

1-20-2006

## Competitive Usability Studies of Virtual Environments for Shipbuilding

Kurt Satter  
*University of New Orleans*

Follow this and additional works at: <https://scholarworks.uno.edu/td>

---

### Recommended Citation

Satter, Kurt, "Competitive Usability Studies of Virtual Environments for Shipbuilding" (2006). *University of New Orleans Theses and Dissertations*. 331.  
<https://scholarworks.uno.edu/td/331>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by ScholarWorks@UNO with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Dissertation has been accepted for inclusion in University of New Orleans Theses and Dissertations by an authorized administrator of ScholarWorks@UNO. For more information, please contact [scholarworks@uno.edu](mailto:scholarworks@uno.edu).

COMPETITIVE USABILITY STUDIES OF VIRTUAL ENVIRONMENTS FOR  
SHIPBUILDING

A Dissertation

Submitted to the Graduate Faculty of the  
University of New Orleans  
in partial fulfilment of the  
requirements for the degree of

Doctor of Philosophy  
in  
Engineering and Applied Science

by

Kurt M. Satter

B.S. Louisiana Tech University, 1967  
M.S. Southern Methodist University, 1998  
M.S. Southern Methodist University, 1999

December 2005

© 2005, Kurt M. Satter

## Dedication

This dissertation is dedicated to my daughter, Stephanie Anne Wilburn, and my grandsons: Riley Aiden (12/24/03) and Logan Avery (06/06/05).

“The best I’ll ever be.”



## Acknowledgements

I am most grateful to the Office of Naval Research (ONR) for their financial support of the research through GCRMTC via project numbers 327-02-5138 and 27441 and to the Naval Research Laboratory (NRL) for the VR equipment provided to the University of New Orleans through Department of Defense (DoD) Property Disposal. Navy Laboratory personnel at NRL, Washington (Gregory Schmidt), Hydrodynamic/Hydroacoustic Tech Center, NSWC, Carderock (William Smith), and Naval Undersea Weapons Center, Newport (Douglas Maxwell) were helpful, and their cooperation and assistance is acknowledged.

Several University of New Orleans staff and graduate assistants must be credited with much of the CAVE systems installation, technical support, and the seemingly endless task of test administration. These exceedingly helpful individuals include: Edward Williams, Steve Attaya, David Bryan, Carl Lubrano, and Warren Cansello.

I must also recognize Dr. Betsy Pearman of the University of Northern Colorado for her coaching, support, and direction on all things “statistical.” Her assistance was invaluable and sincerely appreciated.

I thank my advisor, research supervisor, and friend, Dr. Alley Butler. Without his help and encouragement little, if any, progress would have been made. I would also be remiss if I did not also thank Dr. Butler’s wife, Kathryn, for her patience and understanding of intrusions into family weekends for project and research discussions.

I also thank the Engineering Management Department Chair, William Lannes, for championing the PhD program and for his support and direction into the administivia of terminal degree pursuit. Without his guidance and support little progress would have been made.

Lastly, I acknowledge the expertise and efforts of my brother Robert “Bob” Lafaye in “large document generation and management” and my daughter, Stephanie Anne Wilburn, Esq., for the endless encouragement and her mantra of: “You can do it Dad!”

# Table of Contents

<b>LIST OF TABLES .....</b>	<b>IX</b>
<b>LIST OF FIGURES .....</b>	<b>XI</b>
<b>ABSTRACT .....</b>	<b>XII</b>
<b>KEYWORDS .....</b>	<b>XII</b>
<b>1 INTRODUCTION.....</b>	<b>1</b>
1.1 VIRTUAL REALITY AS A NEW TECHNOLOGY IN SHIPBUILDING .....	1
1.2 USABILITY ANALYSIS AND USABILITY ENGINEERING .....	3
1.3 RESEARCH QUESTION .....	4
1.4 DISSERTATION ROADMAP .....	5
<b>2 REVIEW OF THE LITERATURE.....</b>	<b>7</b>
2.1 VIRTUAL REALITY.....	7
2.1.1 <i>Interface and System Studies</i> .....	9
2.1.2 <i>Collaborative VR Design Studies</i> .....	12
2.2 EVALUATION METHODS.....	15
2.2.1 <i>Hueristic Evaluation</i> .....	16
2.2.2 <i>Formative Evaluation</i> .....	16
2.2.3 <i>Summative Evaluation</i> .....	16
2.2.4 <i>Usability in MMI Evaluations</i> .....	18
<b>3 COMPETITIVE USABILITY STUDY OF SHIPBUILDING VE'S.....</b>	<b>25</b>
3.1 NEED .....	25
3.2 USABILITY TESTS FOR SYSTEMS AND INTERFACES .....	25
3.2.1 <i>Specify Standard User Configurations</i> .....	25
3.2.2 <i>Usability Tests of Systems and Interfaces</i> .....	26
3.2.3 <i>Identify and Document Basic User Functions as Benchmark Scenarios</i> .....	26
3.3 ENVIRONMENTS/INTERFACES .....	26
3.4 SCENARIOS/PASSES/GROUPS/DATA.....	27
3.5 ROLE OF STATISTICAL ANALYSIS.....	31
3.5.1 <i>Kolmogorov-Smirnov Normality Testing</i> .....	33
3.5.2 <i>Homogeniety of Variance (Homoscedasticity)</i> .....	34
3.5.3 <i>F-Statistic</i> .....	35
3.5.4 <i>Analysis of Variance (ANOVA)</i> .....	36
3.5.5 <i>Reliability - Cronbach's <math>\alpha</math></i> .....	38
<b>4 BENCHMARK 1 (NAVIGATION).....</b>	<b>40</b>
4.1 DESCRIPTION.....	40
4.2 PASS-TO-PASS IMPROVEMENTS IN ELAPSED TIMES.....	40
4.2.1 <i>Novice Users</i> .....	40
4.2.2 <i>Experienced Users</i> .....	41
4.2.3 <i>All Users</i> .....	43
4.3 ELAPSED TIMES DETAILED STATISTICAL ANALYSIS .....	44
4.3.1 <i>Benchmark 1 - Pass 3 Statistics</i> .....	46
4.4 USER SUBJECTIVE OVERALL ENVIRONMENT RATINGS .....	48

