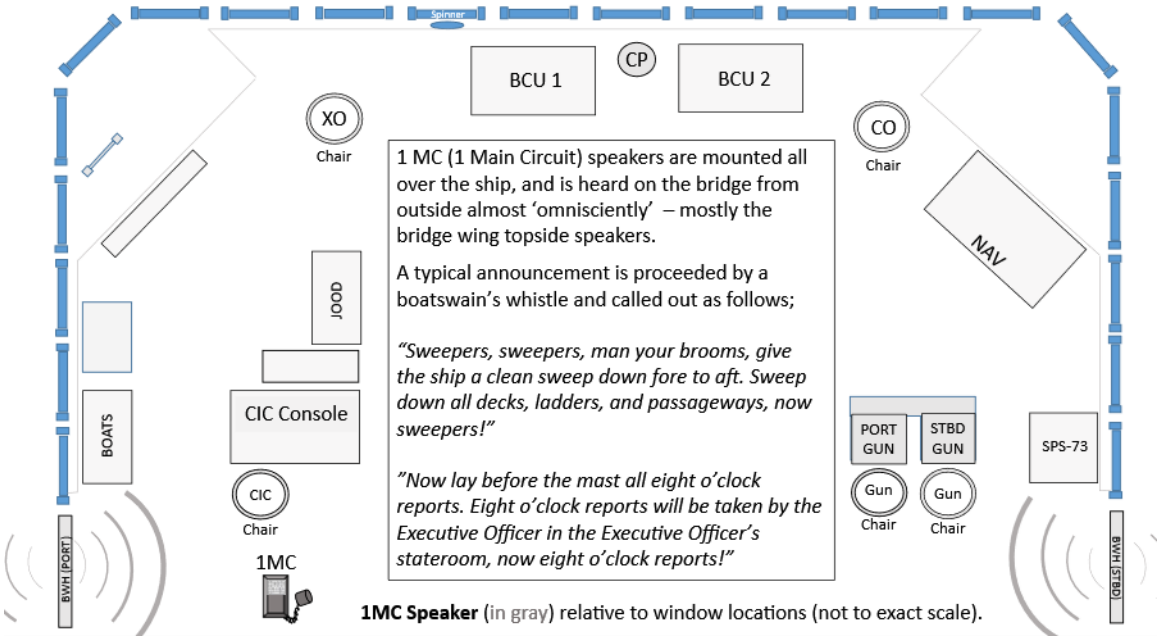


Figure A.18. Bridge 1 main circuit microphone and exterior sound source.

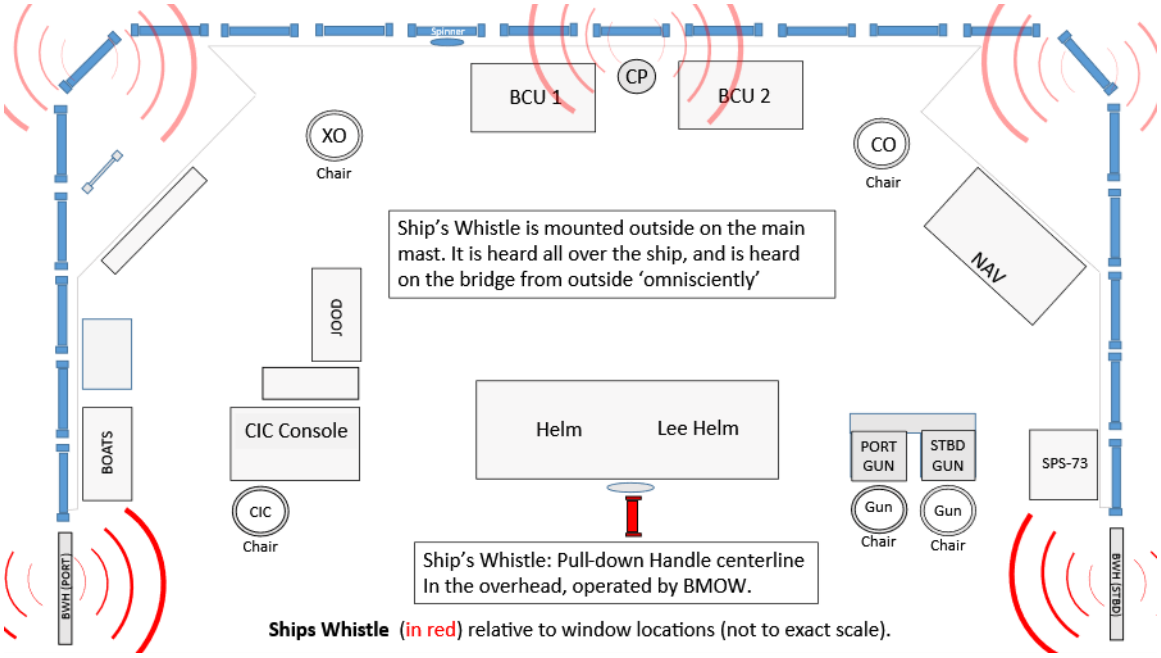


Figure A.19. Ship's whistle handle and sound source locations.

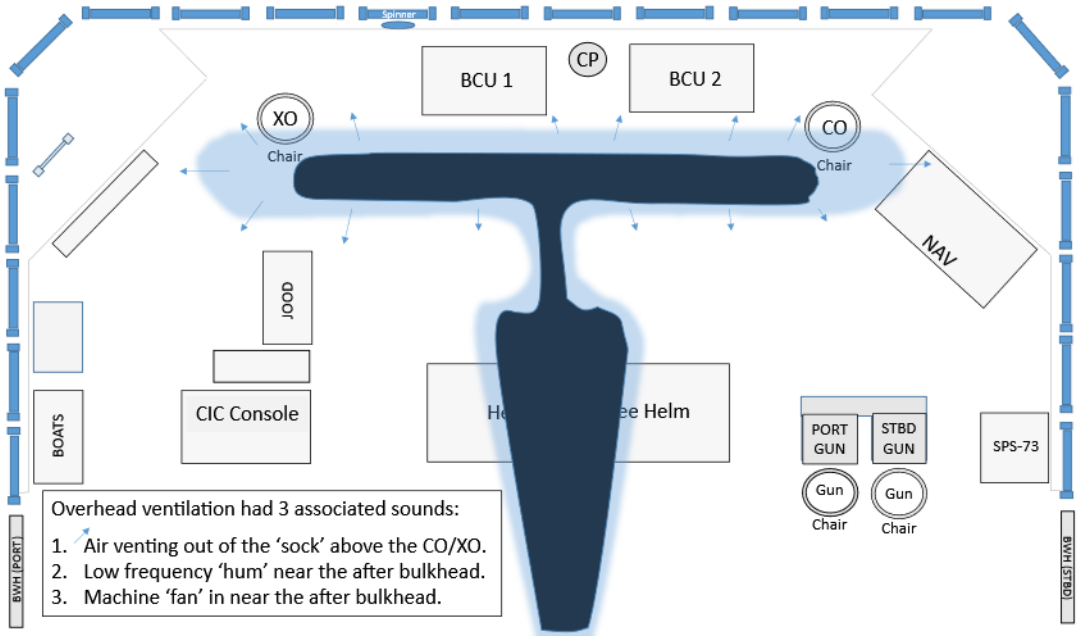


Figure A.20. Bridge overhead ventilation sound source location.

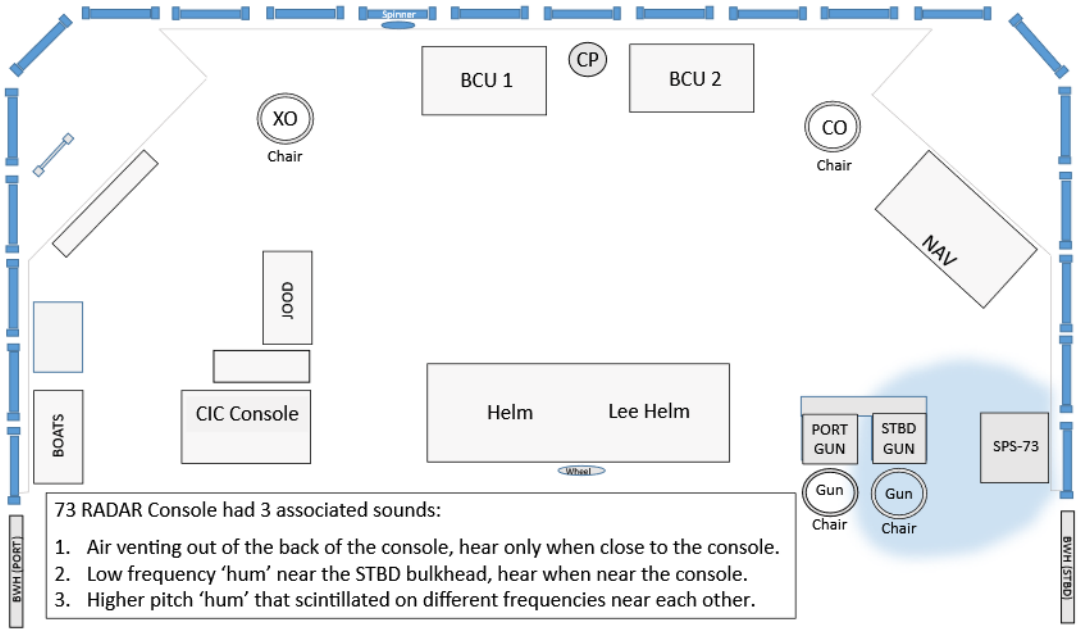


Figure A.21. Bridge SPS-73 radio detection and ranging sound source location.

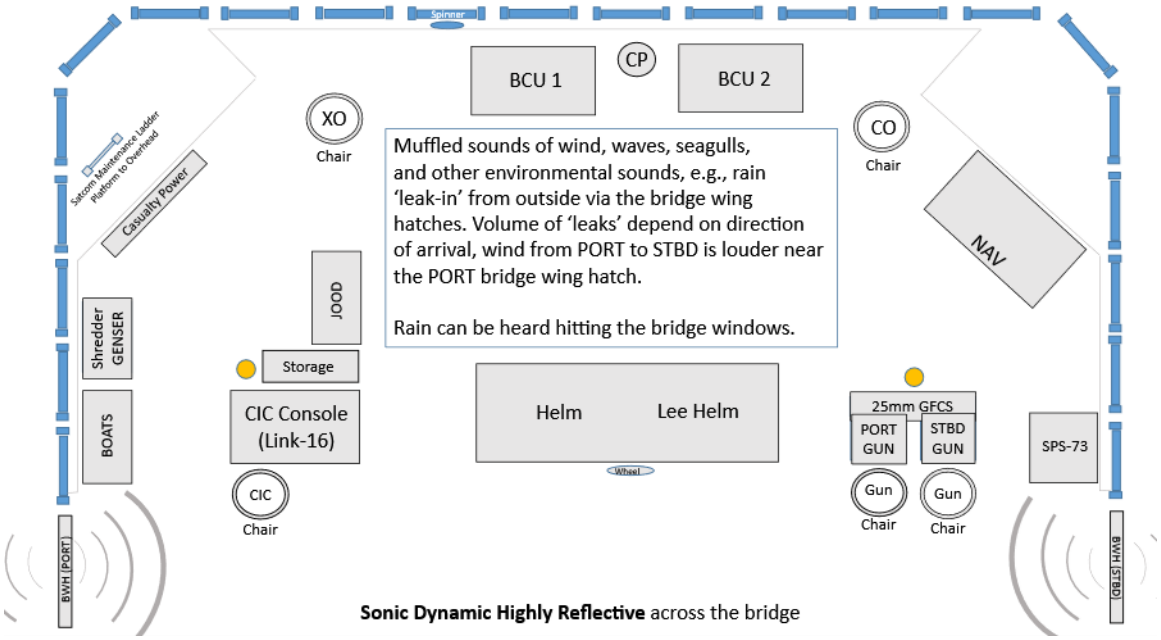


Figure A.22. Bridge external environmental sound scape locations.

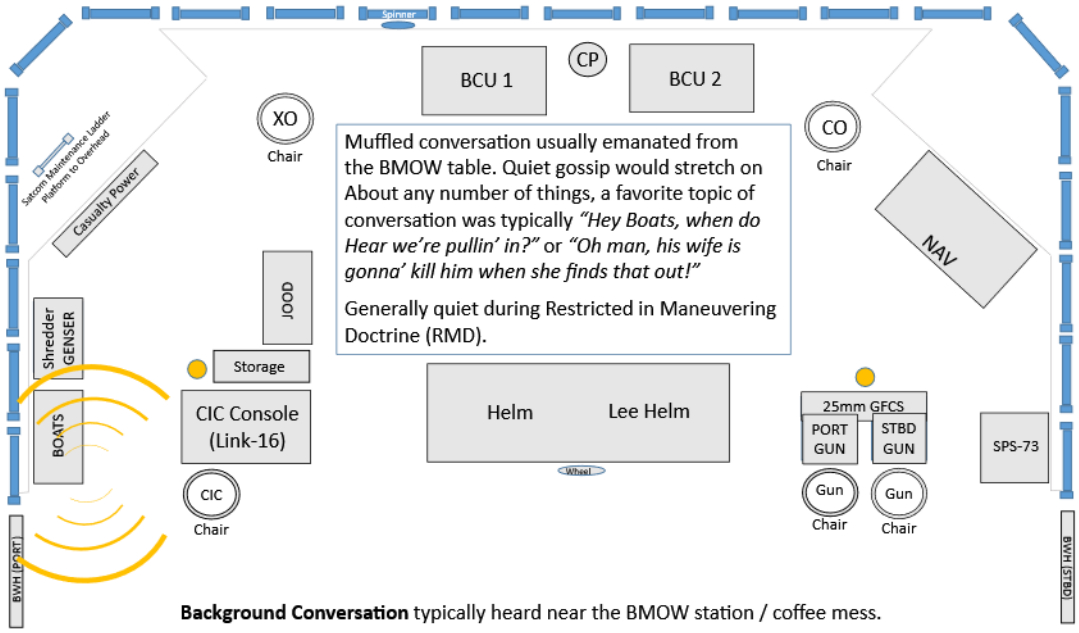


Figure A.23. Typical conversational background noise.

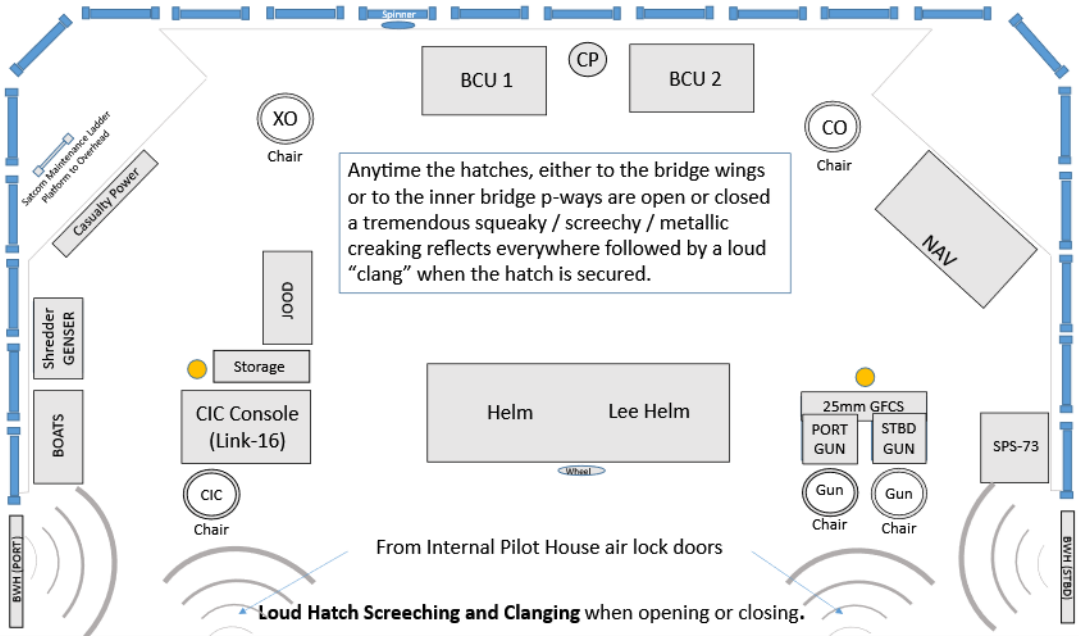


Figure A.24. Background bridge hatch noise.

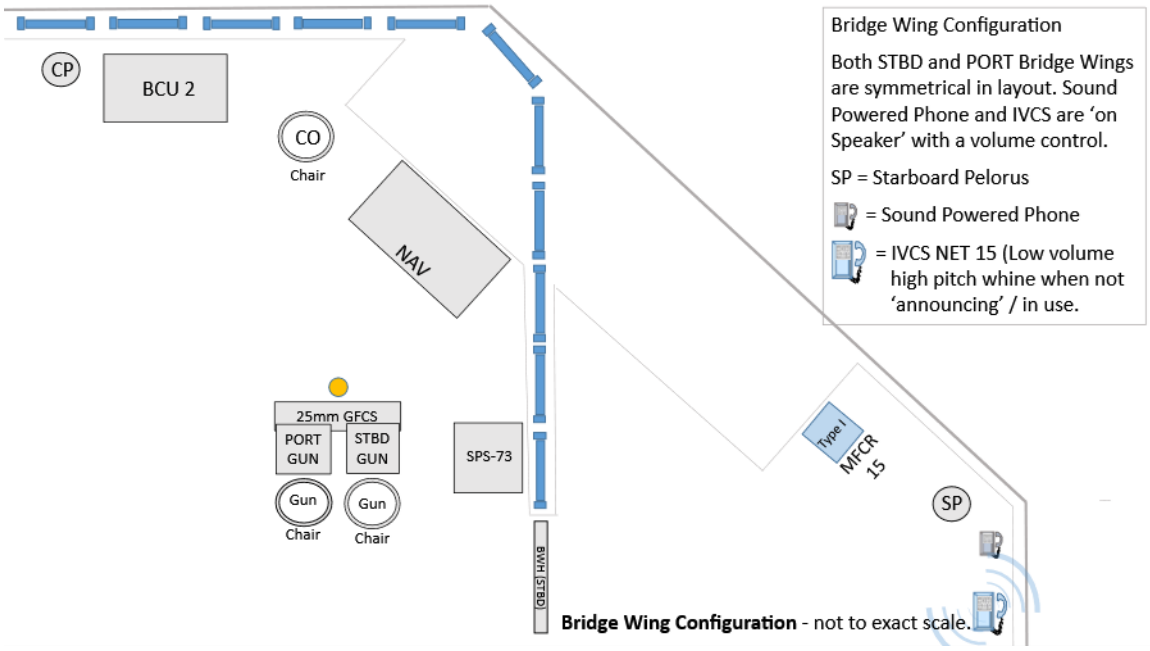


Figure A.25. Bridge wing configuration and internal voice communications system location.

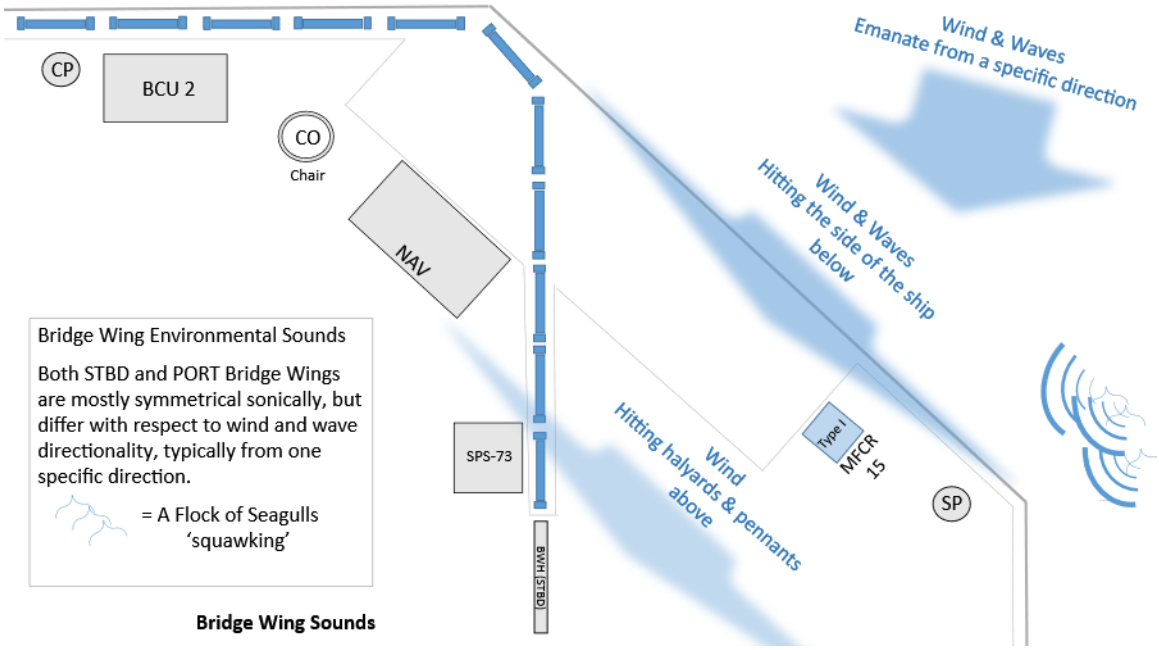


Figure A.26. Bridgewing environmental soundscape.