4.4.1	<i>Novice Users</i> .....	49
4.4.2	<i>Experienced Users</i> .....	50
4.4.3	<i>All Users</i> .....	51
4.5	USER ENVIRONMENT OVERALL RATINGS STATISTICS.....	52
4.5.1	<i>B1Np3R – Novice User Overall Impressions Statistics</i> .....	53
4.5.2	<i>B1Ep3Ovr – Experienced User Overall Impressions Statistics</i> .....	53
4.5.3	<i>B1Allp3Ovr – All Users Overall Impressions Statistics</i> .....	54
4.6	BENCHMARK 1 RELIABILITY .....	55
<b>5</b>	<b>BENCHMARK 2 (FIND AND REPAIR MANIPULATION) .....</b>	<b>56</b>
5.1	DESCRIPTION.....	56
5.2	PASS-TO-PASS IMPROVEMENTS IN ELAPSED TIMES.....	57
5.2.1	<i>Novice Users</i> .....	57
5.2.2	<i>Experienced Users</i> .....	58
5.2.3	<i>All Users</i> .....	59
5.3	ELAPSED TIMES DETAILED STATISTICAL ANALYSIS .....	60
5.3.1	<i>Benchmark 2 - Pass 3 Statistics</i> .....	60
5.4	USER SUBJECTIVE OVERALL ENVIRONMENT RATINGS .....	63
5.4.1	<i>Novice Users</i> .....	63
5.4.2	<i>Experienced Users</i> .....	64
5.4.3	<i>All Users</i> .....	66
5.5	USER ENVIRONMENT OVERALL RATINGS STATISTICS.....	67
5.5.1	<i>B2Np3Ovr – Novice User Overall Impressions Statistics</i> .....	67
5.5.2	<i>B2Ep3Ovr – Experienced User Overall Impressions Statistics</i> .....	68
5.5.3	<i>B2Allp3Ovr – All Users Overall Impressions Statistics</i> .....	69
5.6	BENCHMARK 2 RELIABILITY .....	69
<b>6</b>	<b>BENCHMARK 3 (SPATIAL AWARENESS).....</b>	<b>71</b>
6.1	DESCRIPTION.....	71
6.2	PASS-TO-PASS PLACEMENT IMPROVEMENTS IN LOCATING ICONS .....	71
6.2.1	<i>Novice Users</i> .....	72
6.2.2	<i>Experienced Users</i> .....	72
6.2.3	<i>All Users</i> .....	73
6.3	SPATIAL AWARENESS DETAILED STATISTICAL ANALYSIS .....	74
6.3.1	<i>Benchmark 3 - Pass 3 Statistics</i> .....	75
6.4	USER SUBJECTIVE OVERALL ENVIRONMENT RATINGS .....	77
6.4.1	<i>Novice Users</i> .....	77
6.4.2	<i>Experienced Users</i> .....	78
6.4.3	<i>All Users</i> .....	78
6.5	USER ENVIRONMENT OVERALL RATINGS STATISTICS.....	79
6.5.1	<i>Pass 3 Statistics</i> .....	80
6.6	BENCHMARK 3 RELIABILITY .....	82
<b>7</b>	<b>COLLABORATION.....</b>	<b>83</b>
7.1	ENHANCED STEREOSCOPIC VOICE/GLOVE ENVIRONMENT .....	83
7.2	COLLABORATIVE BENCHMARK 2 TIMINGS .....	83
7.3	COLLABORATIVE BENCHMARK 2 RATINGS.....	86
7.4	COLLABORATIVE BENCHMARK 3 OFFSETS .....	88
7.5	COLLABORATIVE BENCHMARK 3 RATINGS.....	90
<b>8</b>	<b>BENCHMARK 4 (FAULT IDENTIFICATION) .....</b>	<b>93</b>
8.1	DESCRIPTION.....	93
8.2	BENCHMARK 4 - STATISTICS .....	94
8.3	BENCHMARK 4 - PASS 1.....	94

8.4 BENCHMARK 4 - PASS 2.....	96
8.5 BENCHMARK 4 - PASS 3.....	97
8.6 BENCHMARK 4 - 3 PASS AVERAGE.....	98
8.7 BENCHMARK 4 - ALL TRIALS .....	100
8.8 BENCHMARK 4 FAULT ID TIMINGS .....	100
8.9 ECONOMIC FACTORS .....	101
8.9.1 VR System Capital and Operating Costs .....	101
8.9.2 VR System Cost Savings.....	102
<b>9 CONCLUSIONS AND RECOMMENDATIONS .....</b>	<b>104</b>
9.1 COMPETITIVE ANALYSIS OF ENVIRONMENTS .....	104
9.2 STEREOSCOPIC VS NON-STEEOSCOPIC ENVIRONMENT COMPARISONS .....	105
9.3 SPATIAL AWARENESS.....	106
9.4 COLLABORATIVE STEREOSCOPIC VOICE/GLOVE INTERFACE .....	107
9.5 FAULT IDENTIFICATION.....	107
9.6 COMPETITIVE USABILITY .....	107
9.7 ADDITIONAL TESTING .....	108
9.8 ENHANCEMENTS FOR FURTHER STUDY .....	108
9.8.1 Tracking Map.....	108
9.8.2 Notes/Annotations Log.....	109
9.8.3 CAVE-to-ImmersaDesk and CAVE-to-CAVE Collaboration .....	109
9.9 CLOSURE .....	110
9.9.1 Benchmarks, Measurements, and Ratings .....	110
9.9.2 Statistical Analysis for Proof of Significance .....	111
9.9.3 Summary .....	111
9.9.4 Conclusion .....	112
<b>BIBLIOGRAPHY .....</b>	<b>114</b>
<b>APPENDIX A .....</b>	<b>123</b>
NOVICE USER BENCHMARK 1 (NAVIGATION) DETAIL .....	123
<b>APPENDIX B .....</b>	<b>148</b>
EXPERIENCED USER BENCHMARK 1 (NAVIGATION) DETAIL.....	148
<b>APPENDIX C .....</b>	<b>173</b>
ALL USERS BENCHMARK 1 (NAVIGATION) DETAIL.....	173
<b>APPENDIX D .....</b>	<b>198</b>
NOVICE USER BENCHMARK 2 (FIND/REPAIR) DETAIL .....	198
<b>APPENDIX E.....</b>	<b>223</b>
EXPERIENCED USER BENCHMARK 2 (FIND/REPAIR) DETAIL .....	223
<b>APPENDIX F.....</b>	<b>248</b>
ALL USERS BENCHMARK 2 (FIND/REPAIR) DETAIL.....	248
<b>APPENDIX G .....</b>	<b>273</b>
NOVICE USER BENCHMARK 3 (SPATIAL AWARENESS) DETAIL.....	273
<b>APPENDIX H .....</b>	<b>306</b>
EXPERIENCED USER BENCHMARK 3 (SPATIAL AWARENESS) DETAIL.....	306
<b>APPENDIX I.....</b>	<b>339</b>

ALL USERS BENCHMARK 3 (SPATIAL AWARENESS) DETAIL .....	339
<b>APPENDIX J .....</b>	<b>372</b>
BENCHMARK 1 STATISTICALLY SIGNIFICANT ENVIRONMENT COMPARISONS .....	372
<b>APPENDIX K .....</b>	<b>376</b>
BENCHMARK 2 STATISTICALLY SIGNIFICANT ENVIRONMENT COMPARISONS .....	376
<b>APPENDIX L.....</b>	<b>380</b>
BENCHMARK 3 STATISTICALLY SIGNIFICANT ENVIRONMENT COMPARISONS .....	380
<b>APPENDIX M .....</b>	<b>384</b>
CRONBACH’S ALPHA RELIABILITY CALCULATIONS DETAIL .....	384
<b>APPENDIX N .....</b>	<b>391</b>
HUMAN SUBJECTS PROTECTION CERTIFICATION.....	391
<b>VITA.....</b>	<b>393</b>