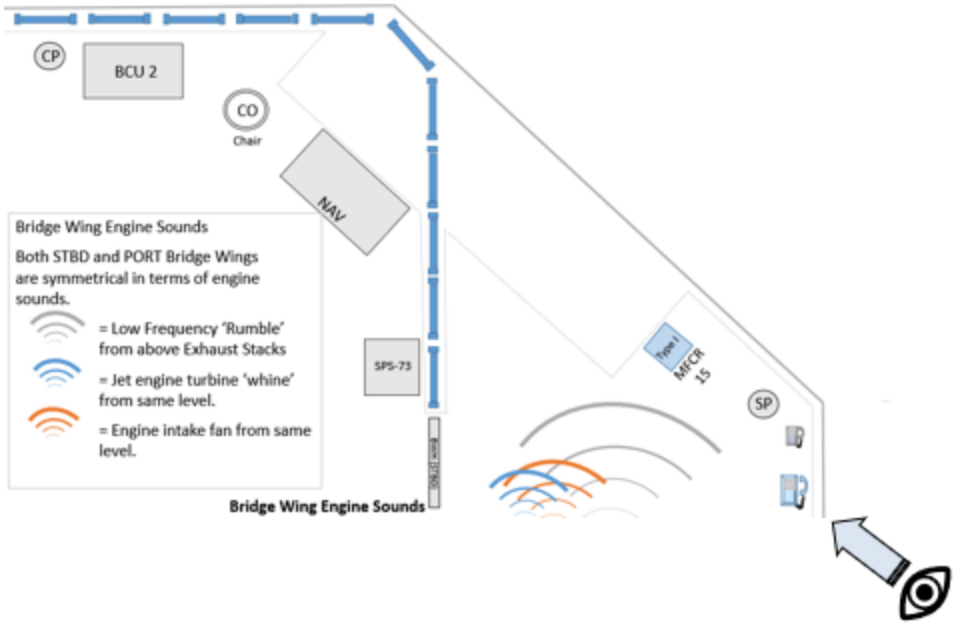


Figure A.27. Ship's powerplant and engine soundscape.

VR Grey Scale Projection



Figure A.28. Initial bridge virtual environment construct based on bridgewing description as seen from the starboard side of the ship looking into the bridge control units. The point of view is shown in Figure A. 27 using an "eye icon" and an arrow to indicate view direction.

APPENDIX B. EXTENDED REALITY EXTERNAL VIRTUAL ENVIRONMENT CONFIGURATION AND CONTROL

The SCC PC is used to create any kind of needed maritime scenario, e.g., entering a traffic separation scheme, and then adjusting the scenario during run time as needed to facilitate desired learning in the XR VE. This includes numbers of vessels, time of day, weather environments and currents, etc.

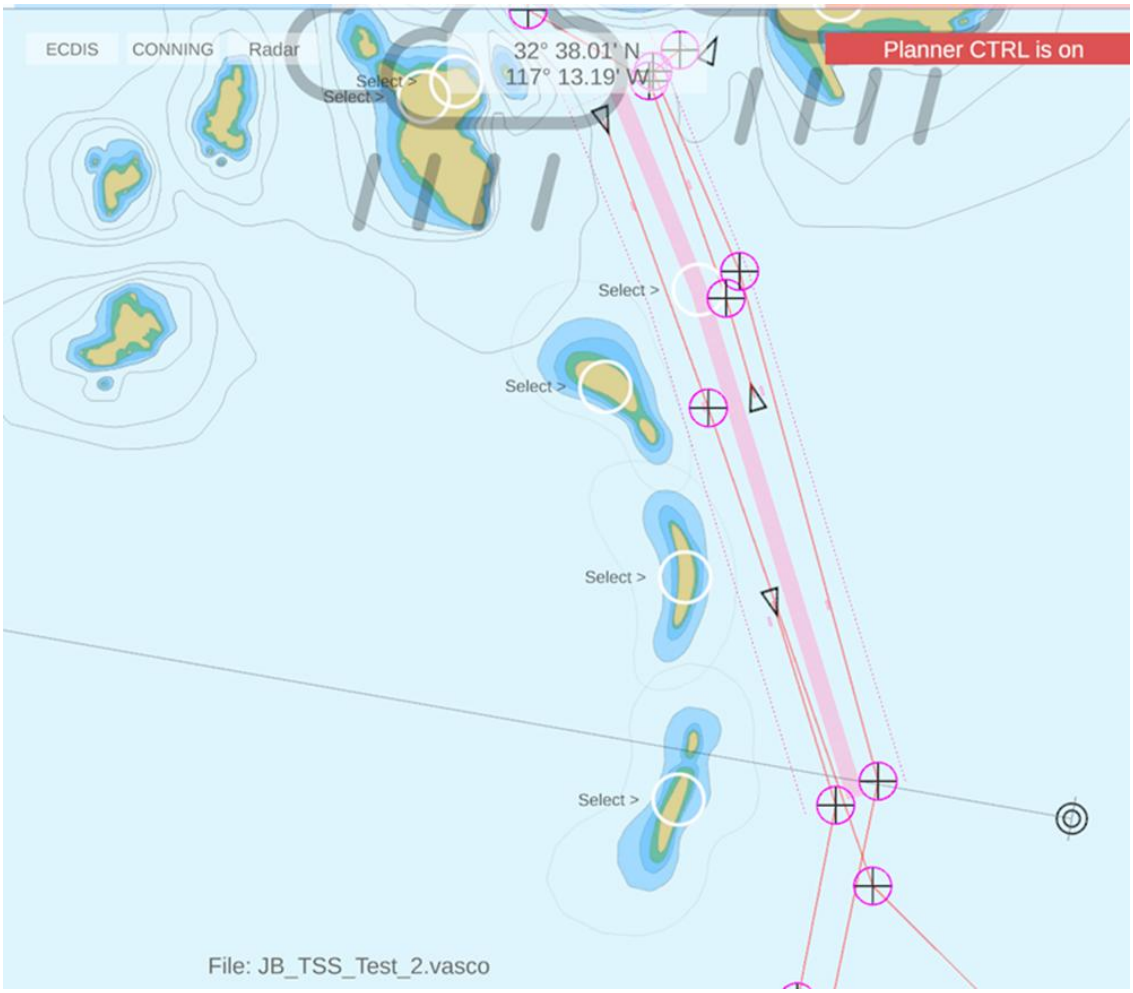
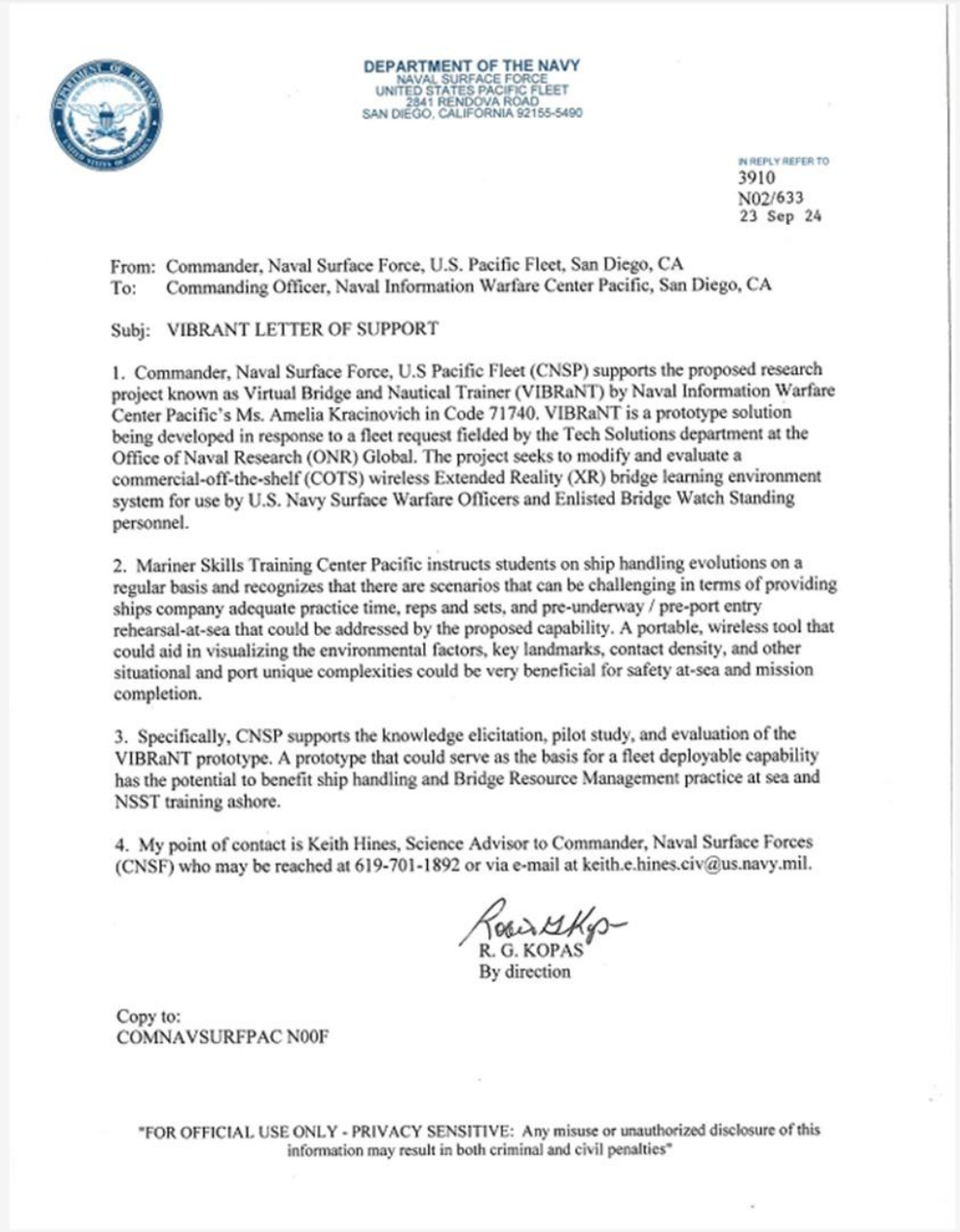


Figure A.29. Bridge virtual environment Scenario Creation and Control Personal Computer display of a constrained channel transit.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. COMMANDER NAVAL SURFACE FORCES SEPTEMBER 23, 2024 LETTER OF SUPPORT

Commander Navy Surface Forces Letter of Support (LoS) to conduct VIBRaNT testing at MSTCPAC. The LoS was required to conduct the experiment and collect data.



DEPARTMENT OF THE NAVY
NAVAL SURFACE FORCE
UNITED STATES PACIFIC FLEET
2841 RENDOVA ROAD
SAN DIEGO, CALIFORNIA 92155-5490

IN REPLY REFER TO
3910
N02/633
23 Sep 24

From: Commander, Naval Surface Force, U.S. Pacific Fleet, San Diego, CA
To: Commanding Officer, Naval Information Warfare Center Pacific, San Diego, CA

Subj: VIBRANT LETTER OF SUPPORT

1. Commander, Naval Surface Force, U.S Pacific Fleet (CNSP) supports the proposed research project known as Virtual Bridge and Nautical Trainer (VIBRaNT) by Naval Information Warfare Center Pacific's Ms. Amelia Kracinovich in Code 71740. VIBRaNT is a prototype solution being developed in response to a fleet request fielded by the Tech Solutions department at the Office of Naval Research (ONR) Global. The project seeks to modify and evaluate a commercial-off-the-shelf (COTS) wireless Extended Reality (XR) bridge learning environment system for use by U.S. Navy Surface Warfare Officers and Enlisted Bridge Watch Standing personnel.

2. Mariner Skills Training Center Pacific instructs students on ship handling evolutions on a regular basis and recognizes that there are scenarios that can be challenging in terms of providing ships company adequate practice time, reps and sets, and pre-underway / pre-port entry rehearsal-at-sea that could be addressed by the proposed capability. A portable, wireless tool that could aid in visualizing the environmental factors, key landmarks, contact density, and other situational and port unique complexities could be very beneficial for safety at-sea and mission completion.

3. Specifically, CNSP supports the knowledge elicitation, pilot study, and evaluation of the VIBRaNT prototype. A prototype that could serve as the basis for a fleet deployable capability has the potential to benefit ship handling and Bridge Resource Management practice at sea and NSST training ashore.

4. My point of contact is Keith Hines, Science Advisor to Commander, Naval Surface Forces (CNSF) who may be reached at 619-701-1892 or via e-mail at keith.e.hines.civ@us.navy.mil.

R. G. KOPAS
By direction

Copy to:
COMNAVSURFPAC N00F

"FOR OFFICIAL USE ONLY - PRIVACY SENSITIVE: Any misuse or unauthorized disclosure of this information may result in both criminal and civil penalties"

Figure A.30, Commander Naval Surface Forces letter of endorsement for VIBRaNT testing.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. PRE-EVENT COORDINATION EMAIL

Human Subjects Research IRB approved participant recruitment email.

(IRB Attachment A)

Participant Recruitment Email

Attention all Surface Warfare Officers! (And those seeking SWO qualification.)

The Office of Naval Research seeks Naval Officers and Enlisted Bridge Watch Team Personnel to evaluate a Virtual Bridge and Nautical Trainer (VIBRaNT) system.

Personnel will be competitively selected from those participating in MSTC PAC NSST training events taking place near Pier 7 at 32nd Street Naval Station. Those selected will be provided an opportunity to work with Scientists from Naval Information Warfare Center (NIWC) Pacific to shape the future of Naval training.

Participants will don Meta Quest 3 and Meta Quest Pro Virtual Reality (VR) Head Mounted Displays to join fellow watchstanders and instructors in a 30-minute bridge training experience inside a virtual DDG-51 class bridge environment.

Those interested please contact Ms. Amelia Kracinovich, NIWC PAC Code 71740, amelia.r.kracinovich.civ@us.navy.mil (204) 351–6091 or Jon Brewster, NIWC PAC Code 75672, jon.m.brewster.civ@us.navy.mil (858) 952–9311.

[NOTE: Participation is completely voluntary but because this study utilizes VR, there is the potential for simulator sickness, which is a kind of motion sickness. Therefore, if you regularly experience motion sickness, you SHOULD NOT participate in this study. Similarly, if you are prone to vertigo or seizures, we recommend you do not participate as symptoms such as nausea, disorientation, and sweating have occurred. If at any time you wish to withdraw from the study, you may do so without concern for any sort of repercussion; your safety and comfort are a very important part of this experiment and will remain a priority at all times. Data collected from this research will be anonymous– no data can be traced back to your identity. Additionally, none of the collected data will be used in any of your future performance evaluations.]

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX E. PRIVACY STATEMENT

IRB approved Privacy Act Statement to inform participants of data collection being performed, potential data distribution, and rights to withdraw from the study at any time.

(IRB Attachment B)

Privacy Act Statement

To be read by all voluntary Human Subjects prior to signing the Informed Consent Document

Authority. SECNAVINST 5211.5E, Department of the Navy (DON) Privacy Program, paragraph 10(a) page 24, and paragraph 10(d) page of 26, December 2005

Purpose. Human performance data and other research information will be collected in an experimental project entitled "Virtual Bridge and Nautical Trainer."

Routine Uses. The Departments of the Navy and Defense, and other U.S. Government agencies will use the resulting research data for analyses and reports. Use of the information obtained may be granted to non-Government agencies following the provisions of the Freedom of Information Act or contracts and agreements. I voluntarily agree to its disclosure to agencies or individuals identified above and I have been informed that failure to agree to this disclosure may make the research less useful. The "Blanket Routine Uses" that appear at the beginning of the Department of the Navy's compilation of data bases also apply to this system.

Voluntary Disclosure. Participation in this study and provision of information is voluntary. However, failure to provide requested information may invalidate test data and/or test procedures and could therefore result in removal from the project. Dismissal from the research project will involve no reproach, prejudice or jeopardy to my job or status.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX F. INFORMED CONSENT DOCUMENT

IRB ICD stating approved details of study, study participation expectations, and right to withdraw at any time.

(IRB Attachment C)

INFORMED CONSENT DOCUMENT

Protocol Title: VIrtual BRidge Nautical Trainer (VIBRaNT)
Principal Investigator: Ms. Amelia Kracinovich (Code 71740), amelia.r.kracinovich.civ@us.navy.mil
Co-Principal Investigator: Mr. Jon Brewster (Code 75672), jon.m.brewster.civ@us.navy.mil
Associate Investigators: Dr. Jonathan Mun (NPS), Dr. Mollie McGuire (NPS), and Jacob Anderson (Code 81510)
NIWC Pacific IRB eProtocol Number: NIWCPAC24-06-102

INTRODUCTION

This research is funded by the Office of Naval Research Global (ONR G), being performed, and partially funded by Naval Information Warfare Center Pacific (NIWC Pacific). The intent of the study is to aid Navy efforts to evaluate and characterize Extended Reality (XR) technology for use in ship bridge team training by benchmarking it to currently fielded Navigation, Seamanship, and Shiphandling Training (NSST) mockup training systems.

PURPOSE OF RESEARCH

This research will evaluate and characterize XR technology facilitation of shiphandling and bridge resource management skills training.

You along with your bridge team will be observed in two training sessions in order to: 1) collect data during your 'normal' NSST training, 2) collect data during a 30-minute virtual event (XR) similar to the NSST and, 3) statistically compare the data results of each session to evaluate XR facilitation of bridge training. The XR virtual event will involve wearing Meta Quest Head Mounted Display (HMD) technology that projects a virtual environment obstructing your view of the 'real world'. While you are performing your role on the bridge in XR, your fellow watchstanders and instructors will also be performing their respective roles in the same virtual environment, able to interact with each other just as you would in the NSST mockup simulator. This is to investigate future training capabilities to practice with your crew when not co-located or otherwise not able to travel to a NSST mockup. Additionally, to characterize XR's usefulness, the investigators will provide surveys for you to complete after each observation session.

DURATION OF INVOLVEMENT

In addition to your regularly scheduled NSST training session, you will be participating in one 70-minute session consisting of one 30-minute XR evaluation session, with an additional 40 minutes allotted to complete surveys for both the NSST and the XR training scenarios, pre-briefing and screening, and HMD fitting and safety. Survey responses are limited to circling numbers on a scale,

where the time needed to complete each survey depends on each individual. Consent is being sought to collect data limited to one NSST session and one XR session as noted above.

PROCEDURES

1. Prior to beginning observations, investigators will provide a briefing on the research to include general information about the study, study risks, and provide an opportunity to ask questions and/or express concerns. Expected to last ~ 5 minutes.

2. After the briefing, the investigators will issue (in addition to this Informed Consent Document), 1) a Privacy Act Statement, 2) a Pre-Screening Questionnaire and, 3) a Simulator Sickness Questionnaire (only during XR training). Information you provide will be 'de-identified' using a code and NOT include any Personally Identifiable Information (PII) or Health related information. This code will be randomly assigned and only traceable on a master key which will be stored on an encrypted hard drive and locked in a cabinet at NIWC Pacific. You'll receive your unique code after signing the ICD and write it on every form thereafter, starting with the Pre-Screening Questionnaire. Data collected WILL include your bridge watch team assignment, e.g., "Conning Officer," "Helmsman," "Navigator," etc. Filling out/signing these forms is expected to take ~ 5 minutes.