## List of Tables

Table 1 – Comparisons of Formative and Summative Evaluation [82] .....	17
Table 2 – Usability Hueristics in MMI Testing [79;86] .....	19
Table 3 – Sample Usability Specification Table [88].....	21
Table 4 – B1N Pass-to-Pass Elapsed Time Changes.....	41
Table 5 – B1E Pass-to-Pass Elapsed Time Changes .....	42
Table 6 – B1All Pass-to-Pass Elapsed Time Changes.....	44
Table 7 – B1Np3Tstat - Pass 3 Novice User Timing Statistics.....	46
Table 8 – B1Ep3Tstat - Pass 3 Experienced User Timing Statistics.....	47
Table 9 – B1Allp3Tstat - Pass 3 All Users Timing Statistics .....	48
Table 10 – B1N – Novice User Pass-to-Pass Overall Impression Ratings Changes.....	50
Table 11 – B1E – Experienced User Pass-to-Pass Overall Impression Ratings Changes ....	51
Table 12 – B1All – Pass-to-Pass Overall Impression Ratings Changes.....	52
Table 13 – B1Np3Ovr – Novice Users Overall Impressions Statistics .....	53
Table 14 – B1Ep3Ovr – Experienced Users Overall Impression Statistics.....	54
Table 15 – B1Allp3Ovr – All Users Overall Impressions Statistics .....	55
Table 16 – Benchmark 1 Cronbach’s $\alpha$ Reliability Coefficients.....	55
Table 17 – B2N – Novice User Pass-to-Pass Elapsed Time Changes.....	58
Table 18 – B2E – Experienced User Pass-to-Pass Elapsed Time Changes.....	59
Table 19 – B2A – All Users Pass-to-Pass Elapsed Time Changes.....	60
Table 20 – B2Np3Tstat – Novice User Pass 3 Novice Users - Timing Statistics .....	61
Table 21 – B2Ep3Tstat – Experienced Users Pass 3 Experienced Users Timing Statistics.	62
Table 22 – B2Allp3Tstat – All Users Pass 3 All Users Timing Statistics.....	62
Table 23 – B2N – Novice Users Pass-to-Pass Overall Impression Ratings Changes .....	64
Table 24 – B2E – Experienced User Pass-to-Pass Overall Impression Ratings Changes ....	65
Table 25 – B2All – All Users Pass-to-Pass Overall Impression Ratings Changes.....	67
Table 26 – B2Np3Ovr – Novice Users Overall Impressions Statistics .....	67
Table 27 – B2Ep3Ovr – Experienced Users Overall Impressions Statistics .....	68
Table 28 – B2Allp3Ovr – All Users Overall Impressions Statistics .....	69
Table 29 – Benchmark 2 Cronbach’s $\alpha$ Reliability Coefficients.....	70
Table 30 – B3Np3-TOFF – Novice User Pass 3 – Total Offsets Statistics.....	75
Table 31 – B3Ep3-TOFF – Experienced Users Pass 3 – Total Offsets Statistics .....	76
Table 32 – B3Allp3-TOFF – All Users Pass 3 – Total Offsets Statistics .....	77
Table 33 – B3Np3Ovr – Novice User Pass 3 Overall Impressions Statistics.....	80
Table 34 – B3Ep3Ovr – Experienced Users Pass 3 Overall Impressions Statistics .....	81
Table 35 – B3Allp3Ovr – All Users Pass 3 Overall Impressions Statistics .....	82
Table 36 – Benchmark 3 Cronbach’s $\alpha$ Reliability Coefficients .....	82
Table 37 – Benchmark 2 Timings Comparisons - Individual vs Collaborative .....	85
Table 38 – Collaborative Benchmark 2 Overall Impressions Statistics .....	87
Table 39 – Benchmark 3 Offsets Comparisons – Individuals vs Collaborative.....	89
Table 40 – Collaborative Benchmark 3 Overall Impressions Statistics .....	91
Table 41 – Benchmark 4 Pass1 Fault Counts Statistics.....	95
Table 42 – Benchmark 4 Pass 2 Fault Counts Statistics.....	97
Table 43 – Benchmark 4 Pass 3 Fault Counts Statistics.....	98
Table 44 – Benchmark 4 – 3 Pass Average Fault Counts Statistics .....	99

Table 45 – Benchmark 4 – All Trials Fault Counts Statistics .....	100
Table 46 – Benchmark 4 Elapsed Fault ID Elapsed Times – 86” Non-Stereo Interface ....	101
Table 47 – Benchmark 4 Fault ID Elapsed Times – CAVE Environment .....	101
Table 48 – Prototype Ship Cost Savings Using Virtual Environments .....	103
Table 49 – ANOVA Pair-wise Environment Comparisons Summary .....	104

## List of Figures

Figure 1 – Designer Centered Design and Evaluation Processes [78] .....	15
Figure 2 – Test/Data Dimensions .....	28
Figure 3 – Benchmark Test Groupings and Layers .....	28
Figure 4 – Usability Questionnaire .....	29
Figure 5 –Test/Data/File Naming Conventions .....	31
Figure 6 – Statistical Analysis Test Selection .....	33
Figure 7 – B1Np3Tim – Novice User Pass 3 Navigation Times.....	41
Figure 8 – B1Ep3Tim - Experienced User Pass 3 Navigation Times .....	42
Figure 9 – B1Allp3Tim - All User Pass 3 Navigation Times.....	43
Figure 10 – B1Np3Ovr – Novice User Pass 3 Overall Impression Ratings .....	49
Figure 11 – B1Ep3Ovr – Experienced User Pass 3 Overall Impression Ratings .....	50
Figure 12 – B1Allp3Ovr – All Users Pass 3 Overall Impression Ratings.....	52
Figure 13 – B2Np3Tim – Novice User Pass 3 Manipulation Times.....	57
Figure 14 – B2Ep3Tim - Experienced User Pass 3 Manipulation Times.....	58
Figure 15 – B2Allp3Tim - All User Pass 3 Manipulation Times.....	59
Figure 16 – B2Np3Ovr – Novice User Pass 3 Overall Impression Ratings .....	64
Figure 17 – B2Ep3Ovr – Experienced User Pass 3 Overall Impression Ratings .....	65
Figure 18 – B2Allp3Ovr – All Users Pass 3 Overall Impression Ratings.....	66
Figure 19 – B3Np3-Toff – Pass 3 Novice User Total Placement Offset.....	72
Figure 20 – B3Ep3-Toff – Pass 3 Experienced User Total Icon Placement Offsets .....	73
Figure 21 – B3Allp3-Toff – Pass 3 All Users Total Icon Placement Offsets.....	74
Figure 22 – B3Np3Ovr – Novice User Pass 3 Overall Impressions Ratings.....	77
Figure 23 – B3Ep3Ovr – Experienced User Pass 3 Overall Impressions Ratings.....	78
Figure 24 – B3Allp3Ovr – All Users Pass3 Overall Impressions Ratings+.....	79
Figure 25 – Collaborative Benchmark 2 Find/Repair Times by Group.....	84
Figure 26 – Comparison of Individual vs Collaborative Find/Repair Times .....	85
Figure 27 – Collaborative Benchmark 2 Group Overall Impressions Ratings .....	86
Figure 28 – Comparison of Individual vs Collaborative Find/Repair Overall Ratings .....	87
Figure 29 – Collaborative Benchmark 3 Spatial Awareness Offsets by Group.....	88
Figure 30 – Comparison of Individual vs Collaborative Spatial Awareness Offsets .....	90
Figure 31 – Collaboration Benchmark 3 Group Overall Impressions Ratings.....	91
Figure 32 – Comparison of Individual vs Collaborative Spatial Awareness Overall Impressions Ratings.....	92
Figure 33 – Benchmark 4 Pass 1 Fault Counts.....	95
Figure 34 – Benchmark 4 Pass 2 Fault Counts.....	96
Figure 35 – Benchmark 4 Pass 3 Fault Counts.....	98
Figure 36 – Benchmark 4 – 3 Pass Average Fault Counts .....	99
Figure 37 – Potential LPD Prototype Rework Savings .....	103
Figure 38 – Stereoscopic vs Non-Stereoscopic Environment Times Comparisons .....	105
Figure 39 – Stereoscopic vs Non-Stereoscopic Environment Distance Offsets Comparisons .....	106
Figure 40 – Stereoscopic vs Non-Stereoscopic Environment Overall Ratings Comparisons .....	106



## **Abstract**

Establishing usability specifications as measurable attributes in repeatable scenarios has been an essential task in the management and continuous improvement processes [1]. Early studies in Usability Analysis were primarily conducted to assist software developers and hardware designers in improving the Human-Computer Interface (HCI) or Man-Machine Interface (MMI). However, this study was conducted to provide comparative data supporting broad conclusions regarding the comparative merits of one technology (non-stereoscopic, conventional CAD systems) competed against another (tracked, stereoscopic virtual environments). Competing environments to establish usability features and preferences provides a new tool to the interface designer. Benchmark scenarios were designed and executed to measure navigation, fault identification/repair, and spatial awareness through a sequence of choices and to provide user preference of one GUI paradigm over another functionally similar paradigm. This study, performed on a ship design application, included an analysis of the effects of user collaboration in virtual environments.

## **Keywords**

Design collaboration, Usability Analysis, HCI, Human-Computer Interface, Man-Machine Interface, Spatial Awareness, VE's, Virtual Environments, Virtual Reality, VR

# 1 Introduction

## *1.1 Virtual Reality as a New Technology in Shipbuilding*

The use of virtual systems in manufacturing and design has dramatically assisted in the reduction of cycle time and improvements in design quality. Unfortunately, implementation and transfer of the virtual technologies into shipbuilding has seriously lagged behind other industries, especially aerospace and automotive.

In the past, physical mock-ups were developed for critical or significantly dense portions of a ship. Such mock-ups were expensive and have since been replaced by virtual mock-ups based on only an electronically produced representation. This step to virtual prototyping has been incorporated in products ranging from the Boeing 777 aircraft to many of the ships now produced at major shipyards. However, experimentation by even the most advanced shipyards in the use of virtual reality techniques with real time, tracked, stereopsis has been significantly limited. Current efforts at Boeing and in the automotive industries include virtual reality and augmented reality to improve design quality and cycle time. Boeing reports a 50% reduction in design cycle time with virtual systems [2]. NIST (National Institute of Standards and Technology) reports projects with successful implementations in manufacturing that involve widely differing industrial efforts [3]. The uses of virtual systems for manufacturing have been shown to improve factory operations and reduce costs. Dramatic improvements in cycle time, manufacturing effectiveness, design quality, and design producibility are possible with properly constructed virtual environments.

The focus of this project is to verify the benefits of using immersive environments demonstrated in other industries and to design a platform for operational testing and training programs within the shipbuilding industry. Once completed, the project compiles information on implementation and utilization issues with an emphasis on comparative evaluation.

Improvement of Virtual Environments (VE's) for shipbuilding involve's experimentation with virtual environments and interface paradigms to determine and quantify the advantages and limitations of each environment and user interface paradigms.

This work builds on the experience of Maxwell, King, and Butler [4] in the development of virtual environments for ship design and expands upon developments and achievements in Douglas Maxwell's 2001 MSME thesis at Louisiana Tech University [5]. In Maxwell's work, only one environment and one interface method was developed for a single user. However, shipbuilding is often a collaborative process in which there are multiple design and naval construction experts, each of whom must provide important technical input for the design process. As an important expansion of the 2001 work, this research project focuses on development and enhancement of virtual environments to support multiple users. Each environment and interface paradigm has advantages and limitations that need to be examined. To ensure optimal performance, a competitive study is needed. In response to these needs, this project completes both research and development activities designed to answer the technical questions of virtual environment use and to examine stereoscopic environments with collaboration among multiple technical experts.

In shipbuilding, naval architects and engineers produce a design that involves construction of 3D products. The ship as a product is considered the largest and most complex product routinely produced today. The complexity of the product requires coordination among technical experts in structural design, piping design, ventilation system (HVAC) design, electrical design, human factors, marine engineering, production engineering, and other technical specialties. The design of one technical specialist influences the design for another technical specialist, and the various technical specialists must collaborate to ensure that the ship performs as desired. Although in the past this collaboration took place using a drawing as a representation of a portion of the product, with today's technology representation of the product is electronic, and the collaboration occurs through an electronic representation of the product. The objective of this research is to quantify the benefits of state-of-the-art, user friendly, real-time, stereoscopic, virtual environment to facilitate collaboration among technical specialists in shipbuilding.

Virtual environments produce real time, 3D representations of geometry using tracked stereoscopic vision. In conventional CAD, the user perceives and understands the design through a two dimensional interface, the workstation screen or other planar surface. User interaction with the product representation is limited by the two dimensional nature of the user interface. In contrast, with a virtual environment the users experience a "like real," three-dimensional representation of the geometry. The users can interact with the product

design by viewing it from any angle or perspective. The representation in a three dimensional virtual environment gives a designer a chance to experience the design before it is constructed in a way not possible with a planar image.

## ***1.2 Usability Analysis and Usability Engineering***

Methodical approaches to Usability Analysis and Usability Engineering (a discipline that provides structured methods for achieving usability in interface design during product development, [6]) have been evolving since the early 1970's. However, most investigations stem from the work of Gould and Lewis [7] with their discussion of three global strategies:

- 1) early focus on users and tasks,
- 2) empirical measurement, and
- 3) iterative software design.

Usability analysis and usability engineering techniques are gaining such wide acceptance and use that the International Standards Organization (ISO) has released the ISO 9241-11 Guidance on Usability standard [8] to aid practitioners in its use.

Many methods are utilized during the course of development and implementation projects to evaluate suitability of the HCI (Human-Computer Interface) or MMI (Man-Machine Interface). In addition to the work on industrial usability for the Naval Research Laboratory (NRL), Hix et al. have developed a battlefield visualization system using a User-Centered Design and Evaluation process [9]. This process involves improving the user interface for a tactical display on a responsive workbench. This battlefield visualization system includes an extensive effort for user interface design. As an extension of that effort, Gabbard, Hix, and Swan [10] found that "comparatively little effort has gone into user interaction components of VE's. The user interaction components of VE applications are often poorly designed and rarely evaluated with users." Techniques for evaluating the usability of an immersive medical VE were reported by Gabbard et al. [11] in 1999. This study centered on the use of eye-tracking as a mechanism for monitoring the HCI.

As an initial step to development of an effective virtual environment, Gabbard, Hix, and Swan [10], Nielsen [12], and Poupyrev, et al. [13], advocate a user task analysis. This involves the identification and complete description of tasks, subtasks, and methods

required for the use of a system. A top down decomposition on a detailed basis is needed. Failure to accomplish a task analysis can result in poor user interaction design, and for this reason designer task analysis is accomplished with potential system users very early in the development process.