3. Afterwards, the investigators will begin observations of you and your bridge team in either the NSST or in the XR trainer. If you're training first in the NSST, you'll simply report to the NSST simulator at the time designated, perform your training, and fill out the study surveys (see above) after the NSST session is complete. Expected to require ~ 10 minutes.

4. The investigators will provide an XR equipment safety brief, usage guidelines, and go over a general description of what to expect in the XR training environment before beginning the XR portion of this study. In addition, investigators will assist in donning the equipment and ensuring you are OK to proceed with the study session. The expected time is ~ 10 minutes.

   [NOTE: <u>You may ask questions about the system at any time and may spend as much time as necessary to familiarize yourself with the equipment until you feel comfortable proceeding.</u> A study team member will be with you at all times, to ensure your safety while using the equipment.]

5. Once you feel comfortable with the XR equipment, you and your bridge team will perform several shiphandling evolutions using pre-determined scenarios. A study team member will be with you at all times to observe your performance as well as for your safety to ensure you won't trip and/or bump into things when using the XR equipment. You do not need to physically move around to navigate the XR environment, you will instead be using a teleportation feature that the researchers will review with you. Investigators will be collecting performance data, e.g., reaction time, time to complete actions, etc., while you are performing the scenario. The XR scenario will take ~30 min to complete.

6. After each scenario, the XR equipment will be removed, and you will be asked to complete a series of questionnaires to include: 1) a System Usability Scale–10 questions, 2) a NASA Task Load Index–6 questions, 3) Bridge Watch Team Self-Assessment–29 questions and, Exit Interview–3 questions. Expect time to complete is ~ 10 minutes.

[NOTE: For Instructors, we ask you monitor the team performance while watchstanders train in each environment and then complete the following surveys: 1) System Usability Scale–10 questions, 2) a Bridge Watch Team Assessment–33 questions, and 3) an OOD Assessment–28 questions. Expected time to complete ~ 10 minutes.]

7. After both NSST and XR training observation sessions have concluded and surveys have been submitted, your participation is complete.

   [NOTE: At the conclusion of the study, investigators can provide you an out brief of your team's performance in XR at a broad level, but specific performance metrics will not be kept or provided. If desired, you will be provided with a copy of your informed consent document and privacy statement to take home with you.]

SPECIAL PROCEDURES FOR OFFICERS OF THE DECK (OODs)

1. Sessions performed in the NSST will NOT involve any additional equipment unless you are participating as the OOD. Bridge Watch Team Officers of the Deck (OODs) will be asked to wear eye tracking glasses. This equipment is similar to sunglasses in weight and appearance, completely passive, non-corrective, and will not obstruct the OOD's ability to perform in the NSST. Data collected will be used to assess eye gaze vector data and eye-blink rate for analysis in comparison to the eye tracking data collected in the XR condition using a Meta Quest Pro–also a completely passive device.

2. All eye tracking data collected will be de-identified using a code, available only to the NIWC research team, and physically secured in a locked cabinet when not being analyzed.

3. During the XR session, the OOD will wear a Meta Quest Pro HMD which performs nearly identical to the Meta Quest 3 HMDs used by his/her watch team and all data collected will be subject to the same security conditions noted above.

4. Although collection of eye tracking data is highly desired for this study, OOD's who do NOT want to wear eye tracking equipment may at any time opt out of this collection effort for either the NSST or XR data collection.

RISK AND DISCOMFORT

Because this study utilizes HMD systems, there is the potential for simulator sickness, which is a kind of motion sickness. Experiences of symptoms such as nausea, disorientation, and sweating have occurred for some individuals using HMD equipment like the Meta Quest 3. If you are feeling any of these symptoms, you should verbally report this to a NIWC research team member. Any verbal report of symptoms, or a high score on the Simulator Sickness Questionnaire, will result in your withdrawal from the study. This research is considered to be minimal risk. That means that the risks associated with this study are the same as what you face every day. However, because this study utilizes HMD systems, there are small things to be aware of. If you have epilepsy, it is highly recommended that you do not participate in this study. If you are sensitive to extra weight straining your neck or having a strap or band tightened around your head, it is recommended that you do not participate. You will be given the option for your data to be deleted, and an anonymized report of the incident will be made to the IRB committee. Participants who suffer symptoms will be asked

by the research team to remain in the experiment room until they feel better. Attempts will be made to bring them to different room from their fellow participants based on space availability and participants ability to move. Participants will be instructed to focus on one point in space to help alleviate symptoms. Symptoms are only temporary and should subside after 15 minutes. If you wish to withdraw from the study, you may do so without penalty of concern for any sort of repercussion, (Your safety and comfort are a very important part of this experiment and remain a priority at all times) however, the research team CANNOT guarantee anonymity as others on your bridge team may disclose your withdrawal from the study, but anonymity will be encouraged to all participants..

There is a small risk where participants will collide with real world environmental objects while using the HMD system. When using an HMD system, you will be monitored at all times to ensure that you do not collide with real world environmental obstacles (e.g., walls, furniture). The surrounding area around each participant will also be cleared and free of as many environmental objects as possible to prevent any tripping or collision risk. You should not need to move to operate the XR system and instead use a 'teleport' feature to move within the virtual environment. You will be continually monitored by the study team, moving obstructions out of the way, verbally telling participants to stop moving, or touching them to prevent further movement.

POTENTIAL BENEFITS

You will not be paid for your time and there are no known direct benefits to you participating in this study. Potential benefits may include the ability to enable Navy bridge team's opportunity to practice more often and thus enhance their performance, confidence, and team cohesion. Participants will be able to share their expertise with researchers and provide their feedback as potential users of a system that may ultimately benefit the Navy. Participation in system effectiveness evaluations will aid in characterizing XR technology capacities and aid in targeting its effectiveness to 'best suited' instructional practice needed to better the Fleet.

ALTERNATIVES TO PARTICIPATION

The alternative to participation is not to participate in this study. The research team will at no time disclose whether you choose to participate or withdraw, however, the research team CANNOT guarantee anonymity as others on your bridge team may disclose your withdrawal from the study, but anonymity will be encouraged to all participants.

VOLUNTARY PARTICIPATION AND WITHDRAWAL

Participation in this study is voluntary, and you may withdraw from this study at any time for any reason. Participation in this study will not affect your employment nor will your job performance be contingent upon participation in this research in any way. Both your participation (and your withdrawal, if you choose to withdraw) will always remain anonymous.

By volunteering for this study, you have agreed to provide honest and ethical responses and perform to the best of your ability throughout the testing procedure. At any time, for any reason you may withdraw any data provided during the study by informing any of the investigators. If you wish to withdraw after your participation is complete, you may contact the PI, Ms. Amelia Kracinovich at (619) 806–1716 /amelia.r.kracinovich.civ@us.navy.mil, and provide your participant number. This data will then be destroyed and will not be included in any part of the

analysis with the exception for video/audio capture. Once the participation session has been completed, there will be no way for the research team to remove individuals from the recording as the recording is a collective of the whole participation group.

CONFIDENTIALITY

Any data collected from this research will be anonymous–no data can be traced back to your identity. Prior to the collection of any test data, you will be randomly assigned a code number. That number will be the only identifier used when referring to your data, and therefore your personal information will be kept completely separate from the results of the study. No one other than the study investigators will have access to the data. Only data summarizing across all individual data sets will be included in any internal or external reports and these technical reports will be provided to ONR, the Surface Warfare Schools Command, Mariner Skills Training Center Pacific, Navy Postgraduate School, and Surface Forces Pacific. Additionally, none of the collected data will be used in any of your future performance evaluations.

NSST video and audio recordings will not be taken or retained by the research team. The research team may review video and audio information recorded by NSST instructors for XR comparisons, but reviews will only be made for comparison to the XR condition and identity information will NOT be retained. You may opt out of video/ audio analysis of the NSST, but the data will still be collected by the NSST as part of your regular training regime.

Data obtained by the NIWC Research Team, to include XR virtual environment recordings, eye tracking, and survey data, will be retained in locked containers when not in use and held on NIWC Pacific Building 365 premises for 5 years after the research collection efforts have concluded. Only members of the NIWC Research Team have access to this data and all data will only be identifiable via code.

FUTURE DATA USAGE

De-identified data may be used for future research studies or distributed to other investigators without additional informed consent from the subject. At no point will attempts be made to re-identify the data. Records will be retained for three years post study completion. Requests for access to de-identified data should be made to the study PI and will require a data sharing agreement to be established between NIWC Pacific and the interested party. Data analysis will also be included in an NPS Information Science PhD Dissertation by Jon Brewster titled *Enhancing Knowledge Flow with XR. N*o additional data will need to be collected for this effort.

CONTACT INFORMATION

For inquiries regarding your participation in this study or data analysis, please contact the Principal Investigator (PI), Ms. Amelia Kracinovich at (619) 806–1716/ amelia.r.kracinovich.civ@us.navy.mil. For questions about your rights as a volunteer participant or any problem related to protection of research volunteers, please contact the Chair of the NIWC Pacific Institutional Review Board (IRB), Dr. Kara Sorensen, at (619) 553–1340 or via email at kara.c.sorensen.civ@us.navy.mil.

SIGNATURE OF RESEARCH SUBJECT

I have read the Privacy Act Statement, this voluntary consent form and the approval letter for this study issued by the NIWC Pacific Commanding Officer. I have been given the opportunity to discuss any questions or concerns that I might have with the researchers and had all my questions answered to my satisfaction.

_____
Participant's Printed Name

Eye tracking consent (if applicable): Opt-in ☐ Opt out ☐

_____
Participant's Signature Date

_____
Principal Investigator Date

APPENDIX G. PRE-SCREENING QUESTIONNAIRE

IRB approved pre-screening questionnaire to survey participant backgrounds and experience levels and ensure no prior reported history of motion sickness.

(IRB Attachment D)

Pre-Screening Questionnaire

Participant #: _____

1. Gender (circle one): Male Female
2. Age:_____years old
3. Military Status (circle one): Active Duty Retired Civilian

   CONN Qualified? (circle one): Yes No

   JOOD Qualified? (circle one): Yes No

   OOD Qualified? (circle one): Yes No

   SWO Qualified? (circle one): Yes No
4. Length of Shiphandling Experience:_____Years_____Months
5. Experience with the following technologies (1=No Experience, 7=Extensive Experience):

Shiphandling Simulators (circle one):	1	2	3	4	5	6	7
(e.g., the Conning Officer Virtual Environment, COVE, etc.)							
Virtual Reality:	1	2	3	4	5	6	7
Augmented Reality:	1	2	3	4	5	6	7
PC/Console Gaming:	1	2	3	4	5	6	7
Mobile/Phone Gaming:	1	2	3	4	5	6	7
General Mobile Device Usage:	1	2	3	4	5	6	7
General Personal Computer Usage:	1	2	3	4	5	6	7

6. Have you previously experienced (circle one)

Motion Sickness:	Yes	No
Vertigo:	Yes	No
Epilepsy:	Yes	No

APPENDIX H. SIMULATOR SICKNESS QUESTIONNAIRE

IRB approved simulator sickness questionnaire to ensure participants can safely test the VIBRaNT XR system.

(IRB Attachment E)

Simulator Sickness Questionnaire (SSQ)

Participant #: Session #:

Instructions: Circle the items that apply to you **RIGHT NOW**.

SYMPTOM	RATING			
	0	*1*	*2*	*3*
1. General Discomfort	None	Slight	Moderate	Severe
2. Fatigue	None	Slight	Moderate	Severe
3. Headache	None	Slight	Moderate	Severe
4. Eye Strain	None	Slight	Moderate	Severe
5. Difficulty Focusing	None	Slight	Moderate	Severe
6. Salivation Increased	None	Slight	Moderate	Severe
7. Sweating	None	Slight	Moderate	Severe
8. Nausea	None	Slight	Moderate	Severe
9. Difficulty Concentrating	None	Slight	Moderate	Severe
10. "Fullness of the Head"	None	Slight	Moderate	Severe
11. Blurred Vision	None	Slight	Moderate	Severe
12. Dizzy (eyes open)	None	Slight	Moderate	Severe
13. Dizzy (eyes closed)	None	Slight	Moderate	Severe
14. Vertigo	None	Slight	Moderate	Severe
15. Stomach Awareness**	None	Slight	Moderate	Severe
16. Burping	None	Slight	Moderate	Severe

17. Other: Please describe: _____

_____

_____

** "Stomach Awareness" usually is used to indicate a feeling of discomfort that is just short of nausea.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX I. VIRTUAL BRIDGE AND NAUTICAL TRAINER APPARATUS

The basic IT specifications of the VIBRaNT XR system PUK are as follows:

- VIBRaNT SCC PC Laptop

1x HP Z Book Firefly 14 Core i7-1165G7 32GB RAM, 512GB SSD 14IN FHD Win 11 Pro–NVIDIA Quadro T500 (4GB Graphics).

- VIBRaNT Wifi Router

1x NETGEAR RAX20 4-Stream AX1800 Wifi 6 Router.

- VIBRaNT VE Server

1x HP EliteDesk 800G4 Mini i5-Gen8 8GB 256GB SSD Win 10 Pro.

- VIBRaNT HMDs

11x Meta Quest 3 HMDs configured with a Meta Quest 3 Elite Strap with a Battery Extender, loaded with the VIBRaNT VE presentation software.

- VIBRaNT HMD with eye tracking

1x Meta Quest Pro HMD loaded with the VIBRaNT VE presentation and eye tracking software.

- VIBRaNT Equipment Carrying Case

1x Pelican Case with foam inserts was also provisioned to permit safe transportation of all VIBRaNT equipment is also a component of the VIBRaNT 'carry-on' system.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX J. NAVAL INFORMATION WARFARE CENTER PACIFIC VIRTUAL BRIDGE AND NAUTICAL TRAINER INSTITUTIONAL REVIEW BOARD PROTOCOL NIWCP.2024.0003-IR-EP7-A NOVEMBER 19, 2024

NAVAL INFORMATION WARFARE CENTER PACIFIC IRB RESEARCH PROTOCOL

Title: VIrtual BRidge and Nautical Trainer (VIBRaNT)

Planned dates of study: 1 May 2024–31 May 2025

1. Principal Investigator: Ms. Amelia Kracinovich (Code 71740)

2. Associate Investigators: Dr. Jonathan Mun (NPS), Dr. Mollie McGuire (NPS), Mr. Jon Brewster (Code 75672), Mr. Jacob Anderson (81510)

3. Background:

Shiphandling and bridge resource management is a complex, spatial-temporal activity requiring extensive training for Naval Officer's to gain and retain skills required of an Officer of the Deck (OOD) and qualify as a Surface Warfare Officer (SWO). While shiphandling is paramount, OOD's also act on behalf of the Commanding Officer (CO) and must be capable of managing many related tasks to ensure safe operation of the ship. These include but are not limited to: (1) maintaining overall situational awareness of air, surface, and subsurface contacts, (2) monitoring ship's engineering, sensors, weapons systems status, (3) managing internal and external communications, (4) managing deck operations and, (5) ensuring proper orders are given at the right time to meet the task at hand.

Because shiphandling, situational awareness, and bridge resource management are embodied cognitive processes, Naval Information Warfare Center Pacific (NIWC PAC) research team proposes the use of Extended Reality (XR) to support Naval Officer training to gain and maintain proficiency as OOD's and SWO's. XR, or spatial computing, refers to the use of Head Mounted Display (HMD) technology able to a provide embodied three-dimensional (3D) visual and audio Virtual Reality (VR), Augmented Reality (AR), Mixed Reality (MR), or Augmented Virtuality (AV) presentations combined with hand-held haptic controllers, and conventional Information Technology (IT, e.g., laptop computers, network routers, Internet, etc.) to facilitate learning previously limited to physical mockup environments. In this specific case, XR is meant to include the VR HMD, haptic controllers, the Scenario Creation and Control System PC, and Wifi enabled local area network router, server and Internet backbone technologies comprising the system.

To perform this study, NIWC PAC intends to modify and evaluate a low-cost Commercial Off The Shelf (COTS) XR training system (Kilo-Solution's 'Virtual Advancement System for Competent Operations' (VASCO)) for developing and enhancing watch standing competence. Modification of VASCO will include addition of two new 3D models of Norfolk and San Diego port approaches, along with a DDG-51 class guided missile destroyer VR bridge environment. We refer here to the resulting VASCO system modifications as the Virtual Bridge and Nautical Trainer or "VIBRaNT." Once developed, VIBRaNT will be compared to the United States Navy

(USN) Navigation, Shiphandling, Seamanship Training (NSST) system to evaluate its ability to effectively facilitate training of navigation and shiphandling skills. The VIBRaNT research team, working in cooperation with USN shiphandling Subject Matter Experts (SMEs), will perform the VIBRaNT comparison to the NSST at the Maritime Seamanship Training Center located at Naval Base San Diego. Researchers hypothesize the visual and auditory presentation provided by VIBRaNT's XR platform will enable knowledge acquisition and transfer of critical spatial-temporal understanding, in some cases equal to or better than the legacy platform provided by NSST. To perform the comparison, a 'counter-balanced within subjects' experimental design will be employed such that participants will perform the same training scenarios on both training platforms, NSST and VIBRaNT respectively. Training results will be evaluated based on (1) time to task (reaction time), (2) time to task completion, (completion time after initial recognition), (3) cognitive load based on eye gaze and blink patterns observed and NASA Task Load Index (NASA TLX) survey, (4) NSST performance comparison to baseline after VIBRaNT exposure, (5) post-event surveys (see attachments) and, (6) qualitative oral feedback of the training experience in VIBRaNT compared to NSST. These metrics will be statistically analyzed to compare facilitation of bridge training between the VIBRaNT XR system, (experimental condition) and the NSST physical bridge mockup simulator (control condition).

Successful development of VIBRaNT has the potential to improve bridge watch standing by increasing training availability, ultimately leading to improved fleet readiness. Additionally, because the VIBRaNT training platform is affordable, small, software based, and portable it, (1) has a tremendous potential to scale beyond fixed cite training facilities, (2) allows teams to practice from dispersed and remote locations, at odd hours, in a single 'virtual' bridge environment, (3) can be easily customized to represent unique bridge configurations, (4) can mimic unique bridge sound environments and, (5) enables over-the-air software updates to provide different bridge and port models for navigation rehearsal at costs far lower than a physical simulator. Not only will this system aid in bridge watch standing, it will also add to scientific knowledge by objectively demonstrating the utility of XR to enhance knowledge flow in situations requiring spatial-temporal skills and understanding, specifically in cases of 1st person embodied activities where recognition of delayed feedback cues, along with timely management of other co-occurring bridge activities, is critical to safe operation of the ship.

4.Sponsor and Funding:

Sponsorship and funding for VIBRaNT has been provided by the Office of Naval Research (ONR) Tech Solutions for $1,018,000. Major Bid and Proposal internal NISE funding of 62K assisted in creating the 3D audio scape for VIBRaNT.

5. Objectives, Research Plan, and Methodology:

A. Objectives:

i. Use of Data:

Data collection and use will be performed by VIBRaNT personnel (Amelia Kracinovich (71740) as Principal Investigator (PI) and other scientists listed as Associate Investigators to: (1) Iterate on system design based on comments from novice and expert ship-handlers asked to test VIBRaNT. System design modification data will be collected in the form of guided and unstructured feedback, (2) Develop scenario and evaluation criteria from data gathered in interviews with

Navy training SMEs and, (3) Evaluate VIBRaNT's ability to facilitate training in comparison to the NSST mockup simulator to determine viability for use in Navy training. This data will also be collected in part of Jon Brewster's dissertation for NPS, no additional data will need to be collected for this effort.

VIBRaNT system design phase will illicit feedback from Navy Officers, Enlisted Personnel, and Navy Scientists to prepare for the evaluation phase of the research. It is requested that the IRB consider this phase as not HSR as the information being collected is not focused on the individual but rather the system design. No specific information able to be tied back to the individual is being collected, and the only information that would tie info back to an individual would be signature on an informed consent document. Because of this, we are requesting this phase of study to not be considered HSR and the IRB to approve waiving informed consent requirement for this phase of the study.