In 2002, Thomas and Macredie [14] coined the term “New Usability” to include the extended usability required for emerging products and systems being provided to “digital consumers.” Expansion of usability analysis and usability engineering techniques to quantify CAD interfaces in distributed collaborative environments is an important step in production of improved, economically viable manufactured products.

### ***1.3 Research Question***

In 1965 Ivan Sutherland [15] stated the goal for research in computer graphics: to immerse the user in virtual worlds that look real, sound real, feel real, and behave properly as the user interacts with them. This simple statement has driven computer graphics research ever since.

In his 1994 public lecture cosponsored by the Royal Academy of Engineering and the British Computer Society in London [16], then restated and updated in his 1999 IEEE VR Conference Keynote address [17], F. P. Brooks poses two scientific questions that command the attention of the serious researcher in virtual worlds:

1. Can we make systems that will give the realistic simulated experience described in Sutherland's challenge?
2. If we can, what worthy tasks can the user of such tools accomplish that cannot be accomplished as well without them?

On the first question, immense strides have been made over the past 35 years, and systems development seems to be within striking distance of the goal. Or, to put it another way, the virtual environments technology now works, but barely. Today's research can be summarized under four categories that make the virtual environments systems:

- more realistic,
- faster,
- handier, and

- more faithful in model accuracy.

With the use of virtual environments in shipbuilding, usability studies of multiple interfaces and environments can provide valuable measures upon which to statistically compare the viability of such environments in the collaborative engineering design and review processes in shipbuilding applications. The research question becomes: **Are the new, more complex and expensive (compared to traditional, non-stereoscopic CAD) systems worth the additional costs?**

## ***1.4 Dissertation Roadmap***

The remainder of this document discusses this existing research and literature on the use of virtual reality in design environments including background, needs, hardware, software, interfaces, and applications. Usability analysis and engineering techniques are discussed.

Following the literature and research review the document describes the research environments, measuring instruments (Benchmark scenarios), measures, and analysis techniques used to prepare the results presented in the document. Detailed results of 3-pass average and final pass user objective and subjective measures of four user interface environments (two non-stereoscopic and two Virtual Reality or VR), including statistical analysis of the results, are presented.

The focus of this project is to:

1. identify the environment and interfaces to be compared
2. define the usability attributes (e.g., navigation)
3. develop the measuring instruments (scenarios and Benchmarks)
4. define the values to be measured
5. identify the user groups
6. execute the tests and collect the data
7. analyze and report the results.

The document presents overall conclusions from the analysis of the study results and presents topics for additional study. Several appendices are included to present the detailed

(pass-to-pass) Benchmark results for each pass of each Benchmark for each user group (novice and experienced).

## **2 Review of the Literature**

### ***2.1 Virtual Reality***

In 1962 Ivan Sutherland developed a light pen with which images could be sketched on a computer. Sutherland's first computer-aided design program, called Sketchpad, opened the way for designers to use computers to create blueprints of automobiles, cities, and industrial products. By the end of the decade, the designs were operating in real time [18]. In an invited lecture to the 1965 IFIP Congress entitled “The Ultimate Display,” Sutherland [15] was among the first researchers to discuss the possibilities of 3D systems in design environments including a head-mounted display and the use of the 1<sup>st</sup> crude pointing device (computer mouse, patented in 1970) developed by Douglas Engelbart [18] in 1964.

To create the illusion of depth, VR software is constantly calculating the two views of its virtual world to correspond to the way the eyes see the same scene from slightly different angles. The two images, each from a slightly different perspective, are sent to viewing screen(s) in a synchronized fashion. Between the screen(s) and the eyes are shutter glasses; polarizing filters, that allow the left-eye/right-eye views to be seen by alternating the polarization to allow only one view to reach the left or right eye [19]. Shutter glasses alternate viewing between the left/right eye 120 times per second. Synchronization is maintained by an infrared emitter that keeps the computer generated image and the left or right eye of the shutter glasses operating correctly. This shuttering effect happens too quickly for the user to note the change (persistence) and the viewer’s brain processes the two visual inputs as depth.

Interaction with a three dimensional virtual environment can be accomplished using three dimensional wands or flightsticks, using data gloves that can grasp and move part of the design, or with voice commands. Early experimentation in 1998 by Chu, Dani, and Gadh [20] with these techniques shows that by using a virtual world for design processes, speed up in design is accomplished by a factor of five or ten. The research by Chu et al. focused on design synthesis at the conceptual level with wands, gloves, and voice. It did not focus on collaborative processes for a multiple person design review or collaborative sessions. Further, although clearly seminal in nature, the work of Chu et al. stopped



significantly short of a correctly executed usability analysis. The focus of this work includes the type of usability analysis for virtual environments, pioneered by Gabbard, Hix, and Swan [10] at the Naval Research Laboratory (NRL). The results are provably superior methods for the user to interact with the virtual environment.

The data glove is one means of control within the virtual space. At the end of each finger, a light emitting diode (LED) shines light through optical fibers woven into the glove's material. These fibers carry the light up the fingers to phototransistors at the base of each finger. As a finger bends, it compresses the optical fiber so that less light passes through it. The phototransistors constantly measure the varying light intensities and send that information to the VR software.

The wand and data glove each have a sensor to track the movement of the hand through six degrees of freedom (6DOF, x-axis, y-axis, z-axis, roll, pitch, and yaw). The software uses the data from the wand or glove's 6DOF sensor and fiber optics to calculate the position and orientation of either the wand or hand and fingers. Then, the software modifies the screen's display of the user (hand or arrow) to match the position of the glove or wand [21].

Virtual Environments (VE's) offer new possibilities and challenges to HCI design, but have been noted for being significantly more difficult to design and use than traditional CAD interfaces [22]. In 1996 studies by Kaur et al. [23], they showed that designers lacked a coherent approach to interaction design and appeared to be preoccupied with difficult technical issues and thought little about supporting user interaction. This concept was further reinforced by the COVEN Project [24] as part of the European ACTS program addressing generic requirements for Collaborative Virtual Environments (CVE's) to include such problems as disorientation, perceptual misjudgements, and difficulty finding available interactions. These problems were shown to result in user frustration during early training sessions dissipating with further exposure to the environment [23;25].

In a 1998 report on the 21<sup>st</sup> century warship, Baum, Boudreaux, Bourassa, and Jenkins [26] discussed the viability of 3D structural models using the Deneb 3D-visualization tool which allows the designer to take a "virtual walkthrough" of the 3D model. They stressed how the ability to detect and correct any errors before construction begins resulted in lower construction costs and shortened construction schedules. This paper and the ship construction programs to which it related stopped short of the current

research because stereoscopic environments and wand/glove gestural interfaces were not used.

### **2.1.1 Interface and System Studies**

Many studies have been performed to investigate Human-Computer Interface (HCI) actions in both traditional CAD and design environments. As early as 1990, researchers at Matsushita Electric Works, Ltd. were investigating VR for use in industrial applications at Matsushita Electric Works, Ltd. In 1999 Nomura and Sawada [27] reported on these applications including the Virtual Space Support System developed in conjunction with Kansei Engineering.