VIBRaNT system evaluation will observe shiphandling and seamanship task management while maintaining overall situational awareness during training scenarios presented in the NSST and VIBRaNT training platforms. Because this will be analyzed statistically and PII will be retained. Data taken from this phase will include participant's responses from both NSST and VIBRaNT performance for comparison of (1) participant's reaction time, (2) participant's time to task completion, (3) eye gaze patterns, (4) subjective survey's, e.g., cognitive load and, (4) subjective experience.

The evaluation phase will leverage normally scheduled Navy NSST training events at the Mariner Skills Training Center Pacific (MSTCPAC) as the control condition, such that VIBRaNT investigators need only observe regularly scheduled NSST training and be ready to run bridge crews (assigned for NSST training) through the experimental VIBRaNT XR condition either before or after the NSST event has completed. VIBRaNT XR observation and data collection is planned to require no more than 70 minutes over what is normally mandated for ships company in meeting NSST training requirements. In this within subjects design, the same bridge crew will participate in both the NSST and VIBRaNT conditions on the same day, in the same location with minimal impact to the crew's regular schedule. Performance data taken from training events in both the NSST and VIBRaNT platforms will be controlled for a consistent set of baseline factors: presented environmental conditions (e.g., wind and current), ship physics (e.g., rate of turn, speed over ground, engine/rudder state), density of contacts (e.g., air, surface, subsurface), constraining maritime passages, (e.g., San Diego Bay) crisis points, (e.g., engine failure) numbers of other trainees and instructors interacting in the training, (e.g., in NSST and VIBRaNT), etc. 'Within Subjects' data collection will be de-identified using assigned participant identification codes (completely anonymized) and counter-balanced for a statistical analysis of variance comparison of the same scenarios and stimulation performed in the NSST versus VIBRaNT. The VIBRaNT team will seek to eliminate instrumentation and other bias by closely controlling potential confounding variables. Experimental design, variable operationalization, and resultant data are intended to either support or refute the research hypothesis (H_a) that the VIBRaNT training platform is equivalent or better to the control condition of the NSST physical bridge mockup (H_o) in facilitating shiphandling and bridge resource management knowledge acquisition and transfer.

(NOTE: All data analysis will aggregate measures across individual performers without any Personally Identifying Information (PII) links. At no time will this information be paired to participants, see paragraph 13 Private/Sensitive Data for details.)

ii. Importance for the Navy:

Recent developments in XR IT are presenting unprecedented opportunities for the Navy to advance its personnel training capabilities and beyond. XR provides a wide range of opportunities to increase the availability and effectiveness of naval training and operations, particularly in situations needing knowledge transfer of 1st Person, special-temporal knowledge such as shiphandling and bridge resource management. Given the high cost of navigation and shiphandling related accidents, the need for a readily available, cost-effective training platform is critical. The only opportunity for shiphandling and bridge management training ashore is currently in fixed and logistically limited facilities, e.g., NSSTs. If successful, VIBRaNT will allow training anywhere in the world, any time it's required. This could prove advantageous when difficult evolutions are planned to take place away from fleet concentration areas, e.g., on deployment, and when refresher training is needed to hone bridge watch standing skills. Given recent developments in the XR technology market, introductions of Meta Quest 3, Apple Vision Pro, and the Alliance for Open Universal Scene Description (OUSD), indications are that XR IT will continue to advance while decreasing in price. This provides an opportunity for the Navy to provide '1st Person' spatial-temporal training in many different contexts to enable 'embodied' knowledge acquisition, bridge training being but one of several possibilities. If VIBRaNT succeeds, a cost benefit analysis could be performed to determine total lifetime costs of VIBRaNT compared to legacy trainers and costs to deploy the capability across the fleet.

Future longitudinal studies may also show clear benefits from training accessibility of a VIBRaNT like capability where watchstanders repeat several 'sets and reps' over time and gain far great ability to recognize and respond to emergent situations in complex operational/combat environments.

In addition, VIBRaNT research opens opportunities to employ XR in areas well beyond knowledge transfer. The possibility of employing AR, MR, AV and VR in other Navy situations, e.g., overlaying critical combat information such as track status, equipment conditions, weapons warning and status changes, changing threat sectors, latest indications and warning, commander's intent, etc., is limitless. In addition, overlaying information into the persistent field of view of an officer standing watch, in the ship's wardroom, off duty in a stateroom, or off the ship in a fleet command center holds potential to accelerate decision speed. VIBRaNT could even be used as a vehicle to introduce new XR IT, providing a virtual test platform to eliminate risk of testing novel XR solutions at sea. This kind of test platform could also allow SMEs to join capability evaluation from remote locations. Potential applications are limitless.

iii. Expected outcomes:

Study outcomes will include qualitative and quantitative data from human subject participants (novice to experts) in a range of shiphandling evolutions. Outcomes include, (1) generate design and performance measures to clarify XR application in training bridge teams. (2) inform effectiveness and application of XR in training shiphandling, seamanship, and bridge resource management. This data will also benefit future design patents and/or scientific publications summarizing the work. Anonymized results will be summarized in reports to ONR and be submitted for publication to the Defense Technical Information Center (DTIC). Suggestions for further development and potential fleet transition opportunities derived from this research will also be included in these reports. The raw data will only be retained by NIWC for a period not to

exceed 5 years. General guidelines and potential future development will be based on analysis available in DTIC.

B. Research Plan:

i. Participants:

Participants will include a range of civilian and military personnel, novice ship-handlers, master mariners, qualified USN OODs, senior SWOs, and schoolhouse instructors. Participants will be needed for (1) the design phase and (2) the evaluation phase. The design phase will be broken into two 'sprints' to prepare the VIBRaNT XR system for the evaluation phase. The design phase participants will include civilian and military personnel from NIWC PAC. In addition, civilian and military personnel will also come from the Mariner Skills Training Center Pacific (MSTC PAC) in San Diego CA, with backup participants coming from MSTC Atlantic (MSTC LANT) in Norfolk VA and Naval Postgraduate School (NPS) in Monterey CA. An IRB is not being sought for this phase of the project as no intervention will be applied and no PII will be collected or retained. Participants (for the design phase) will be recruited by the investigators based on already held relationships with SMEs, Program Offices, TYCOMs, and NPS professors where participation will be recruited on completely voluntary, and ad hoc basis.

For the evaluation phase, participants will be recruited mainly from bridge crews reporting to MSTCPAC for NSST training. The investigators will coordinate with MSTCPAC schoolhouse to gain volunteers based on availability and operational tempo of those scheduled to participate in NSST training. All bridge crew participants will have the basic training and qualifications required to operate a ship from the position they are normally assigned to perform their duties from, e.g., Officer of the Deck (OOD), Junior Officer of the Deck (JOOD), Conning Officer (CONN), Navigator (NAV), Helmsman (HELM), Lee Helmsman (LEE HELM), Boatswain's Mate of the Watch (BMOW) or Quartermaster of the Watch (QMOW) and will be observed as a group along with their assigned instructor. This 'typical bridge crew' (typically comprised of the nine individuals listed above) represents the sample configuration for the evaluation phase of the effort. A Letter of Support (LoS) from Commander, Naval Surface Forces, U.S. Pacific Fleet (SURFPAC) has been obtained and is included as Attachment L. SURFPAC has authority over MSTC PAC. This protocol is being carried out with co-investigators from NPS and will have an IAIR in place for them to participate. The IAIR once in place will give authorization for recruitment at NPS. If the PI is required to recruit from other military commands, additional LoS's will be added to the protocol through an amendment request. All participation will be voluntary, and participants will be informed they can withdraw from the study at any time without penalty or negative repercussions. Participants don't need to participate in both phases (design and evaluation) of the research to be included, Participants in the evaluation phase will be asked to indicate in their pre-screening demographics questionnaire if they participated in the design phase. As noted, the evaluation phase of the research, to include the study, data collection and recruitment location planned is MSTC PAC, with backup locations to include MSTC LANT and Navy Surface Warfare Schools Command (SWSC) depending on logistics and sample needs. Should these alternate locations be required, an amendment shall be submitted with an authorization from competent authority to test at the backup location.

1) Participants in VIBRaNT System Design:

The PI and Co-PI will recruit civilian personnel for design feedback as needed from NIWC PAC, Naval Postgraduate School (NPS), MSTC PAC, MSTC LANT and other commands as available. Recruitment will be based on connections to known NIWC SME's, Naval Officers, NPS Professors, and Program Officers who have expertise in the field. There is no extended exposure to XR or risk to these recruits during this phase. Civilians (aside from civilian instructors and VIBRaNT Research Team) will only be needed during the design phase and will be recruited from Naval Postgraduate School or NIWC Pacific via PI and Co-PI as noted above.

For design feedback sessions, a statistical power analysis of needed participants is not appropriate as the measures are qualitative. Design feedback will be divided into two 'sprints', and participants with a diverse range of expertise will be sought to review software builds and provide feedback, particularly in sprint 1. For modifying the VASCO COTS system into VIBRaNT we will need feedback from each role a watchstander assumes on the ship bridge. These roles include: the OOD, JOOD, NAV, CONN, Helmsman, Lee Helmsman, QMOW, and BMOW. Investigators will also obtain feedback from schoolhouse instructors. An optimal number of 50 participants will be desired, with a greater range not to exceed 100 participants, accounting for dropouts, greater than planned recruitment response, etc. 50 participants is the optimal number as eight positions occupy the ship bridge meaning feedback from each position will be received about 10 times. Since there are two ports and we are likely to only have enough time with each participant to review one port, each position in one port is likely to be viewed five unique times. This is in accordance with research from the User-Centered Design and Engineering Branch (UCD&E, Code 53621) suggest that 4- 6 participants will suffice.

For sprint 2, participants with similar backgrounds or expertise will advise. Participants in sprint 1 will also be able to participate in sprint 2. We will ask participants in sprint 2 to identify if they were also in sprint 1. SMEs will receive flyers by the schoolhouses. The schoolhouses or the participants will contact our team if ship teams wish to participate. Only the researchers on this protocol will be actively interacting with participants. We will aim to recruit a similar number of participants as in sprint 1 for the reasons listed above involving unique responses from each bridge position and port models. Roundtable feedback sessions will be planned, and ad hoc feedback sessions will also occur on a one on one or small group basis, with a range of 4 to 20 SME participants reviewing the holistic application of the VIBRaNT training platform in each session. It is anticipated that about 50 participants will be needed, with the same caveat above for drops, recruiting response, etc., not to exceed 100 participants.

It is requested that the IRB consider this phase as not HSR as the information being collected is not focused on the individual but rather the system design. No specific information able to be tied back to the individual is being collected, and the only information that would tie the info back to the individual would be signature on an informed consent document. Because of this, we are requesting this phase of study to not be considered HSR and the IRB to approve waiving informed consent requirement for this phase of the study.

2) Participants in VIBRaNT System Evaluation:

The PI and Co-PI will recruit military personnel from MSTCPAC in San Diego CA, with backup participants coming from MSTCLANT in Norfolk VA and NPS in Monterey CA. A Letter of Approval from SURFPAC is staffed and included as Attachment L. (NOTE: If the PI is required to recruit from other military commands, e.g. MSCTLANT an IRB amendment request to add

these will be submitted for approval.) PI and Co-PI recruiting of bridge team participants shall be managed in concert with the MSTCPAC Scheduler and Instructors, the Surface Warfare Schools Command representative, and Ship's Commanding Officer for a participation period not to exceed 70 minutes in the experimental condition (VIBRaNT). Coordination of participants will be routed through administrative assistants at the schoolhouses who are solely responsible for bridge teams coming to the trainers and do not have influence over their outcome in the schoolhouses. In addition to Bridge Team Watchstander recruiting, the PI will also recruit MSTCPAC Instructors to participate and observe during the VIBRaNT observation sessions. Email (see Attachment A) requesting participation will be coordinated by the PI/Co-PI in advance to ships and individuals who have up-coming training events scheduled with the NSST facility. These requests will be sent by the NSST Staff to ensure not only students, but also MSCTPAC instructors are available for XR VIBRaNT and NSST observations. The staff coordinating the availability of participants only take on an administrative role and have no authority over the participants. Since this is a within subjects' design and this testing does not interfere with the normal schedule of the NSST, the participants recruited will be scheduled through the schoolhouse and their willingness to participate will be relayed through leadership to our team. Investigators will only go to scheduled NSST training when it is previously determined ahead of time that the bridge team is willing to participate in the experimental condition. The Informed Consent Document (ICD), Recruitment email, and PI/Co-PI VIBRaNT indoctrination script also relates all participants can withdraw from the VIBRaNT evaluation study at any time without penalty or negative repercussion.

In addition, in-person recruitment is also possible based on timing, equipment set up, and other activity of the PI/Co-PI at MSTCPAC and other locations (MSTCLANT), but any group/ individual recruited for the evaluation of VIBRaNT will be pre-screened for XR and provided the same Privacy Act and ICD as those recruited via email (as noted above). All of the pre-screening questionnaires will be completed before conducting VIBRaNT observations/evaluations. No pre-existing data or materials will be requested.

Participants in the evaluation phase will first complete the ICD, privacy statement and pre-screening questionnaires and be asked to participate in both the control and experimental condition. Before starting the control condition (NSST), the OOD participant will have the eye tracking glasses calibrated to them. The control condition will be participants being observed during their routine (previously scheduled) training at the NSST which isn't controlled in any way by the VIBRaNT investigators. NSST training is a normally required event for USN bridge crews to maintain their qualifications and the only data recorded by the VIBRaNT team during the NSST observation period is eye tracking of Officer of the Deck (OOD) done passively and concurrently with participation in NSST and post-event surveys provided to the bridge crew after the NSST event is over. VIBRaNT researchers will come in after hours to watch video captures by the NSSTs to determine reaction times to key events. (Note: NSST training is managed by MSTCPAC and is a normally required event for USN bridge crews to maintain their qualifications.)

The VIBRaNT investigator team will be leveraging NSST training events on a 'Not to Interfere' basis with regularly scheduled MSTCPAC bridge training sessions. If a participant chooses not to participate in the VIBRaNT study, it has no impact on his/her participation in the Navy mandatory MSTCPAC scheduled training session as it is a regularly scheduled event. The

VIBRaNT one-hour session will take place either immediately before or immediately after the NSST scheduled event where, as stated in the informed consent form, each recruit has no obligation to participate in the VIBRaNT study and can withdraw at any time without negative repercussions. Should the bridge team being observed lose a member for any reason, the investigator may continue the observation planned with the instructor taking the team members' place and note differences in the statistical analysis. If a second participant needs to drop then one of our team members, Jon Brewster, a qualified SWO, will take their place. If more than two participants need to drop, then this group will need to be excluded.

For the evaluation phase, it is hypothesized VIBRaNT will be greater than or equal to the NSST in facilitating shiphandling and bridge resource management training. Therefore, expected mean performance of experimental condition (VIBRaNT) will not differ and may, for certain mapped skill sets, exceed training facilitation performance found in the control condition (NSST). Assuming a statistical value of alpha=0.05 and a power of (1-β) = 0.80, A maximum of 30 bridge teams (USN, eight members per team and one instructor), will be recruited for participation in both the experimental condition (VIBRaNT) and the control condition (NSST). Recruits will be chosen from ships company bridge watch teams scheduled for regular training taking place at MSTCPAC. A sample size of n=243 participants would be optimal for the evaluation. (243 participants were subjected to both the control and experimental conditions). The target sample is purposive of junior grade Naval Officers, ranking from O1 to O4 and schoolhouse instructors. This will be a convenience sample, depending on ships schedules, available trainees for NSST training, etc. This dependency may confound obtaining the target n=243 for the study. The contingency plan is to travel to other trainers outside MSTCPAC, e.g., MSTCLANT, to reach a desired population of 243participants performing in the desired experimental condition. Should Navy logistics prevent the optimal participation of n=243, the available number of participants approaching an optimal n=243 will be statistically analyzed to determine the study 'power' and ability to generalize based on participant availability. Each bridge team will be counted as one data observation point in our statistical analysis, meaning we need 27 bridge teams. If, after statistical analysis has been performed, the available n still proves insufficient, a 'Between-Subjects' experimental design will be utilized to broaden the participation base to include participants from other locations, e.g., NPS. Throughout the evaluation phase, the research team will leverage a broad range of statistical tests, to include non-parametric tests, to assess if required numbers of participants have provided reliable results from the data obtained. If another population is required, or a modification in the experimental design be desired, an amendment with a CO Letter of Support will be submitted to the IRB committee.

There will be targeted calls for participants for the evaluation phase, see Attachment A "Participant Recruitment Email." Participants will mainly be recruited through the NSST school house; however, participants will be pre-screened to ensure they meet the inclusion/ exclusion requirements. It is not required that participants volunteer for both design and evaluation phases, nor is participation in one phase required, or exclusionary, from another.

ii. Inclusion/Exclusion Criteria:

Included participants for this study must be 18 years of age or older and United States citizens but will be excluded from and system effectiveness evaluation sessions if they have a prior history of motion sickness, vertigo, epilepsy, or similar condition. This is because participants may be asked to wear an HMD (e.g. Meta Quest 3) which has the potential to cause motion sickness. For the

effectiveness evaluation sessions, included participants must meet an additional requirement of having experience navigating/conning large ships, e.g., Navy combatants and have the basic training and qualifications required to operate a ship from the position they are normally assigned to perform their duties from. This information is also referenced in attachments.

iii. Research Design:

The research study design will consist broadly of two phases, (1) a system design feedback phase where comments are solicited on the modified COTS software (VIBRaNT), (2) a system evaluation phase, where the modified COTS (VIBRaNT) is experimentally observed and analyzed for effectiveness.

1) VIBRaNT System Design

The design phase will be subdivided into two sprints. Sprint 1 will focus on developmental features in isolation from each other, e.g., the VR model of a USN DDG-51 class bridge, including the imbedded displays inside the virtual bridge, e.g., Multi-Function Control Repeaters (MFCRs), Voyage Management System (VMS), RARAR displays, helm controls, etc., the virtual model of the port of San Diego and Norfolk. The first sprint will involve sending screenshots, e.g.,.ppt files, pictures, etc., of specific imagery to the vendor, Office of Naval Research, NIWC PAC XR SMEs, and Naval Officers across the US.

Sprint 2 will iterate on overall system performance in the context of a training scenario, e.g., entering the virtual port of San Diego, as perceived via a DDG-51 class destroyer virtual bridge environment, during a pre-planned training scenario. This sprint will ensure VIBRaNT features, e.g., San Diego and Norfolk port renderings, the DDG Bridge Model with all its imbedded displays, the Graphical User Interface (GUI) for instructor input, scenario design events in the Controller PC, etc., are ready for evaluation. However, there will be no physical data collected from or about the individuals participating in this sprint. Information gathered will strictly focus on user interface input to ensure the system is ready for the evaluation phase of the study.

Each design sprint will involve Naval Officers periodically wearing a VIBRaNT HMD, with instructors using the VIBRaNT laptop computer to provide feedback on the scenario design GUI and ability to act as a virtual Helmsman and Lees Helmsman.

As new developmental software features are released, feedback will be elicited from instructor SMEs, Naval Officers, and qualified SWOs. Sprint feedback for each software developmental element will be consolidated and written as action items to the COTS vendor, Kilo-Solutions, to improve their design.

Sprint 2 feedback will end when all design elements are complete and integrated. The research team will finalize the design phase by asking instructors and qualified SWOs to speak aloud of their thought process while running VIBRaNT training scenarios and practicing event injects. This will be conducted in-person.

Participants in the design phase are not required to participate and will not suffer penalty if they withdraw from Sprint 1 or 2. These sprints are for the purpose of commentary only, are part of the design phase and none of these individuals are planned to take part in the sampled population for the evaluation phase of the VIBRaNT study at MSTCPAC. In the rare event that they do take part

in the evaluation phase, all pre-screening, ICD and other procedures will be followed as noted for the sampled population.

2) VIBRaNT Effectiveness Evaluation

The effectiveness evaluation will begin with pilot testing to verify data analysis operationalization readiness before collecting experimental data. Pilot testing will involve a bridge team who is not participating in the evaluation. All participants (in each of their bridge team assigned roles, OOD, JOOD, CONN, etc., will be assigned a participant ID code, this number will be used on all data collected. PII linking to the participant ID will NOT be collected and any related potentially identifying data, e.g., ship assignment, will be stored in an external hard drive and locked in a secure cabinet.

Data collection for both platforms is planned to take place at MSTCPAC, with MSTCLANT and NPS as an alternative site if the optimal number n=243 27 participant groups (8 personnel per bridge team group and one instructor) cannot be met at MSTCPAC. In addition to housing the NSSTs, MSTCPAC also has large classrooms that can be used for collecting VIBRaNT data from participants using VIBRaNT in a classroom setting. VIBRaNT effectiveness evaluations are anticipated to last no more than one-hour and be performed either just before or just after the regularly scheduled NSST training event.

a) VIBRaNT Pilot Testing

The research team plans to use one bridge team/scenario (either San Diego or Norfolk) for pilot testing to ensure desired data can be collected and analyzed according to VIBRaNT's effectiveness metrics. Video and audio recordings are typically captured during NSST training evolutions; the research team will leverage these recordings i during the pilot study, and evaluation phase to compare time stamps on the instructor screen and note when key events occur in the scenario. Video data recording is a normal MSTCPAC function of NSST Instructors, VIBRaNT team members may view this data, but it will not be recorded or retained by the VIBRaNT team. These observations will inform when to measure 'reaction time' and 'time to task completion' to minimize instrumentation bias and enable a performance comparison based on platform presentation (VIBRaNT or NSST). If changes to the test procedures for video capture is needed, it will only be done after IRB approval as an amendment.

Voice printing or other audio identifying markers will not be used. All subject audio observations will be identified by anonymous codes where audio study responses in XR will be limited to associated behavioral actions observed in each condition, Mockup or XR for comparison purposes only.

All information for both pilot and effectiveness evaluation will be 'de-identified' at the point of observation using a participant number for each participant applied to both survey data and eye tracking data. Data analyzed and reported will not have any kind of identification markings. Any video of subject faces will not be retained by this research effort. Video being observed is a normal function of the NSST training system already in place at MSTC PAC and is subject to Navy operational training regulations. Investigators will view these videos in house at the NSSTs and not retain any NSST recordings at any time. De-identified data will be used only to map study scenario activities to equivalent NSST training events. Recordings of VIBRaNT interaction will not have face capture, and only anonymous aviators with participant number in place of

participant name. In addition, eye tracking data of the Officer of the Deck will be collected but only identified by code at the time of recording. These recordings will be de-identified, and any recorded media will be encrypted on an external hard drive and securely stored in a locked cabinet- only anonymized results will be presented to external parties, to include the sponsor.

b) VIBRaNT Effectiveness Observations

VIBRaNT XR set up and observations will be made on a not to interfere basis with MSTCPAC training. The research team will arrange observations based on MSTCPAC schedule and travel to NSST when evaluation events are approved. Training participants and scenarios are normally determined by MSTCPAC two weeks prior to training, and the research team will schedule a MSTCPAC classroom for VIBRaNT XR set up and observation periods as directed by MSTCPAC staff SMEs. Bridge teams of the same size performing in the control condition (NSST) will train in VIBRaNT. VIBRaNT HMDs will communicate to each other via a standalone Control PC and wireless LAN.

Experimental condition

For the experimental condition (VIBRaNT XR), the breakdown of activities will include approximately 5 minutes for ICD, 5 minutes of pre-screening, 10 minutes of introductory VIBRaNT training/ safety, 30 minutes of activity in VIBRaNT, 10 minutes of completing the survey forms after finishing their VIBRaNT session, and 10 minutes for survey after NSST for a total of approximately 70minutes.

Upon arrival, recruits will be provided with the Privacy Statement, see Attachment B, an Informed Consent Document, see Attachment C, a Pre-Screening Questionnaire, see Attachment D, and a Simulator Sickness Questionnaire, see Attachment E.

The PI and Co-PI will review the questionnaire responses and if the recruit is not screened out by the PI based on the questionnaire answers, he/she will be assembled in the VIBRaNT testing classroom for introductory training with the VIBRaNT HMD and Controllers expected to last no more than 10 minutes. Based on the initial exposure to XR, the PI will check to ensure the recruit is Ok to proceed, if they are, the PI will admit them into the effectiveness evaluation period.

The 30-minute evaluation period will include a virtual reality presentation of a DDG-51 Bridge where each watchstander performs training via an HMD presentation of scenario events taking place in the vicinity of either San Diego Bay or Norfolk Naval Station. The scenario events presented will be matched to scenarios normally provided by the NSST to train bridge teams how to respond in a variety of emergent situations. These simulated scenarios will all take place in Meta Quest 3, all participants will wear this headset minus the instructor who will sit behind the laptop computer to monitor the planner mode. Researchers will ask participants if they need help being fitted for VR HMD. The headsets, laptop and all equipment will be disinfected immediately after use.

During the 30-minute period, the Officer Of the Deck (OOD) Watchstander will wear a Meta Quest Pro HMD which will track and record his/her eye gaze and blink rate. (This is included in the Informed Consent Form) for analysis. The study team will have reviewed the ICD as part of the pre-screen evaluation period. If the OOD chooses to not participate in the eye tracking portion

of the study, then the research team will not turn-on eye tracking capability in the Meta Quest Pro Headset. All data collected will be anonymized via a participant identifier code.

When the 30-minute VIBRaNT observation period is over, each recruit shall complete the following surveys as appropriate: NASA Task Load Index (TLX) Workload Questionnaire, see Attachment F, System Usability Scale, see Attachment G, Bridge Watch Team Self-Assessment Rubric, see Attachment H, Bridge Watch Team Instructor Assessment Rubric, see Attachment I, OOD Instructor Assessment, see Attachment J, and Exit Interview Questions, see Attachment K.

Control Condition

Before any experimentation takes place, participants first complete the ICD, privacy statement and pre-screening questionnaires. Upon completing or (in the counter-balanced design) before beginning VIBRaNT XR experimental observations, recruits will conduct their normally scheduled bridge training in the NSST. 5 minutes for ICD, 5 minutes of pre-screening, 10 minutes of introductory VIBRaNT training/ safety, 30 minutes of activity in VIBRaNT, 10 minutes of completing the survey forms after finishing their VIBRaNT session, and 10 minutes for survey after NSST for a total of approximately 70 minutes outside of their typical training time. If a participant decides they do not want to be in the data set before beginning the scheduled training, then none of the video recording will be analyzed, only the surveys will be issued to the participants willing to be included in the study. The instructor will be asked to fill the role of any students who don't wish to participate, and our team member Jon Brewster will fill in if a second participant does not wish to participate. If more than two participants drop from the study during data collection, then the session will end, and their data will be excluded. The only difference to a regular NSST event is as follows:

NSST Students will complete the same post-event forms as completed for VIBRaNT to include NASA Task Load Index (TLX) Workload Questionnaire, see Attachment F, System Usability Scale, see Attachment G, Bridge Watch Team Self-Assessment Rubric, see Attachment H, Bridge Watch Team Instructor Assessment Rubric, see Attachment I, OOD Instructor Assessment, see Attachment J.

During the 30-minute period where the training scenario overlaps with that presented in the VIBRaNT condition, the OOD Watchstander will wear non-corrective, non-interfering eye tracking glasses and his/her eye gaze and blink rate will be tracked and recorded. (This is included in the Informed Consent Form) for analysis. The OOD will have the option to opt out of eye tracking if desired and will not have their eyes calibrated or data collected for eye tracking. Researchers will ask participant if they need assistance putting on eye tracking device, this is very light weight and fits like a normal eyeglass would. The eye tracker will be disinfected immediately after use. All data collected will be anonymized via an assigned participant identifier code and any PII will be on an encrypted hard drive locked in a cabinet at NIWC PAC.

3) Treatment (experimental) conditions:

Treatment conditions apply only to the evaluation phase of the study. VIBRaNT is considered the treatment condition, with participants (OODs and bridge teams) being 'treated' (experimental condition) as compared to the legacy 'treatment' of the NSST training platform (control condition).

4) Measures:

a) VIBRaNT System Design

While there are no 'measures' per se for the VIBRaNT design phase, qualitative feedback will be obtained. Sprint 1 will include open 'speak aloud' feedback and Likert scale survey rubrics used during the evaluation phase. For Sprint 2, qualitative feedback will be like Sprint 1 but also include overall system performance and scenario design feedback from MSTCPAC NSST SMEs and instructors.

b) VIBRaNT Effectiveness Evaluation

Measures for the VIBRaNT effectiveness evaluation will consist of (1) event reaction time, (2) task completion time, (3) rubric surveys, (4) eye gaze pattern data, (5) NASA TLX survey, (6) system usability survey and, (7) subjective experience exit interview. Quantitative measures will leverage the qualitative system design feedback results to determine event injects and when best to observe and record performer reactions during the training scenario(s). Eye gaze data collection will be performed during the control condition (NSST) as well as the experimental condition (VIBRaNT XR) to compare gaze vector patterns and blink rate to events injected during both the control and experimental conditions. The same scenario event injects will occur across both the control and experimental conditions.

5) Hypothesis:

a) VIBRaNT System Design

Hypothesis testing is not applicable to the VIBRaNT design qualitative feedback sessions.

b) VIBRaNT Effectiveness Evaluation

The hypotheses to be tested is during the VIBRaNT effectiveness evaluation include:

H_1–Knowledge Acquisition and Transfer facilitated via VIBRaNT will (for mapped skills/tasks) be greater than or equal to the facilitation provided via NSST.

H_2–Extraneous Cognitive Load produced by VIBRaNT will (for mapped skills/tasks) be less than or equal to the extraneous cognitive load produced by NSST.

H_3–Facilitation of critical Situational Awareness (SA) skills via VIBRaNT will greater than or equal to SA skills facilitation via NSST.

H_4–The Knowledge acquired via the VIBRaNT will, when preceding NSST training sessions enhance performance observed in the NSST.

Hypotheses listed represent H_a for the experimental condition (VIBRaNT) with comparison to H_o of the control condition (NSST).

C. Methodology

i. Task Description:

1) VIBRaNT System Design:

Design feedback will be collected in guided and unstructured sessions during both the 1st and 2nd design sprints. Participant feedback collected in the 2nd sprint will involve 'talking aloud' methodology where the research team will ask individual training SME's and bridge trainees to explain their actions and thought processes while going through the VIBRaNT facilitated training scenarios. The same methodology will be utilized with instructors while they set up the training scenario and watch the trainees progress and interaction with VIBRaNT throughout the final design feedback sessions with 'open-ended' questions being asked of trainees and instructors.

2) VIBRaNT System Effectiveness:

VIBRaNT system effectiveness testing will be conducted concurrently with MSTCPAC, (or MSTCLANT) scheduled NSST training events in San Diego (or Norfolk) with DDG-51 bridge crews. The research team will attend these sessions looking to enable a 'within subjects' experimental design, to include counterbalancing VIBRaNT and NSST training sessions such that bridge crew trainees will either train first with VIBRaNT followed by NSST or begin with NSST followed by VIBRaNT. The counterbalance design accounts for maturation/learning bias based on the sequence of training platform exposure to similar training scenarios. Regardless of which condition the bridge team begins with, informed consent and pre-screening will be reviewed and signed for both the control and experimental condition before beginning experimentation. If a subject decides to withdrawal during the data collection or after data collection is complete, their individual data will be destroyed from the combine data set and analysis of that data discontinued. If a participant withdraws during the experimental condition, the instructor will virtually fill that role to finish experimental data collection. and our team member Jon Brewster will fill in if a second participant does not wish to participate. If more than two participants drop from the study during data collection, then the session will end, and their data will be excluded.

Each role on the bridge team is defined by the training schoolhouse based on the position they fill during actual bridge operation; hence they will hold the same position in the control and experimental condition. The remainder of the data from the bridge team will be retained and analyzed. At the conclusion of training events, surveys will be issued consisting of Likert scales and training rubrics specific to performance criteria of team and individual trainee performance aligned to requirements for each member of the bridge team (Attachment H), the OOD (Attachment J) and instructors (Attachment I). Instructors and trainees will fill out a VIBRaNT System Usability Survey (SUS) (Attachment G). Trainees will also be asked to fill out a NASA Task Load Index (TLX), workload questionnaire (Attachment F), and a simulator sickness questionnaire will be issued at the end of the testing condition (Attachment E). During NSST simulator downtime, the VIBRaNT team will review video and audio capture of bridge team interactions and compare them to instructor scenario views to determine 'reaction time' and 'task completion time' for key events presented during each scenario. These control scenarios will bound the development of the VIBRaNT experimental testing scenarios.

Eye gaze tracking will also be employed in each NSST and VIBRaNT training session. In the case of the experimental condition (VIBRaNT XR) gaze vector data will be captured from the OOD watchstander via a Meta Quest Pro (in VIBRaNT XR) and compared to the eye gaze data captured via Pupil Labs' eye tracking equipment (in NSST). This data will be analyzed and compared for blink rate, gaze vector change rate, gaze location and duration rates, time to/within desired Area of Interest (AOI) zones. Comparison will be limited to the OOD watchstander for ease of comparison.

The experimental VIBRaNT testing scenarios will leverage, as closely as possible, the same metrics derived from the NSST control condition. All scales and surveys will be completed at the conclusion of the training. The measurement of reaction time and task completion time will be done on the self-contained video replay of the VIBRaNT system. It is anticipated that each VIBRaNT training session will not exceed 70 minutes of testing.

ii. Apparatus/Tools/Materials:

As discussed, the apparatus, tools, and materials of the experimental phase of this study will consist primarily of IT, specifically the COTS VASCO system militarized into a USN version referred to as VIBRaNT.

Meta Platforms Meta Quest 3 Head Mounted Display, a scenario and controller PC, and a wireless router for bringing up to 12 individuals into the same virtual environment. The Meta Quest 3 headsets will be configured with 512GB of storage, 2 color cameras and 3D spatial audio. A Meta Quest Pro and Pupil Labs Eye tracker will be used to collect eye tracking data, where the Meta Quest Pro provides essentially the same presentation fidelity as the Meta Quest 3. In addition, the VIBRaNT system uses the Controller PC to provide instructors the ability to design and configure scenario parameters for training and to control scenario input stimuli during runtime. The control PC will also show eye gaze heat map overlays to aid in determining AOI look-up and gaze time. The instructor-based hardware consists of a VIBRaNT laptop PC hardware and software, 3D models of San Diego, Norfolk, and the DDG-51 bridge configuration loaded into the system. Meta Quest 3/Meta Quest Pro headsets communicate to the controller PC via an 'off network' Local Area Network (LAN) using a VASCO network router. The specifications of the instructor PC Controller hardware suite are:

Laptop–HP Z Book Firefly 14 Core i7-1165G7 32GB RAM, 512GB SSD 14IN FHD Win 11 Pro–NVIDIA Quadro T500 (4GB Graphics)

Router NETGEAR RAX20 4-Stream AX1800 Wifi 6 Router

1x HP EliteDesk 800G4 Mini i5-Gen8 8GB 256GB SSD Win 10 Pro

A Pelican Case with foam inserts that permit safe transportation of all equipment will be provided as part of the VIBRaNT system.

At the start of this protocol, Meta Quest 3 and Meta Quest Pro HMDs will be used. These HMDs are full VR headsets that block out the outside world and present an entirely computer-generated environment into the participant's visual field. They also have High Definition color-pass-through modes that allow the operator to 'see' the actual physical environment where the testing is being performed. (This 'pass-through' mode is an important safety feature that allows the participant to quickly and easily 'exit' the virtual environment without needing to remove the HMD.) The Meta Quest Pro enables eye tracking via onboard cameras which are completely passive, have no impact on user experience, allows full color 'pass-through', and presents scenarios with nearly identical levels of visual fidelity as the Meta Quest 3. The subjects selected for eye tracking will be OOD's, who will be notified and sign a release form for collecting their gaze vector data. All gaze vector data will be anonymized such that the only identification made will be to a particular watch standing position, e.g., OOD Subject Nr. 1.

During the control condition (NSST), the OOD's will wear Pupil Labs Eye Tracking equipment, essentially a pair of glasses relatively unnoticeable to the measured individual, that will passively track eye gaze and blink rate while performing training in the NSST. This eye tracking data will be collected and analyzed in comparison to the VIBRaNT Meta Quest Pro eye gaze data. Meta Quest 3's, Meta Quest Pro's, and Pupil Labs equipment will be used in accordance with manufacturer's specifications and fact sheets.

Unity 3D VR enabling software will provide the virtual environment/synthetic experience needed to support at sea navigation environment. This will include: (1) simulation of current, wind, and other environmental factors, (2) simulation of own ship hydrodynamic forces, including engine and rudder, (3) simulation of other ships, including tugs, logistics ships for underway replenishment, and other private, commercial, and military ships in the nearby environment, (4) models of common ports, to include San Diego and Norfolk port approaches, (5) a DDG-51 class bridge configuration to include it's control systems, control interfaces, space layout, and audio environment, and (6) DDG-51 class shiphandling and maneuver characteristics, etc.

Surveys will be administered and include: a simulator sickness questionnaire (for VIBRaNT, see Attachment E), a NASA Task Load Index (TLX) workload questionnaire (Attachment F), a System Usability Survey (Attachment G), Trainee (Attachment H), OOD (Attachment J), and instructor surveys (Attachment I).

MSTCPAC (or MSTCLANT) NSST simulators will be used in the control condition. These legacy standalone physical room simulators use video projection to mimic the sea environment. Physical controls and consoles are constructed for participant input. Based on the participants input, the video simulation (projected onto the simulator windows) alters in accordance with programed simulation responses dictated by modeled ship hydrodynamics, environmental conditions, and bridge crew inputs, e.g., HELM, Lee HELM rudder, course, speed inputs, etc.

iii. Procedures:

1) VIBRaNT System Design:

Procedures employed during the VIBRaNT design phase, (Sprint 1/Sprint 2) design feedback and system effectiveness evaluations will be obtained during independent software testing sessions. Participant recruitment will occur independently of involvement in any other design or evaluation session. Involvement in design feedback doesn't preclude participation in an evaluation session. The pilot condition and feedback sessions will occur immediately upon IRB approval.

The control condition will be constrained to MSTCPAC/LANT training schedules of the NSSTs and thus system effectiveness evaluations will be scheduled around availability of the MSTCPAC and LANT. There will be three segments in the phase one design Sprint 1, one for each development feature (San Diego port, Norfolk port, and DDG-51 Class bridge). Collecting pilot data early will allow the research team to analyze what key events occur during training to replicate them in the VIBRaNT facilitated conditions. Once all developmental software features are completed and released, design phase Sprint 2 feedback sessions will occur. Participants in Sprint 2 will provide feedback regarding the training environment and scenario as a whole.

It is anticipated the research team will travel to participants located at 32nd Street Naval Station to observe control conditions in the building that houses the NSSTs and collect VIBRaNT system

design feedback. Design prototypes will be presented, which will consist of a Meta Quest 3 HMD and VIBRaNT software to display the ship navigation training scenario. The first design feedback sprint, which focuses on each development feature separately, will be conducted in-person and virtually to reach a maximum number of Naval Officers and qualified SWOs. As noted, the second design feedback sprint will be entirely in-person to explore the scenarios with all the developmental features integrated.

VIBRaNT system design and final delivery is currently scheduled to be completed by September of 2024, with the effectiveness evaluation (experimental condition) data collection phase occurring from September to May 2025.

2) VIBRaNT System Effectiveness:

Procedures employed during the VIBRaNT system effectiveness evaluations occur at Naval Base San Diego. The experimental condition employed for system evaluation will be in a controlled environment such as an empty classroom at MSTCPAC near a NSST simulator. The space will be cleared and made safe for the testing to allow participants to roam freely without objects obstructing their paths while wearing VIBRaNT HMDs.

Participants will first be briefed on the project and then asked to read the Privacy Act (Attachment B) and then read and sign the Informed Consent Document (Attachment C). The reading level of the instructions will be tested to ensure that it is no greater than an 8th grade reading level. Participants will then be asked to complete the pre-screening questionnaire (Attachment D). If participants answer that they have experienced motion sickness, vertigo, or epilepsy, they will be excluded from the experiment, and all data will be destroyed. Experimental sessions are expected to last no longer than an hour. If the control condition scenario lasts longer than an hour, the research team will only analyze the data from within the first hour and replicate the first hour of scenario events in the experimental condition/VIBRaNT system.