In an early VR study (1993) by Ware et al. [28], head tracking in the desktop environment was shown to improve the user's ability to understand complex 3D structures. The study was primarily performed to demonstrate the use of normal mid-range workstations, thereby, enabling a wide range of affordable 3D systems. As early as 1992, Bolt and Herranz [29], discussed two-handed gestural interfaces and in a 1995 ACM paper, Wexelblat [30] described a gesture analyzer to capture natural, empty-handed gestural commands that may be translated into any appropriate action in the VR space.

An early CAVE (Cave Automatic Virtual Environment) was first reported as an implementation of VR technology by a PhD student, Carolina Cruz-Niera. The 1992 Cruz-Niera et al. [31] work first described the application of mirrors and folded-optics for immersive display as opposed to the traditional HMD (head-mounted displays) previously used in VR work.

In their 1999 study, Gabbard et al. [11] used a medical visualization testbed for evaluating eye tracking as a mechanism for monitoring user activities in fully immersive virtual environments. The study centered on the use of eye-tracking as a mechanism for monitoring human activity in VE's, and more specifically, as an instrument to facilitate formative usability evaluation of fully immersive VE interfaces.

Several testbed and developmental study systems were first reported in 1997. The need for both navigation and control (manipulation) for both mouse-based and 3D input devices was discussed by Hand [32]. VRMAT (Virtual Reality Manipulation Assessment Testbed), a test system for studies of measurement techniques in virtual environments by

Poupyrev et al. [13], stressed a systematic task analysis of immersive object manipulation techniques and user spatial awareness.

In their 1997 report to the ACM, Stork and Maidhoff [33] discussed the uses of 3D input devices in virtual environments to do sketching and the need to introduce a new paradigm to bridge the gap between sketching and detailed 3D design. Feature-based CAD in virtual environments was addressed from the viewpoint of process-planners by Trika et al. [34]. Their method of feature specification implicitly enforces feature accessibility constraints, and by including navigation and collision detection, provides a possible method for implying the order in which features may be manufactured.

In other 1997 studies, Tushar et al. [20;35] reported on the COVIRDS (Conceptual VIRtual Design System) which stressed the use of hand-tracking devices and voice commands in VR design environments. This work demonstrated the relative efficiency of the interfaces in specifying shapes and dimensions in product design.

In 1998, Jayaram et al. [36] reported on another VR design environment, VADE (Virtual Assembly Design Environment), that focuses on using virtual interfaces in commercial CAD systems. VADE's features included realistic user interaction with parts within the design space, collision detection in real-world engineering models. The VADE system was based on an earlier study (1997) by Jayaram et al. [37].

A call for more studies in the use of VR and augmented reality technology in product design and realization was made in 1999 by Lu et al. [38]. Specifically the authors called for new paradigms to efficiently deliver new products to society across time and space. In their 1999 work, Bowman et al. [39] describe testbed methods for evaluation of common VE tasks including test results that provide an empirical basis for choosing interaction techniques in VE applications that produce measurable usability gains.

Evans, Vance, and Dark [40] reported on a 1999 study in which users rated their ability to interact with design images using a traditional CAD and a VE interface; their preferences; and their subjective feelings about the interfaces. The report stresses that participants preferred a traditional interface for interaction tasks and a VR interface for visual tasks.

A detailed description and analysis of voice/glove VR controls was the basis of the 1999 Brown University, Computer Science Master's thesis of Joseph LaViola [21]. His

research and thesis examines the two modalities (hand and voice) individually and in combination. He also describes two application prototypes (Multimodal Scientific Visualization Tool and Room Designer) which were used as a basis for the Voice/Glove interface presented in this study. In 2000, Rosenfield et al. [41], also reported encouraging results for HCI speech-based interfaces which were further refined in a 2002 implementation of speech recognition in VR presented by Dorkjkinw and Vance [42].

SeamlessDesign, a flexible, collaborative, virtual workspace for rapid prototyping, was reported by Kiyokawa et al. [43]. Their 2000 study utilized a shared VR environment stressing parallel, collaborative activities in the design space. Another design VR study in 2000 by Kuester et al. [44], documents DesignersWorkbench, using a two-handed virtual interface allowing “collaborative development in a semi-immersive” virtual environment. Also in 2000 Omata et al. [45] discussed the use of gestural-based interfaces for international communications between real and virtual environments. The Virtual Round Table is another collaborative, augmented, multi-user, VE that was reported in 2000 to the ACM by Broll and Schardt [46]. The Virtual Round Table was designed to support location-independent mixed reality applications and preserve traditional verbal and non-verbal (gestural) communications mechanisms.

The 2001 literature also shows much interest in adding haptic feedback to VR design systems. Vance and Volkov [47] report on a study of the effects of adding a haptic device to a VR design environment using 76 participants. Their results indicate that the use of haptic devices resulted in faster decisions, but the group did not make more accurate or precise evaluations. VE remote collaboration techniques using visual (head-mounted), aural, and haptic interfaces (tele-presence) were reported for a laptop PC docking station design and for medical robotics by Ansar et al. [48].

CU-Ergo, the Cornell University Ergonomics Web [49], is an excellent source of current information of the ergonomics of VR systems. The site presents information from research studies and class work by students and faculty in the Cornell Human Factors and Ergonomics Research Group (CHFERG), directed by Professor Alan Hedge of the Department of Design and Environmental Analysis at Cornell University. CHFERG focuses on ways to enhance usability by improving the ergonomic design of hardware, software, and workplaces, to enhance user comfort, performance and health in an approach dubbed Ergotecture.

### 2.1.2 Collaborative VR Design Studies

In 1997 Connor et al. [50] presented TeamCAD as one of the first uses of a Wide-Area Network (WAN) as an immersive tool for collaborative design. Later studies (Sky and Buchal [51], Kan et al. [52], and Woo et al. [53;54]) discuss the use of the Internet as the mechanism for distributing the collaborative design tasks. Also in 1997, Lea et al. [55] demonstrated the Wide Area Virtual Environment (WAVE) system collaborative Virtual Environment (CVE) based on the Internet. The globalization of VR design environments based on an internet communications infrastructure was discussed in a presentation by Horvath et al. [56] before the 2002 International Design Conference in Dubrovnik. The presentation stressed that the internet-based CAD/E system they examined showed poor performance compared to conventional standalone systems. They concluded that new infrastructure, methods and knowledge are needed in the form of collaborative virtual design environments (CVDE's).

The ergonomics of VR collaborative techniques with one- and two-person teams in a virtual environment to move large objects through a cluttered environment was reported by Gill and Ruddle [57] in 1998. Topics investigated by the project include real-world vs. VE object rotation, movement interfaces for cluttered VE's, methods for providing body force feedback when handling "heavy" virtual objects, and one- and two-person versions of the piano movers problem (moving a large object through a cluttered environment).

In 2003, Anderson et al. [58] demonstrated a VE DesignStation providing a virtual environment application for conceptual design for architectural projects based on a collaborative 3D design process that included not only imagery and a critiquing process as well, but also included the use of a toolbox and the ability to work in more than one scale simultaneously.