Except for the OOD using Pupil Labs Eye tracking glasses, the control condition will be conducted as ordinarily done in the NSST. OOD participants will be instructed on how to wear the eye tracking equipment and sign a release form prior to beginning an NSST training session. After completing training, participants will be asked to fill out the simulator sickness questionnaire (Attachment E), NASA TLX (Attachment F), SUS (Attachment G), Watchstander Self-Assessment Rubric (Attachment H), the Bridge Team Training Instructor Assessment Rubric (Attachment I), Watchstander OOD self/ instructor–Assessment Rubric (Attachment J).

Researchers will go to the training facility during NSST simulator downtime to match up the video and audio recordings to the instructor's view of key scenario events. Reaction time and task completion time will be evaluated. The same exact metrics will be used to evaluate both experimental conditions. Reaction and task completion time will be captured in the self-contained VIBRaNT system for playback.

6. Anticipated Benefits:

The potential benefits of XR IT to the Navy are vast. VIBRaNT study participants, along with Navy Instructors, and Navy Program Managers will have the opportunity to assess a mobile, highly scalable virtual reality training environment designed for USN use, to include a VR model of a Navy DDG-51 class bridge and VR environments of the USNs two largest ports, San Diego

and Norfolk. These virtual environments will provide bridge watchstanders, from junior officer to senior SWO's, training and shiphandling practice in a virtually limitless number of situations, from simple to extremely complex. It is anticipated VIBRaNT will increase bridge watch standing and shiphandling skills by providing additional time and flexibility to practice their skills. Additionally, the evaluation rubrics may provide trainees and instructors with a better format to debrief after training exercises which may further enhance performance. Participants will be able to share their expertise with researchers and provide their feedback as potential users of a system that will benefit the Navy. Participation in system effectiveness evaluations will aid in quantifying the benefits these technologies can bring to the Fleet and ensure end users are involved in technology development.

7. Other Reviews:

Review of VIBRaNT beyond the ONR Technical Solutions sponsor review will include reviews by NAVSEA PMS 339 PAPM, Mr. Cesar Perez. Data analysis will also be included in an NPS Information Science PhD Dissertation by Jon Brewster titled *Enhancing Knowledge Flow with XR*, and a review publication of Project VIBRaNT will either be submitted as a NIWC Pacific technical report and/or submission as a conference presentation.

8. Other Regulations:

N/A

9. Participant Recruitment:

Participant recruitment of active duty personnel with Surface Warfare Officer (SWO) qualifications currently assigned to ships at 32nd Street Naval Station and those at MSTCPAC will be managed by the PI via visits to MSTCPAC and/or direct email. Any involvement from coordination at the schoolhouse will be from admins who do not have any authority over participant's careers. This coordination will only serve as a liaison role between us and the participants. Any PII will only be retained by the researchers listed on this document.

The designated POC for MSTCPAC is LCDR Jeff Steiner, Force Navigator SWO Programs Manager for Commander, Naval Surface Force, U.S. Pacific Fleet. He may be reached at 619–437-3322 or via email jeffrey.c.steiner.mil@us.navy.mil.

No supervisor will be asked to conduct recruitment, whether civilian or military. A copy of the recruitment email is included in Attachment A.

For the testing phase, the intent is to recruit subjects from MSTC PAC and perform data collection (for both the experimental condition (XR) and the Control Condition (Mockup/NSST)) in the MSTC PAC NSST facility in San Diego. Other study testing locations mentioned are backup only in the event VIBRaNT cannot recruit enough subjects at MSTC PAC.

The Informed Consent Document (Attachment C) and Recruitment email (Attachment A) relates to participant recruitment and subject populations information.

A. VIBRaNT System Design

The PI and Co-PI will recruit civilian personnel for design feedback as needed from NIWC PAC and the Naval Postgraduate School (NPS). Recruitment will largely be based on connections to

known NIWC SME's, Naval Officers, and NPS professors who have expertise in the field. There is NO extended exposure to VR or risk to these recruits during this phase. For the design feedback sessions, a maximum of 200 total personnel (military) will be recruited by the PI.

B. VIBRaNT Effectiveness Evaluation

Experimental Condition

The PI and Co-PI will recruit military personnel from the Mariner Skills Training Center Pacific (MSTCPAC) in San Diego CA, with backup participants coming from MSTC Atlantic (MSTCLANT) in Norfolk VA and NPS in Monterey CA.

PI and Co-PI recruiting of bridge team participants shall be managed in concert with the MSTCPAC Scheduler and Instructors, the Surface Warfare Schools Command representative, and Ship's Commanding Officer for a participation period not to exceed one-hour in the experimental condition (VIBRaNT). In addition to Bridge Team Watchstander Recruiting, the PI will also recruit MSTCPAC Instructors to participate and observe during the VIBRaNT observation sessions, all the pre-screening questionnaires will be completed before conducting VIBRaNT testing.

(NOTE: If the PI is required to recruit from other military commands, e.g. MSTCLANT, an IRB amendment request to add these will be submitted for approval.)

A Participant Recruitment email has been created and is included as Attachment A.

A Letter of Approval from MSTC PAC has been obtained and is included as Attachment L.

A maximum of 27 bridge teams (USN, 9members per team), will be recruited for participation in both the experimental condition (VIBRaNT XR) and the control condition (NSST). Recruits will be chosen from ships company bridge watch teams scheduled for regular training taking place at MSTCPAC.

Upon arrival, recruits will be the provided with the Privacy Statement, see Attachment B, an Informed Consent Form, see Attachment C, a Pre-Screening Questionnaire, see Attachment D, and a Simulator Sickness Questionnaire, see Attachment E.

The PI will review the questionnaire responses and if the recruit is not screened out by the PI based on the questionnaire answers, he/she will be assembled in the VIBRaNT testing classroom for introductory training with the VIBRaNT HMD and Hand Controllers expected to last no more than 10 minutes. Based on the initial exposure to XR, the PI will check to ensure the recruit is Ok to proceed, if they are, the PI will admit them into the effectiveness evaluation period.

Control Condition

All participation will be voluntary, and participants will be informed they can withdraw from the study at any time with no repercussions. Before starting the 'control' or 'experimental' condition observations, the ICD and pre-screening questionnaires will first be conducted. Upon completing or (in the counter-balanced design) before beginning VIBRaNT experimental observations, recruits will conduct their normally scheduled bridge training in the NSST. The PI is not involved in NSST recruiting or scheduling as that is a MSTCPAC schoolhouse function.

10. Potential Risk for Participants:

Potential risk from participant exposure to VR may occur in either the design or evaluation phases of the VIBRaNT effort. Participants who have a prior history of mission sickness, suffer from vertigo, epilepsy, seizures, sensitive to extra weight straining their neck etc., will be excluded. These exclusion criteria will be included in the recruitment email (Attachment A), the Informed Consent Document (Attachment C), and will be asked about again in the Pre-Screening Questionnaire (Attachment D).

VR HMDs can induce simulator sickness, a form of motion sickness that occurs in some participants when using Augmented Reality/Virtual Reality/Mixed Reality (AR/VR/MR) HMDs. Symptoms can include dizziness and nausea. Participants will be instructed to notify the researchers immediately if symptoms of motion sickness emerge. All participants who use these headsets will be asked to complete a Simulator Sickness Questionnaire at the end of the session. Participants can withdraw from the study at any time with no repercussions.

Prior to use of any of VIBRaNT HMDs, the relevant safety considerations published by the company (i.e. Meta Platforms for Meta Quest 3/Meta Quest Pro) will be reviewed with the participants.

A. VIBRaNT System Design

No risks to personnel are anticipated for either the VIBRaNT Sprint 1 design review or Sprint 2 holistic study sessions. As these are strict question-and-answer interviews, with potentially light VR HMD use, potential risks to participants are assessed to be minimal.

B. VIBRaNT Effectiveness Evaluation

During the VIBRaNT system effectiveness evaluation sessions, risks and mitigation plans are listed below.

i Risk 1: Participants verbally report feeling symptoms of motion sickness during the study OR score moderate to high (greater than 24/48) on the Simulator Sickness Questionnaire. Or the researchers notice symptoms of simulator sickness such as disorientation. Mitigation 1: The session will be immediately ended, participant's data will be deleted, and a report of the incident will be made to the IRB committee. Participants will be kept in the experiment room until they feel better. Attempts will be made to bring them to different rooms from their fellow participants based on space availability and participants' ability to move. Participants will be instructed to focus on one point in space to help alleviate symptoms. Symptoms are only temporary and should subside after 15 minutes.

ii Risk 2: VR headsets are totally enclosed (i.e., one cannot see the outside world), and AR headsets may cause a lack of attention to the outside world. Therefore, participants may not know if they are about to collide with or trip over objects in the environment. Mitigation 2: Participants will also be continually monitored by the experimenters; participants should not need to move to operate the XR system and instead use a 'teleport' feature to move within the virtual environment.

iii Risk 3: Performance data released and tied to individuals could cause damage to individual careers, especially regarding active duty participants who might perform poorly or need to be re-trained/re-qualified for watch positions aboard ship. Mitigation: No PII will be released regardless of performance. The raw data, unique subject identifier number, and key linking identifier number to PII will be kept separately and stored separate encrypted hard drives in locked cabinets. For active duty participants, their identity and data will be anonymized in the same manner as civilians (see Section 15, Precautions re: Data). It is critical that participation information and/or performance data remain confidential and anonymous. No PII will be released including leadership. Only de-identified data and aggregated analysis may be released after proper approvals through the IRB. Participants will be encouraged to not divulge others participation in attempts to maintain anonymity.

11. Informed Consent:

Participant consent will be obtained and documented after first reading the Privacy Act (Attachment B). Then, participants will read, sign, and date an informed consent (Attachment C) form prior to participation in the study. This form is written at no greater than an 8th grade reading level. The consent form will provide information about the procedures used in the study, any risks and benefits, as well as any other information that will help each participant make an informed choice about whether they agree to participate in the study. This consent form will receive prior approval from the NIWC Pacific IRB. The participant will be given a copy of the informed consent document and privacy statement to take with them.

12. Questions Asked:

A. VIBRaNT System Design

During Sprint 1, participants will be asked to evaluate specific design features in an open and guided format. In Sprint 2, participants to be asked to 'think aloud' as they navigate through a scenario. If there are any points of contention, they will be asked to be elaborate. Any sticking points will be expanded upon after the SMEs are finished reviewing the scenario. Instructors will be asked the same kinds of questions but focused on the instructor GUI on the laptop which creates training scenarios, monitors bridge team performance during training, and can act as a virtual Helmsman and Lee Helmsman.

B. VIBRaNT Effectiveness Evaluation

Before beginning system effectiveness testing, participants will be familiarized with the symptoms related to simulator sickness per the simulator sickness questionnaire (Attachment E). Participants will be instructed to alert a researcher if they exhibit simulator sickness symptoms. The scenario will be paused, and the participants will be asked to fill out the simulator sickness questionnaire. Participants will be instructed to discontinue their involvement in testing if their symptoms exceed a total score of 24.

After each training scenario participants will complete the simulator sickness questionnaire (Attachment E), the NASA TLX workload questionnaire (Attachment F), the System Usability Scale (SUS, Attachment G), and performance effectiveness rubric (Attachment H, I and J). The purpose of each questionnaire is given below.

Attachment D. Pre-Screening questionnaire. This questionnaire was custom-developed for this study and asks (1) demographic information such as age, qualification level, and shiphandling experience, (2) familiarity with different kinds of technology, and (3) if they have previously experienced health symptoms that are relevant to the VR equipment used in this study. If a participants have previous experience with issues in section 3 or if they do not meet the appropriate shiphandling qualifications, they will be excluded from the study. Demographic and technology familiarity data will be kept and may be used in data analysis if it appears anecdotally that a post-hoc analysis would be useful.

Attachment E. Simulator Sickness Questionnaire. This questionnaire is a self-report instrument for participants to quantify if they are feeling simulator sickness. Each symptom is given a score of 0–3 depending on severity. If the total score is 24 points greater, the participant's session will be halted.

Attachment F. NASA Task Load Index (NASA TLX) workload questionnaire. This questionnaire is a self-report instrument for participants to report their subjective workload across six workload dimensions.

Attachment G. System Usability Scale (SUS). This questionnaire is a self-report instrument that allows participants to report the amount of enjoyment they felt from the system, how easy it was to use, and whether they would continue using it. This enables system designers to understand which aspects of the system may require additional work.

Attachment H. Watchstander Self-Assessment Rubric is based off rubrics used at the NSST and customed made for this study. This rubric will be given to each SWO after the control condition and experimental condition. They will be asked to rate their performance and situational awareness during the training. The rubric is on a Likert scale where the trainee rates how much they agree or disagree with the statement provided.

Attachment I. Bridge Team Training Instructor Assessment Rubric is based off rubrics used at the NSST and customed made for this study. This rubric will be given to instructors at the end of each control condition and experimental condition. The instructors will be asked to rate the team's performance based on a variety of elements and their experience operating the XR system. The rubric is on a Likert scale where the instructor rates how much they agree or disagree with the statement provided.

Attachment J. Watchstander OOD Self/ Instructor Assessment Rubric is designed with statements targeted specifically for the OOD. The rubric is based off rubrics used at the NSST and customed made for this study. The rubric is on a Likert scale where the OOD and instructor rates how much they agree or disagree with the statement provided. The rubric will be given after the control and experimental condition.

13. Private/Sensitive Data:

Privacy will be ensured by assigning participants a unique participant identification number (PIN) that will be associated with solely their anonymous data and nothing else. The participant's signature will be collected on the consent form; the consent forms will be stored in a locked and secure cabinet. Consent forms, surveys, and individual participant study data will be available only to the study PI and associate investigators listed on this protocol. Individual performance

from then on is labeled with this PIN and never using a name or other identifying information. Thus, we can aggregate measures across all individual performers without any Personally Identifying Information (PII) links. At no time will this information be paired up with participants. The VIBRaNT team is critically aware that if performance data were released and tied to individuals, it could damage individual careers, especially regarding active duty participants who might perform poorly or need to be re-trained/re-qualified for watch positions aboard ship. No PII will be released regardless of performance. The raw data, unique subject identifier number, and key linking identifier number to PII will be kept separately and stored separate encrypted hard drives in locked cabinets.

Data collected from the pre-screening questionnaire (Attachment D) may constitute PII. This is due to the nature of questions related to military experience. Because of this, the pre-screening questionnaire will be stored separately on an encrypted hard drive that will be locked in a secure cabinet, any printed forms will have the participant identifier and be locked in a locked cabinet. Only the information necessary for the data analysis will be matched to the study data (e.g., if running an analysis about the effect of previous shiphandling and port virtual environments experience, only the response to the prior shiphandling and port virtual environments experience question would be associated with the study data). Any files where background survey data and study data are combined will not contain PINs (e.g., will be de-identified) to minimize the risk that the combined data could be tied back to a participant's name.

14. Impact of Not Collecting Private/Sensitive Data:

The project's goal is to understand the general needs of a user community and then design, develop, and test an XR IT technology solution. Collection of private, sensitive, or self-identifying data from individual participants for the VIBRaNT evaluation phase will be used only to track participant involvement and ensure every data point is unique. All data will be analyzed in aggregate, averaging performance data across groups. The only private or sensitive data that may constitute PII will be from data gathered in the pre-screening questionnaire (Attachment D). However, this data will be stored separately, away from all other study data, and the utmost precaution will be taken to separate participant identity PII from performance related data.

15. Precautions Regarding Data:

All data results will be maintained in a locked file cabinet in Building 365 Room 223, NIWC Pacific, San Diego, CA, with only the Investigators having access. The raw data, unique subject identifier number, and key linking identifier number to PII will be kept separately and stored separate encrypted hard drives in locked cabinets. Files will be marked "CONTROLLED UNCLASSIFIED INFORMATION" in accordance with NIWC Pacific Instruction 5210.1. Performance is protected using the assigned arbitrary participant numbers. All data results will be retained for five years after completion of the study and will then be transferred to a Federal Records Center (FRC) in accordance with SECNAV M-5210.1.

16. Attachments:

A. Participant Recruitment Email
B. Privacy Statement
C. Informed Consent Form
D. Pre-Screening Questionnaire
E. Simulator Sickness Questionnaire
F. NASA Task Load Index (TLX) Workload Questionnaire
G. System Usability Scale
H. Self-Assessment Rubric: Bridge Watchstander
I. Instructor Assessment Rubric: Bridge Watchstander
J. Instructor Assessment Rubric: OOD
K. Exit Interview Questions
L. Letter of Support from the Commander Navy Surface Forces Pacific

APPENDIX K. NATIONAL AERONAUTICS AND SPACE ADMINISTRATION TASK LOAD INDEX SURVEY

IRB approved NASA TLX Survey Questionnaire.

(IRB Attachment F)

NASA Task Load Index (TLX) Questionnaire

Participant #:_____ Session #:_____

Please place an "X" along each scale at the point that best indicates your experience with the system.

Mental Demand: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the mission easy or demanding, simple or complex, exacting or forgiving?

Low |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| High

Physical Demand: How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the mission easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Low |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| High

Temporal Demand: How much time pressure did you feel due to the rate or pace at which the mission occurred? Was the pace slow and leisurely or rapid and frantic?

Low |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| High

Performance: How successful do you think you were in accomplishing the goals of the mission? How satisfied were you with your performance in accomplishing these goals?

Low |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| High

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Low |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| High

Frustration: How discouraged, stressed, irritated, and annoyed versus gratified, relaxed, content, and complacent did you feel during your mission?

Low |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| High

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX L. SYSTEM USABILITY SCALE

IRB approved SUS Survey Questionnaire.

(IRB Attachment G)

SYSTEM USABILITY SCALE

Session #: _____ Participant #:_____

Please circle the number that best represents your agreement/disagreement with each statement.

1. I think that I would like to use this system frequently.

Strongly Disagree	**1**	**2**	**3**	**4**	**5**	*Strongly Agree*

2. I found the system unnecessarily complex.

Strongly Disagree	**1**	**2**	**3**	**4**	**5**	*Strongly Agree*

3. I thought the system was easy to use.

Strongly Disagree	**1**	**2**	**3**	**4**	**5**	*Strongly Agree*

4. I think that I would need the support of a technical person to be able to use this system.

Strongly Disagree	**1**	**2**	**3**	**4**	**5**	*Strongly Agree*

5. I found the various functions in this system were well integrated.

Strongly Disagree	**1**	**2**	**3**	**4**	**5**	*Strongly Agree*

6. I thought there was too much inconsistency in this system.

Strongly Disagree	**1**	**2**	**3**	**4**	**5**	*Strongly Agree*

7. I would imagine that most people would learn to use this system very quickly.

Strongly Disagree	**1**	**2**	**3**	**4**	**5**	*Strongly Agree*

8. I found the system very cumbersome to use.

Strongly Disagree	**1**	**2**	**3**	**4**	**5**	*Strongly Agree*

9. I felt very confident using the system.

Strongly Disagree	**1**	**2**	**3**	**4**	**5**	*Strongly Agree*

10. I needed to learn a lot of things before I could get going with this system.

Strongly Disagree	**1**	**2**	**3**	**4**	**5**	*Strongly Agree*

Comments:

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX M. BRIDGE WATCHSTANDER SELF-ASSESSMENT SURVEY

IRB approved Bridge Watchstander Self-Assessment Survey.

(IRB Attachment H)

Self-Assessment Rubric: Bridge Watchstander

Session #: _____ Evaluator #:_____

(Circle rating below, 1 = Disagree, 4 = Neutral, and 7 = Agree, and N/A = Not Applicable)
A box is provided at the end of this survey for additional comments or recommendations.