Blue-C, a telecollaboration system including 3D representation of objects with a gestural interface was described by Disz et al. [59] in a 1995 international workshop on graphics and applications. This system was elaborated upon as an application of the telecollaborative network at the ETHZ (the Swiss Federal Institute of Technology) in Zurich in a presentation by Spagno and Kunz [60] at the 2003 IEEE Conference on Virtual Reality. In 1997, Lehner and DeFaranti [61] reported on CAVERN, the CAVE Research Network; an alliance of industrial and research institutions equipped with CAVE-based virtual reality

hardware and high-performance computing resources, interconnected by high-speed networks, to support collaboration in design, education, engineering, and scientific visualization. This network is being developed to support collaborative vehicle design, engineering, training, and visualization techniques over distance. Also in 1997, A. Rowell [62] presented a sound case for the benefits of group (collaborative) VR, while Tromp and Snowden [63] reported on the use of a VR, conferencing system (MASSIVE) to verify the use of networking to aid collaborative design decisions. Good results were reported by Anumba and Duke [64] for a 1997 CICC (Collaborative Integrated Communications for Construction) project, a European pilot project for evaluating the use of the Internet and Intranet as an infrastructure for VR project teams in the construction area.

A Bochenek et al. [65] 2001 study of collaborative use of virtual 3D display systems for US Army design review purposes showed that teams detected design errors faster when using VE interfaces and that subjective perceptions and preferences data analysis indicated that stereoscopic systems were preferred. Also in 2001, Vance and Yeh [66] reported on a study of engineering design sensitivity and optimization in VR. The study centered on the use of the environment to investigate multiple design changes while viewing and manipulating virtual objects and demonstrated “better designs achieved in a timely manner.”

An ACM Conference on Collaborative Virtual Environments paper, presented in 2000 by Fraser et al. [67], discusses many issues in the use of collaboration virtual environments with an emphasis on the problems inherent in the process rather than the actual systems interface itself. At another 2000 ACM VR conference, E. Swing [68] discussed the collaborative capabilities of the Collaborative Virtual Workspace and efforts to augment the system with an immersive display capability rather than adding collaboration to an existing immersive VE. In another design collaboration study reported in 2000 by Kolarevic et al. [69], the Virtual Design Studio (VDS) environment was used by design-team members at three academic institutions. The study showed that team members could successfully work on a building design at any place (over distance), simultaneously (synchronously), or separately (asynchronously), while the latest state of the design would always be available in a shared database. In 2001 Tay and Ming [70] also used the VDS to study VR concurrent engineering techniques across the Internet. Also in 2000, Bergman and Baker [71] reported on the issues of enabling technologies for distributed collaboration

in virtual design environments at the Jet Propulsion Laboratory. The report identified the types of data frequently exchanged to be: project planning data, design data, notes, documentation, communications data, analysis/performance data, verification data, and scientific data. They suggest that existing MIS GroupWare packages can form a basis for this sharing of scientific data.

Campbell and Wells [72] describe the role of VR in schematic design, through design development to presentation and evaluation. Their 2003 study comments on the effects of VR on detailed design in architectural environments and states that VR proved to be advantageous in several phases of the design. They stress the need for designer toolkits specific to the environment and call for closer collaboration techniques including multiple markup capabilities with software enhanced to evaluate alternative designs.

In 2003 Chipperfield and Vance [73] presented to the ASME a unique VR application, VRHose. The application allowed designers to use VR techniques to design hydraulic hoses including routing designs within hose applications by defining points in space which constrain the hose path. VR is used as the HCI to provide 3D viewing and interaction with digital models. Future versions are to include hose material properties which can constrain hose paths to those which are feasible.

An assessment of VR techniques and systems applied to product realization across the product development lifecycle is discussed by Jayaram et al. [74]. The assessment discusses methods beyond virtual design including visualization of virtual manufacturing and virtual assembly processes including engineering analysis and visualization of analysis of results in collaborative environments.

At the 1996 ASME conference on computers in design engineering, Gaisemier et al. [75], discussed the integration of VR design, modeling, manufacturing, and process engineering, and in 1997 Gupta, Whitney, and Zelter [76] reported studies of design prototyping and analysis in virtual environments. In other conference activity, a 2002 ACM conference paper by Thomas and Macredie [14] discussed the “new usability” as a usability engineering technique for evaluating emerging products and systems for digital customers.

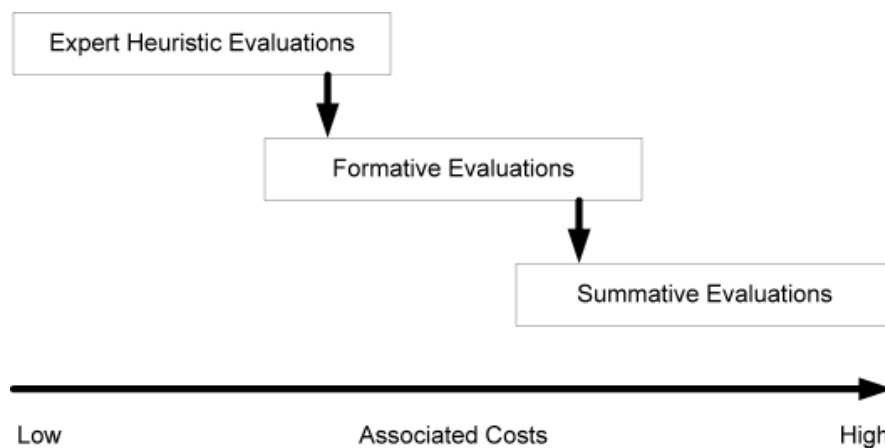
## 2.2 Evaluation Methods

Gray and Boehm-Davis [77] discuss cognitive task analysis as a means of providing descriptions of the declarative knowledge possessed by domain experts. Their work specifically details a family of cognitive process analysis techniques (GOMS) that help describe activities that occur in parallel, eg., multiple user movement through a design space.

The methods used at NRL and advocated by Hix et al. [9] include a three-step evaluation process. This three-step process involves:

- 1) heuristic evaluation,
- 2) formative evaluation, and
- 3) summative evaluation.

Each of the usability evaluations provides input to the next stage of evaluation, and each evaluation technique has a successively higher evaluation cost, as shown in Figure 1. This research project uses these successive evaluation types because of the successful prior implementation at NRL, thereby avoiding the pitfalls of many virtual systems described by Gabbard, Hix, and Swan, [10] where, “many visually compelling VE’s are difficult to use and thus unproductive.”



**Figure 1 – Designer Centered Design and Evaluation Processes [78]**



### **2.2.1 Hueristic Evaluation**

As the first process, an independent expert or experts undertake heuristic evaluation of the user interface. The interface is examined in a two-pass approach [79]. For the two-pass evaluation each expert first gains a general understanding of the flow of interaction and then repeats the review process to identify specific interaction components and conflict [80]. Further guidelines are available specifically for virtual environments [14] including 195 guidelines covering virtual environments.

### **2.2.2 Formative Evaluation**

The second evaluation process is formative evaluation in which users are employed to evaluate the virtual environment interface [81]. There is a usability specialist to proctor the process in which users perform tasks as evaluators collect data. The formative evaluation includes five steps that are conducted iteratively. These steps include:

- development of task scenarios,
- representative users perform the scenarios,
- evaluators collect data,
- VE designers and evaluators suggest improvements,
- VE designers and evaluators refine task scenarios.