1. I was able to react in sufficient time to commands given.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

2. I could **not** make decisions with sufficient time for the action to be carried out.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

3. I was **not** able to accomplish the scenario's mission.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

4. I was able to consistently communicate with CIC.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

5. I could **not** effectively exercise communication procedures with CCS.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

6. I was able to communicate with other vessels via Bridge to Bridge radio when needed.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

7. I could **not** focus on the correct area of interest for the situation at hand.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

8. I was able to correctly apply Rules of the Road.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

9. I was **not able to** provide forceful OOD backup.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

10. I was able to notify the OOD of potential violations of Rules of the Road.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

11. I was able to identify and point out potential collision/near miss situations.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

12. I was **not** able to perceive potential dangers.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

13. My bridge team used standard commands to effectively conn the ship.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

14. My bridge team did **not** perceive set and drift conditions to conn the ship.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

15. My bridge team foresaw situational dynamics needed to conn the ship.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

16. My bridge team was **not** able to operate our ship in accordance with COLREGS.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

17. It was easy for me to use the training system.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

18. I had **trouble** interacting with the simulated VMS.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

19. I could easily interact with the simulated RADAR.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

20. I had **trouble** interacting with the simulated HELM/Lee HELM controls.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

21. It was easy to interact with my fellow bridge team members during the training scenario.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

22. The training system **interfered** with learning of the content in the scenario.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

23. The training system will **not** help refine my seamanship/shiphandling skills.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

24. After the training began, I forgot about the training system and was immersed in the scenario.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

25. The training system was **not** a good representation of a ship's bridge environment.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

26. The training system was a good representation of a DDG-51's hydrodynamics.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

27. Using this training system helps build bridge team cohesion.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

28. The training system does **not** provide a sufficient way to practice proper execution of COLREGS.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

29. The training system did **not** give the sense of motion needed in assessing Set & Drift.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

30. The training system gave the motion sense needed in assessing Advance & Transfer.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX N. OFFICER OF THE DECK SELF-ASSESSMENT SURVEY

IRB approved OOD self-assessment survey questionnaire.

(IRB Attachment J)

OOD Self-Assessment Rubric

Session #: _____ Evaluator #:_____

(Circle rating below, 1 = Disagree, 4 = Neutral, and 7 = Agree, and N/A = Not Applicable)
A box is provided at the end of the survey for any additional comments/recommendations.

1. The OOD was **unable** to vocalize directions in sufficient time for actions to be carried out.

1	**2**	**3**	**4**	**5**	**6**	**7**	**N/A**
Disagree			*Neutral*			*Agree*	

2. The OOD's vocalized directions were easily heard and understood by his/her watchstanders.

1	**2**	**3**	**4**	**5**	**6**	**7**	**N/A**
Disagree			*Neutral*			*Agree*	

3. Watchstanders had to ask for repeated commands due to **not** understanding OOD vocalizations.

1	**2**	**3**	**4**	**5**	**6**	**7**	**N/A**
Disagree			*Neutral*			*Agree*	

4. The OOD was able to consistently confirm watchstander actions.

1	**2**	**3**	**4**	**5**	**6**	**7**	**N/A**
Disagree			*Neutral*			*Agree*	

5. The OOD was able to maintain situational awareness even as updated orders were issued.

1	**2**	**3**	**4**	**5**	**6**	**7**	**N/A**
Disagree			*Neutral*			*Agree*	

6. When problems emerged, the OOD **was unable** to issue new orders needed to ensure an appropriate outcome.

1	**2**	**3**	**4**	**5**	**6**	**7**	**N/A**
Disagree			*Neutral*			*Agree*	

7. The OOD was able to focus on the correct area of interest for the situation at hand.

1	**2**	**3**	**4**	**5**	**6**	**7**	**N/A**
Disagree			*Neutral*			*Agree*	

8. The OOD was able to project confidence to his/her bridge team and superior officers.

1	**2**	**3**	**4**	**5**	**6**	**7**	**N/A**
Disagree			*Neutral*			*Agree*	

9. The OOD was able to direct the JOOD appropriately.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

10. The OOD was **not** able to monitor the CONN appropriately.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

11. The OOD was able to ensure HELM and Lee HELM/Conning Officer used standard commands.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

12. The OOD was **not** able to adhere to the COs standing orders.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

13. The OOD was **unable** to employ VMS system appropriately for navigation.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

14. The OOD was able to use the RADAR system for contact management and relative vectors.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

15. The OOD was able to manage team communication so watchstanders conversed only as needed.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

16. The OOD was able to consistently communicate with CIC.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

17. The OOD was **unable** to communicate casualty reports with CCS in a timely manner.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

18. The OOD easily communicated with CO in terms that a mariner would find useful.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

19. The OOD was **unable** to follow the COs procedures for communicating Contact Reports.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

20. The OOD ensured communication with other vessels via Bridge to Bridge radio when needed.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

21. Bridge team members were empowered and able to question/provide feedback to the OOD.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

22. The OOD and his/her bridge team were **not** able to work effectively together.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

23. The OOD could **not**/did **not** follow the Rules of the Road.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

24. The OOD was **unable** to demonstrate knowledge of effective shiphandling.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

25. The OOD **poorly** managed HELM/Lee HELM shiphandling operations.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

26. The OOD was able to operate in accordance with COLREGS.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

27. The OOD could **not** effectively accomplish the scenario's mission.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

28. The training platform **interfered** with the OOD's ability to execute the mission.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX O. INSTRUCTOR ASSESSMENT SURVEY: BRIDGE WATCH TEAM

IRB approved Instructor Bridge Watchstander Assessment Survey Questionnaire.

(IRB Attachment I)

Instructor Assessment Rubric: Bridge Watch Team

Session #: _____ Evaluator #:_____

(Circle rating below, 1 = Disagree, 4 = Neutral, and 7 = Agree, and N/A = Not Applicable)
A box is provided at the end of this survey for additional comments or recommendations.

1. The bridge team had ample cues presented to react in sufficient time to the situation.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

2. The bridge team was **unable** to make decisions in time for action to be carried out.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

3. The bridge team was able to accomplish the scenario's mission.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

4. The bridge team did was **not** able to consistently communicate with CIC.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

5. The bridge team was able to effectively exercise communication procedures with CCS.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

6. The bridge team was **not** able to communicate via Bridge to Bridge radio when needed.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

7. The bridge team was able to correctly apply Rules of the Road.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

8. Bridge Team members were **not** able to provide forceful OOD backup.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

9. The bridge team was able to notify the OOD of potential violations of the Rules of the Road.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

10. The bridge team was able to point out potential collision/near miss situations.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

11. The bridge team was **not** able to perceive potential dangers.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

12. The bridge team used standard commands to effectively conn the ship.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

13. The bridge team could **not** perceive set and drift conditions to conn the ship.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

14. The bridge team foresaw environmental dynamics needed to conn the ship.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

15. The bridge team was **not** able to operate their ship in accordance with COLREGS.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

16. It was easy for the bridge team to use the training system.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

17. The bridge team had **trouble** interacting with the simulated VMS.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

18. The bridge team easily interacted with the simulated RADAR.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

19. The bridge team had **trouble** interacting with the simulated HELM/Lee HELM controls.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

20. The bridge team easily interacted with fellow team members during the training scenario.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

21. The training system **interfered** with learning of the content in the scenario.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

22. The training system will **not** help refine the bridge team's seamanship/shiphandling skills.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

23. The training system was a good representation of a ship's bridge environment.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

24. Using this training system helped to build bridge team cohesion.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

25. The training system did **not** provide a sufficient way to practice proper execution of COLREGS.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

26. I could easily build a custom scenario from scratch in this training system.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

27. I could **not** easily load a pre-existing scenario into the training system.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

28. It was **difficult** to add on the fly adjustments to the scenario in this training system.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

29. I could easily monitor the bridge team in real-time.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

30. I could easily interact as a virtual HELM/ Lee HELM from the planner mode.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

31. It was **difficult** to replay the training scenario.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

32. Debriefing trainees on their performance would be effective with this system.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

33. The training platform **did not** interfere with the team's ability to make sound and timely decisions.

1	2	3	4	5	6	7	N/A
Disagree			*Neutral*			*Agree*	

APPENDIX P. EXIT INTERVIEW QUESTIONS

IRB approved Exit Interview Questionnaire.

(IRB Attachment J)

VIBRaNT Exit Interview Questions:

(These are to be addressed and transcribed via MS Teams.)

Session #: _____ Participant #:_____

1. Please briefly describe your experience standing watch in VIBRaNT.

2. Please describe anything notable that stood out to you in using VIBRaNT.

3. After experiencing VIBRaNT, please describe other uses/use case for XR that come to mind.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX Q. EXIT INTERVIEW WORD CLOUDS

The following LLM generated word clouds, highlight the sentiments of bridge watchstanders using XR for training. These word clouds were created using Open AI's GPT4 LLM after analysis of 152 participant responses to the questions asked using the VIBRaNT IRB Attachment K, Exit Interview Questions (Appendix P).

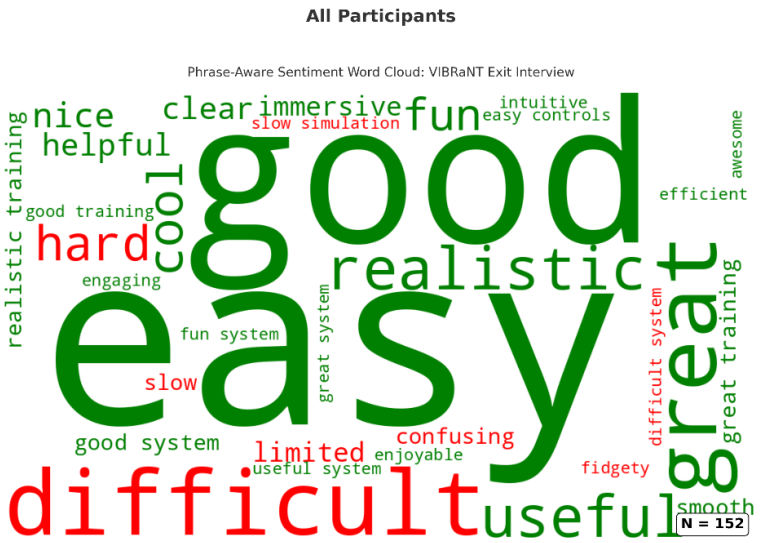


Figure A.31. Exit interview word cloud for all participants response.



Figure A.32. Exit interview word cloud for officer of the deck only response.



Figure A.33. Exit interview word cloud for junior officer of the deck only response.



Figure A.34. Exit interview word cloud for conning officer only response.

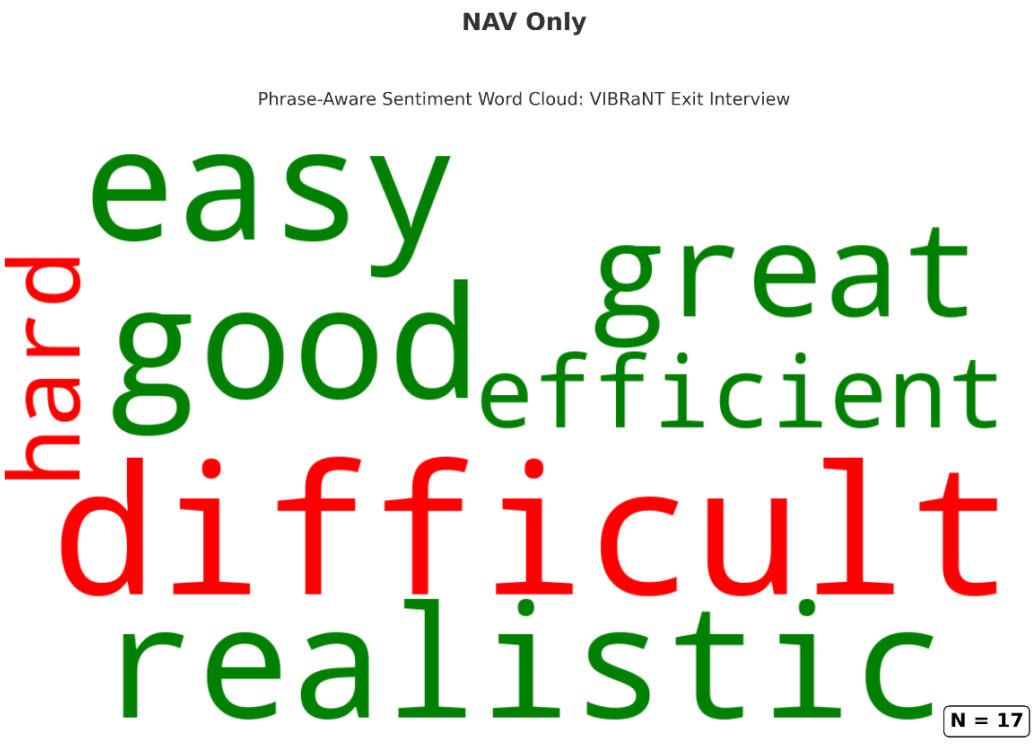


Figure A.35. Exit interview word cloud for navigator only response.



Figure A.36. Exit interview word cloud for helmsman only response.



Figure A.37. Exit interview word cloud for lee helmsman only response.

In addition to participant subgroup responses, word clouds were generated based on other filtered data; these included "Male Participants Only," "Female Participants Only," "VIBRaNT (XR) First," and "NSST First." Experimental (XR) condition first vs. the control (NSST) condition first examined how XR exposure impacted NSST perceptions as reflected in the associated word cloud responses.

A notable observation is that participants who trained first in the control condition, then the experimental condition seemed to reflect "difficulty" in an inverse way compared to those who experienced the reverse order; (those who trained first in the experimental condition first, then trained in the control condition). The purpose of the research was to evaluate and characterize XR, wherein impacts of XR on the NSST perceptions were of primary concern, not the impacts of NSST—the legacy accepted training method–on the experimental condition. Of note here is the appearance of a negative impact of NSST exposure on XR perceptions, which may warrant further study to fully appreciate the consequences of introducing XR into the overall Navy training system.

Male Participants Only



Figure A.38. Exit interview word cloud for male participants only response.

Female Participants Only



Figure A.39. Exit interview word cloud for female participants only response.

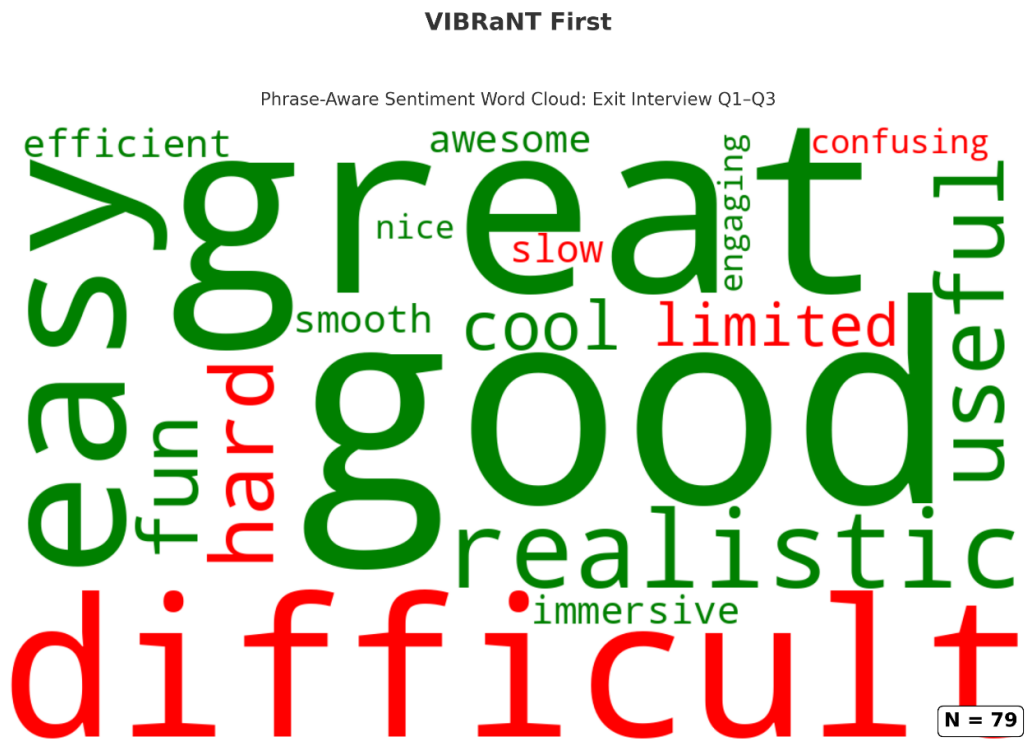


Figure A.40. Exit interview word cloud for extended reality first response.

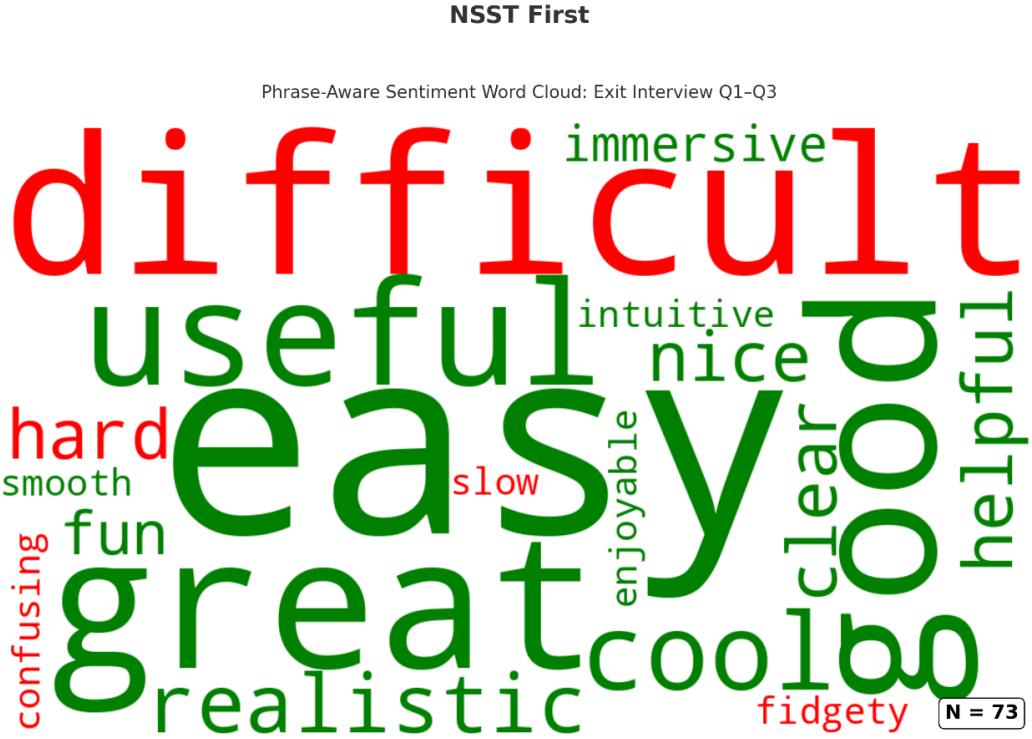


Figure A.41. Exit interview word cloud for Navigation, Seamanship, Shiphandling Trainer first response.

LIST OF REFERENCES

Adriaans, P. (2024). Information. In *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2024/entries/information/

Aguayo, C., Videla-Reyes, R., & Veloz, T. (2023). Entangled cognition in immersive learning experience. *Adaptive Behavior*, *31*(5), 497–515. https://psycnet.apa.org/doi/10.1177/10597123231183996

Aksu, Ş. H. (2025). An examination of the interrelationships among NASA-TLX. *Frontiers in Psychology*, *16*. https://pmc.ncbi.nlm.nih.gov/articles/PMC11968110/

Alexias, G., & Dimitropoulou, E. (2011). The body as a tool: Professional classical ballet dancers' embodiment. *Research in Dance Education*, *12*(2), 87–104. https://psycnet.apa.org/doi/10.1080/14647893.2011.575221

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, *21*(2), 247–286. https://doi.org/10.1080/10508406.2011.611446

Allen, L. (1993, August 10). *Evolution of flight simulation.* [Paper presentation]. Flight Simulation and Technologies, Monterey, CA. https://doi.org/10.2514/6.1993-3545

American Productivity Quality Center. (2023, August 23). *American Productivity and Quality Center's Knowledge Flow Process Framework*. https://www.apqc.org/resource-library/resource-listing/apqcs-knowledge-flow-process-framework

Aristotle. (1989). *Aristotle in 23 volumes* (H. Tredennick, Trans.; Vol. 17). Harvard University Press.

Bailenson, J. (2019). *Experience on demand: What virtual reality is, how it works, and what it can do*. W.W. Norton & Company.

Bangor, A., Kortum, P., & Miller, J.. (2008). An empirical evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Bapuji, H., & Crossan, M. (2007). Knowledge types and knowledge management strategies. In M. Gilbert & T. Durand (Eds.), *Strategic Networks, Learning to Compete* (pp. 8–25). Blackwell Publishing Ltd.

Barrett, L. F. (2020). *Seven and a half lessons about the brain*. Houghton Mifflin.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Psycholological Review*, *58*(1), 373–403. https://doi.org/10.1146/annurev.psych.58.110405.085709

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2010). Grounded cognition: Past, present, and future. *Topics in Cognitive Science*, *2*(4), 716–724. https://doi.org/10.1111/j.1756-8765.2010.01115.x

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364. https://doi.org/10.1038/nn.4502

Block, N. (1993). Review of Consciousness Explained. *The Journal of Philosophy*, *90*(4), 181–193. https://doi.org/10.2307/2940970

Borko, H. (1968). Information science: What is it? *American Documentation*, *19*(1), 3–5. https://doi.org/10.1002/asi.5090190103

Braarud, P. Ø. (2021). Investigating the validity of subjective workload rating (NASA TLX) and subjective situation awareness rating (SART) for cognitively complex human–machine work. *International Journal of Industrial Ergonomics*, *84*(103233). https://doi.org/10.1016/j.ergon.2021.103233

Brown, S. (2009). Paper prototypes and beyond. *Visible Language*, *43*(2), 198.

Buchweitz, A., & Prat, C. (2013). The bilingual brain: Flexibility and control in the human cortex. *Physics of Life Reviews*, *10*(4), 428–443. https://doi.org/10.1016/j.plrev.2013.07.020

Caso, S. (2024). Emerging technologies in military space operations: Current applications and future research for educational and training purposes. *International Journal of Training Research*, *23*(2), 140–155. https://doi.org/10.1080/14480220.2024.2431482

Chalmers, D. (2019). Andy Clark and His Critics. In M. Colombo, E. Irvine, & M. Stapleton (Eds.), *Extended cognition and extended consciousness* (pp. 9–20). Oxford University Press.

Chandramouli, M., Zafar, A., & Williams, A. (2025). A detailed review of the design and evaluation of XR applications in STEM education and training. *Electronics*, *14*(19), 3818. https://doi.org/10.3390/electronics14193818

Chua, A. (2002). The influence of social interaction on knowledge creation. *Journal of Intellectual Capital*, *3*(4), 375–392. https://doi.org/10.1108/14691930210448297

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, *3*(9), 345–351. https://doi.org/10.1016/S1364-6613(99)01361-3

Clark, A. (2010). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press.

Clark, A. (2015). Embodied prediction. In T. Metzinger & J. Windt (Eds.), *Open Mind*. Frankfurt am Main: MIND Group. http://dx.doi.org/10.15502/9783958570115

Clark, A. (2023). *The experience machine: how our minds predict and shape reality*. Pantheon.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19. http://www.jstor.org/stable/3328150?origin=JSTOR-pdf

Collins, H. (2019). *Tacit and explicit knowledge*. University of Chicago Press.

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (Vol. 1195). Houghton Mifflin Boston, MA.

Coope, U. (2020). Aristotle on Productive Understanding and Completeness. In T. K. Johansen (Ed.), *Productive knowledge in ancient philosophy: the concept of technê* (pp. 109–130). Cambridge University Press. https://dx.doi.org/10.1017/9781108641579.006

Dennett, D. (2001). Are we explaining consciousness yet? *Cognition*, *79*(1-2), 221–237. https://doi.org/10.1016/s0010-0277(00)00130-x

Dennett, D. C. (2018). *Conciousness explained* (1st ed.). Little, Brown and Company.

Dickie, I. (2012). Skill before knowledge. *Philosophy and Phenomenological Research*, *85*(3), 737–745. https://doi.org/10.1111/j.1933-1592.2012.00638.x

Draschkow, D., Anderson, N. C., David, E., Gauge, N., Kingstone, A., Kumle, L., Laurent, X., Nobre, A. C., Shiels, S., & Võ, M. L.-H. (2023). Using XR (extended reality) for behavioral, clinical, and learning sciences requires updates in infrastructure and funding. *Policy Insights from the Behavioral and Brain Sciences*, *10*(2), 317–323. https://doi.org/10.1177/23727322231196305

Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(3), 319–325. https://doi.org/10.1002/wcs.1226

Fridland, E. (2017). Skill and motor control: Intelligence all the way down. *Philosophical Studies*, *174*(6), 1539–1560. https://doi.org/10.1007/s11098-016-0771-7

Gardner, H. E. (2011). *Frames of mind: The theory of multiple intelligences*. Basic Books.

Geisler, E., & Wickramasinghe, N. (2015). *Principles of knowledge management: Theory, practice, and cases*. Routledge.

Gilson, M., Burkitt, A., & van Hemmen, J. L. (2010). STDP in recurrent neuronal networks. *Frontiers in Computational Neuroscience*, *4*, 23. https://doi.org/10.3389/fncom.2010.00023

Grandin, T. (2023). *Visual Thinking*. Riverhead Books.

Grant, S. C., & Magee, L. E. (1998). Contributions of proprioception to navigation in virtual environments. *Human Factors*, *40*(3), 489–497. https://doi.org/10.1518/001872098779591296

Greco, J. (2010). *Achieving knowledge: A virtue-theoretic account of epistemic normativity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511844645

Greenwold, S. (2003). *Spatial computing* (Publication Number 53484083) Massachusetts Institute of Technology]. Cambridge, MA. http://dspace.mit.edu/handle/1721.1/7582

Hart, S. G. (2006). *NASA-task load index (NASA-TLX); 20 years later* [Paper presentation]. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, https://www.researchgate.net/publication/235705634_Nasa-task_load_index_Nasa-TLX_20_years_later

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (Vol. 52, pp. 139–183). Elsevier. https://doi.org/10.1016/S0166-4115(08)62386-9

Hays, R. T., & Vincenzi, D. A. (2000). Fleet assessments of a virtual reality training system. *Military Psychology*, *12*(3), 161–186. https://psycnet.apa.org/doi/10.1207/S15327876MP1203_1

Hebb, D. O. (2005). *The organization of behavior: A neuropsychological theory*. Psychology Press.

Huang, W., Chen, H., Liu, Z., Dong, X., Feng, G., Liu, G.,..., & Su, L. (2025). Individual variability in the structural connectivity architecture of the human brain. *Journal of Neuroscience*, *45*(5). https://doi.org/10.1523/jneurosci.2139-23.2024

Huddle, J. (2024). Thinking in Pictures: Temple Grandin's Contribution to Visual Literacy. In *Connecting Visual Literacy to Theory* (pp. 139–150). Routledge.

Huff, C. (2023). *An Empirical Analysis to Determine a Common Augmented Reality or Mixed Reality Solution to Improve Training and Operational Capabilities for the Marine Corps* [Masters, Naval Postgraduate School]. Monterey, CA. https://dair.nps.edu/handle/123456789/5230

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Hutchins, S. G., Bordetsky, A., Kendall, T., & Garrity, M. (2007). Evaluating a model of team collaboration via analysis of team communications. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *51*(4), 415–419. https://doi.org/10.1177/154193120705100456?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle

Johansen, T. K. (2017). Aristotle on the logos of the craftsman. *Phronesis*, *62*(2), 97–135. https://doi.org/10.1163/15685284-12341321

Joyce, R. A. (2007). Embodied subjectivity: gender, femininity, masculinity, sexuality. In L. Meskell & R. Preucel (Eds.), *A Companion to Social Archaeology* (pp. 82–95). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470693605.ch4

Kahneman, D. (2011). *Thinking, fast and slow*. Random House.

Kanellopoulos, A., Koutsouba, M., & Giossos, Y. (2021). Proposition for the introduction of the concept telemathesis in videoconferencing in distance education. *European Journal of Open, Distance and E-Learning*, *23*(2), 83–98. https://www.researchgate.net/publication/347928075

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220. https://psycnet.apa.org/doi/10.1207/s15327108ijap0303_3

Khan, R. et al.. (2025). Psychometric evaluation of the System Usability Scale. *Bmc Medical Research Methodology*, *25*. https://doi.org/10.1186/s12874-025-01234-5

Klug, B. (2017). An overview of the System Usability Scale in library website and system usability testing. *Weave: Journal of Library User Experience*, *1*(6). https://doi.org/10.3998/weave.12535642.0001.602

Lakoff, G. (2012). Explaining embodied cognition results. *Topics in Cognitive Science*, *4*(4), 773–785. https://doi.org/10.1111/j.1756-8765.2012.01222.x

Langendorf, L. N., & Khalid, M. S. (2025). Systematic literature review on usability and training outcomes of using digital training technologies in industry. *Computers in Human Behavior Reports*, *17*, 100604. https://doi.org/10.1016/j.chbr.2025.100604

Leitch, J. M., & Rosen, P. W. (2001). Knowledge management, CKO, and CKM: The keys to competitive advantage. *The Manchester Review*, *6*(2), 9–13. https://www.researchgate.net/publication/237068037

Lewis, J. R. (2018). The System Usability Scale: Past, present, and future. *International Journal of Human–Computer Interaction*, *34*(7), 577–590. https://doi.org/10.1080/10447318.2018.1455307

Lewis, J. R., & Sauro, J.. (2009, July 19–24, 2009). *The factor structure of the System Usability Scale* [Paper presentation]. Human Centered Design: First International Conference, HCD 2009, San Diego, CA. https://doi.org/10.1007/978-3-642-02806-9_12

Lewis, J. R., & Sauro, J.. (2017). Revisiting the factor structure of the System Usability Scale. *Journal of Usability Studies*, *12*(4), 183–192. https://uxpajournal.org/revisit-factor-structure-system-usability-scale

Lorenz, H., & Morison, B. (2019). Aristotle's empiricist theory of doxastic knowledge. *Phronesis*, *64*(4), 431–464. https://doi.org/10.1163/15685284-12341975

Loring, D. W., Meador, K., Allison, J. D., & Wright, J. (2000). Relationship between motor and language activation using fMRI. *Neurology*, *54*(4), 981–983. https://doi.org/10.1212/wnl.54.4.981

Ludwig, D. (2015). Extended cognition and the explosion of knowledge. *Philosophical Psychology*, *28*(3), 355–368. https://doi.org/10.1080/09515089.2013.867319

Magee, J. C., & Johnston, D. (1997). A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. *Science*, *275*(5297), 209–213. https://doi.org/10.1126/science.275.5297.209

Maldonado, P. E., Maturana, H., & Varela, F. J. (1988). Frontal and lateral visual system in birds: Frontal and lateral gaze. *Brain, Behavior and Evolution*, *32*(1), 57–62. https://doi.org/10.1159/000116532

McCoy-Fisher, C., Mishler, A., Bush, D., Severe-Valsaint, G., Natali, M., & Riner, B. (2019). *Student naval aviation extended reality device capability evaluation*. Naval Air Warfare Center Training Systems Division. Orlando, FL Retrieved from https://apps.dtic.mil/sti/pdfs/AD1103227.pdf

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). *Augmented reality: A class of displays on the reality-virtuality continuum* [Paper presentation]. Telemanipulator and Telepresence Technologies, Boston, MA. https://doi.org/10.1117/12.197321

Mingers, J. (2001). Embodying information systems: The contribution of phenomenology. *Information and Organization*, *11*(2), 103–128. https://doi.org/10.1016/S1471-7727(00)00005-1

Mingers, J., & Willcocks, L. (2014). An integrative semiotic framework for information systems: The social, personal and material worlds. *Information and Organization*, *24*(1), 48–70. https://doi.org/10.1016/j.infoandorg.2014.01.002

Miton, H., & DeDeo, S. (2022). The cultural transmission of tacit knowledge. *Journal of the Royal Society Interface*, *19*(195), 20220238. https://doi.org/10.1098/rsif.2022.0238

Moss, J. D., & Muth, E. R. (2011). Characteristics of head-mounted displays and their effects on simulator sickness. *Human Factors*, *53*(3), 308–319. https://doi.org/10.1177/0018720811405196

Mouloua, S. A., Matthews, G., French, J., & Mouloua, M. (2023). Extended Reality in training environments. In D. Vincenzi, M. Mouloua, P. Hancock, J. Pharmer, & J. Ferraro (Eds.), *Human Factors in Simulation and Training: Application and Practice* (pp. 151). Taylor & Francis Group. https://doi.org/10.1201/9781003401353

Mun, J. (2014). *Real Options Valuation BizStats*. In ROV Press. https://www.realoptionsvaluation.com/product/rov-biz-stats/

Nissen, M. E. (2002). An extended model of knowledge-flow dynamics. *Communications of the Association for Information Systems*, *8*(1), 251–266. https://doi.org/10.17705/1CAIS.00818

Nissen, M. E. (2006). *Harnessing knowledge dynamics: Principled organizational knowing & learning* IGI Global. https://10.4018/978-1-59140-773-7

Nissen, M. E. (2014). *Harnessing Dynamic Knowledge Principles in the Technology-Driven World*. IGI GLobal.

Nissen, M. E., & Bordetsky, A. (2011). Leveraging mobile network technologies to accelerate tacit knowledge flows across organisations and distances. In G. Trentin (Ed.), *Technology and Knowledge Flow: The Power of Networks* (pp. 1–25). Chandos Publishing. https://doi.org/10.1016/B978-1-84334-646-3.50001-0

Nonaka, I. (2009). The knowledge-creating company. In T. Siesfeld, J. Cefola, & D. Neef (Eds.), *The Economic Impact of Knowledge* (pp. 175–187). Routledge. https://doi.org/10.4324/9780080505022

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How japanese companies create the dynamics of innovation*. Oxford University Press.

Nonaka, L., Takeuchi, H., & Umemoto, K. (1996). A theory of organizational knowledge creation. *International Journal of Technology Management*, *11*(7-8), 833–845. https://doi.org/10.1504/IJTM.1996.025472

Pavese, C. (2017). Know-how and gradability. *Philosophical Review*, *126*(3), 345–383. https://doi.org/10.1215/00318108-3878493

Pavese, C. (2022). Knowledge How. In *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/fall2022/entries/knowledge-how/

Perry, M. (2003). Distributed cognition. In J. Carroll (Ed.), *Hci Models, Theories, and Frameworks: Toward a Multidisciplinary Science* (pp. 193–223). Elsevier.

Polanyi, M. (1967). *The tacit dimension*. University of Chicago Press.

Popp, J. L., Thiele, J. A., Faskowitz, J., Seguin, C., Sporns, O., & Hilger, K. (2024). Structural-functional brain network coupling predicts human cognitive ability. *Neuroimage*, *290*(120563). https://doi.org/10.1016/j.neuroimage.2024.120563

Prat, C. (2022). *The neuroscience of you: how every brain is different and how to understand yours*. Penguin.

Prat, O., Petrucco, L., Štih, V., & Portugues, R. (2022). Comparing the representation of a simple visual stimulus across the cerebellar network. *Biorxiv*, *11*(7), 1–15. https://doi.org/10.1523/ENEURO.0023-24.2024

Pritchard, D. (2012). Anti-Luck Virtue Epistemology. *Journal of Philosophy*, *109*(3), 247–279. https://doi.org/10.5840/jphil201210939

Ross, G., & Gilbey, A. (2023). Extended reality (xR) flight simulators as an adjunct to traditional flight training methods: a scoping review. *Ceas Aeronautical Journal*, *14*(4), 799–815. https://doi.org/10.1007/s13272-023-00688-5

Ryle, G. (1945, 5 November 1945). Knowing How and Knowing That: The Presidential Address.

Ryle, G. (1949). *The Concept of Mind*. University of Chicago Press.

Salinas, D., Munoz-La Rivera, F., & Mora-Serrano, J. (2022). Critical analysis of the evaluation methods of extended reality (XR) experiences for construction safety. *International Journal of Environmental Research and Public Health*, *19*(22), 15272. https://doi.org/10.3390/ijerph192215272

Sallam, A., Eid, A., Elfaramawy, A., & Khodier, L. (2018). The Academic Research Community Publication. https://doi.org/10.1051/matecconf/201821804010

Shapiro, L. (2019). *Embodied cognition*. Routledge.

Simon, H. A. (1996). *Models of my Life*. MIT Press.

Small, C. T., & Sage, A. P. (2006). Knowledge management and knowledge sharing: A review. *Information Knowledge Systems Management*, *5*(3), 153–169. https://doi.org/10.3233/IKS-2006-00089

Smart, P. (2022). Minds in the metaverse: Extended cognition meets mixed reality. *Philosophy & Technology*, *35*(4), 87. https://doi.org/10.1007/s13347-022-00580-w

Sosa, E. (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge* (Vol. 1). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199297023.001.0001

Sosa, E. (2009). *Reflective Knowledge: Apt Belief and Reflective Knowledge* (Vol. 2). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199217250.001.0001

Sosa, E. a. L. F. C. (2020). Skill and Knowledge. In E. a. C. P. Fridland (Ed.), *The Routledge Handbook of Philosophy of Skill and Expertise* (pp. 146–156). Routledge. https://doi.org/10.4324/9781315180809

Sporns, O. (2010). *Networks of the Brain*. MIT Press.

Stanley, J. a. T. W. (2001). Knowing How. *Journal of Philosophy*, *98*(8), 411–444. https://doi.org/10.2307/2678403

Stanley, J. a. T. W. (2017). Skill. *Noûs*, *51*(4), 713–726. https://doi.org/10.1111/nous.12144

Stavros, S., & Ioannis, A. (2013). Graph Theoretic Analysis of Knowledge Networks. *Arxiv Preprint Arxiv:1305*(1852). https://doi.org/10.48550/arXiv.1305.1852

Sweller, J. (2024). Cognitive load theory and individual differences. *Learning and Individual Differences*, *110*, 102423. https://doi.org/10.1016/j.lindif.2024.102423

Uriarte Jr, F. A. (2008). *Introduction to knowledge management*.

Vakil, T. S. (2023). *The use of extended reality in maritime education and training: a case study of India* World Maritime University]. Sweden. https://commons.wmu.se/all_dissertations?utm_source=commons.wmu.se%2Fall_dissertations%2F2259&utm_medium=PDF&utm_campaign=PDFCoverPages

Varela, F. J., Thompson, E., & Rosch, E. (2017). *The embodied mind, revised edition: Cognitive science and human experience*. MIT Press.

Verona, G. (1999). A resource-based view of product development. *Academy of Management Review*, *24*(1), 132–142. https://doi.org/10.2307/259041

Wang, J., Shi, R., Zheng, W., Xie, W., Kao, D., & Liang, H.-N. (2023). Effect of frame rate on user experience, performance, and simulator sickness in virtual reality. *Ieee Transactions on Visualization and Computer Graphics*, *29*(5), 2478–2488. https://doi.org/10.1109/tvcg.2023.3247057

Wang, L., & Wang, L. (2024). Relationships between motor skills and academic achievement in school-aged children and adolescents: a systematic review. *Children*, *11*(3), 336. https://doi.org/10.3390/children11030336

Willman, P. (1996). "Wellsprings of Knowledge," by Dorothy Leonard-Barton (Book Review). *Sloan Management Review*, *37*(2), 111. https://nps.idm.oclc.org/login?url=https://www.proquest.com/scholarly-journals/wellsprings-knowledge-building-sustaining-sources/docview/224976272/se-2?accountid=12702

Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*, *4*, 58. https://doi.org/10.3389/fpsyg.2013.00058

Wu, J., He, Y., Liang, S., Liu, Z., Huang, J.,..., & Lee, T. M. (2023). Effects of computerized cognitive training on structure–function coupling and topology of multiple brain networks in people with mild cognitive impairment: a randomized controlled trial. *Alzheimer's Research & Therapy*, *15*(1), 158. https://doi.org/10.1186/s13195-023-01292-9

Zagzebski, L. (2003). The Search for the Source of Epistemic Good. *Metaphilosophy*, *34*(1-2), 12–28. https://doi.org/10.1111/1467-9973.00257

Zhang, C. (2020). *Investigation on motion sickness in virtual reality environment from the perspective of user experience* [Paper presentation]. 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE), Dalian, China. https://doi.org/10.1109/ICISCAE51034.2020.9236907

Zhang, T., Hua, C., Chen, J., He, E., & Wang, H. (2021). Study of Human Tacit Knowledge Based on Electroencephalogram Signal Characteristics. *Frontiers in Neuroscience*, *15*, 690633. https://doi.org/10.3389/fnins.2021.690633

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Fort Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California



DUDLEY KNOX LIBRARY

NAVAL POSTGRADUATE SCHOOL

WWW.NPS.EDU

WHERE SCIENCE MEETS THE ART OF WARFARE