Typically, critical incidents occur in which quantitative and qualitative data is developed. The quantitative data shows that a problem occurred, and the qualitative data tends to indicate where the problem occurred.

### **2.2.3 Summative Evaluation**

As the most expensive form of evaluation, summative evaluation is used to statistically compare final forms of VE design, typically after completion of formative evaluation. Scenarios developed during the formative evaluation process are refined for use in evaluating final virtual environment interfaces through summative evaluation. The

results are a specific and quantitative answer to questions regarding which interface performs better [9].

In their textbook on evaluation and measurement, Sarvela and McDermott [82], compare the purposes and issues concerning the various stages of the evaluation process in usability analysis. Table 1 presents their analysis of formative and summative evaluations.

It is apparent that the science of user centered interface design for virtual environments (VE) is developing. As a part of this development Hix and Gabbard [83], authored a taxonomy for usability characteristics as their contribution to the science of VE interface design. They argue that the day of “let’s build it and see what happens” is over, and future research should be focused on the use of user centered design methods.

Others that have studied interface design for virtual environments [84] have also argued for user-centered design of the interfaces [10]. Theoretical models, such as those reported by Kaur [85], have also been developed to support design of virtual environments, and competitive studies have been conducted by Evans, et al. [85] that are similar to the competitive study offered in this proposal. However, all possible manipulation tasks cannot be assessed. Thus, as discussed by Gabbard, et al. [13], it is important to identify a small and representative set of tasks from which to assess the system design, and this argues for the methods developed by Gabbard, Hix, and Swan [10], in which scenarios are constructed for typical cases as part of their user centered methods.

**Table 1 – Comparisons of Formative and Summative Evaluation [82]**

<b>Issue</b>	<b>Formative Evaluation</b>	<b>Summative Evaluation</b>
<b>Purpose</b>	Program improvement – ongoing process of providing feedback about the quality and effectiveness of the program	Program achievement – process of assessing the degree to which the program has accomplished predetermined goals
<b>Overall Evaluation Questions</b>	Is the program being implemented as planned to accomplish its intended goals?	Are efforts having a positive impact on user achievement?
<b>Specific Evaluation Questions</b>	What is working? What should be improved? How should it be changed?  Which interventions are being used? Is the intervention being implemented with fidelity?	What has happened? Who/what was affected? What was the most effective mechanism?  Was it cost-effective?
<b>Measurement Methods</b>	Surveys Interviews Observation Review documents and artifacts	Standards & benchmarks Achievements Local assessments Performance records

As an important adjunct to the development of virtual environment interface paradigms, the use of testbeds is suggested. Poupyrev, et al. [13] argue for in depth experimental studies because “there is still insufficient understanding of the essential characteristics and parameters of VR manipulation.” Their VRMAT system allows in depth studies, but it does not consider voice, gesture, and gaze. In a similar testbed, Bowman et al. [39] advocate evaluation of interaction techniques based on detailed empirical studies. They contrast this evaluation process with usability studies, and it is clear that information from a testbed study would be helpful in the development of virtual environments. However, the use of a usability analysis for interface design focuses on generation of a system for a specific purpose, such as Design Evaluation. Therefore, testbed studies are useful but do not fit the current research due to their limited scope.

#### **2.2.4 Usability in MMI Evaluations**

In the preparation of formal usability methods and evaluation techniques “rule of thumb” examinations are included in the development of Benchmarks and scenarios. As early as 1991, user interface studies by Jeffries et al. [81] showed that heuristic evaluation methods identify significantly more problems than any other method. Jacob Nielsen [79;86] presents a list of 10 usability heuristics as an aid to MMI usability testing development, repeated here as Table 2.

Evaluations of the usefulness of the interface methods are based on standard, Benchmark user scenarios defined in terms of the user action notation described in the 1988 work of Chin, et al. [87]. These Benchmark scenarios present a set of baseline standards for comparing these interfaces.

Usability specifications are “quantitative usability goals, that are used as a guide for knowing when an interface is good enough” [78]. The phrase “good enough” indicates that the goals are set high and that iterative refinement approaches are used to continuously improve the hardware, software, interfaces, and training to converge toward a successful result. However, without a usability metric, changes may not result in a more usable interface. Therefore, based on the maxim “*if you can’t measure it, you probably can’t*

*manage it,”* establishing usability specifications as measurable attributes in a repeatable scenario is an essential task.

**Table 2 – Usability Hueristics in MMI Testing [79;86]**

Visibility of system status	The systems should always keep users informed about what is going on, through appropriate feedback within reasonable time.
Match between system and the real world	The system should speak the user's language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
User control and freedom	Users often chose system functions by mistake and need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo (keystroke removal of input) and redo (keystroke re-insertion of deletions).
Consistency and standards	Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
Error prevention	Even better than good error messages is a careful design, which prevents a problem from occurring in the first place.
Recognition rather than recall	Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever possible.
Flexibility and efficiency of use	Accelerators, unseen by the novice user, may often speed up the interaction for the user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
Aesthetic and minimalist design	Dialogues should not contain irrelevant or rarely needed information. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
Help users recognize, diagnose, and recover from errors	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
Help and documentation	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

#### 2.2.4.1 Usability Analysis Techniques

Typically, Usability Analysis is used to evaluate and/or develop user-friendly software. This study employs Usability Analysis techniques to prove the research hypothesis and appears to be a unique application of Usability Analysis techniques.

Early studies in Usability Analysis were primarily conducted to assist software developers and hardware designers to improve the Human-Computer Interface (HCI) or Man-Machine Interface (MMI). For example, Benchmark scenarios were designed and executed to choose a means for users to navigate through a sequence of choices and to provide confirmation of a preference of one GUI paradigm over another functionally similar paradigm.

This study was conducted to provide comparative data supporting broad conclusions regarding the comparative merits of one technology (conventional, non-stereoscopic CAD) competed against another (tracked, stereoscopic virtual environment) in a ship design

application. This is an apparent new use for Usability Analysis, the development of a competitive evaluation.

Usability Analysis techniques establish usability metrics as measurable attributes in repeatable scenarios in order to provide a measure that is useful in the context of the comparative study. It is used to produce numerical, qualitative, and anecdotal information for comparative evaluation of the conventional and virtual environments under observation.

#### **2.2.4.2 User-Centered Development/Usability Specification Tables**

A major portion of this project is focused on evaluation of the usability of the user interfaces with conventional and immersive virtual environments; to iteratively refine the interfaces based on standard, Benchmark user scenarios defined in terms of the user action notation described in the work of Chin, et al. [87].

A Usability Specification Table provides a tabulated summary of usability characteristics, measuring instruments, and acceptable/planned/best values associated with each characteristic. Such tables were suggested by Whiteside et al. [80], and further refined by Hix and Hartson [78]. Each area to be measured is quantified with attributes, measuring instruments (normally Benchmarks and scenarios) that can be standardized with measures for acceptable characteristics with observed results for many classifications of users from novice (training modes) to experienced, expert users and trainers.

Table 3 presents a sample Usability Specification Table. Here the Usability Attribute to be measured is navigation in a traditional CAD system. The Benchmark test, in this sample which is Benchmark 1, measures navigation using an 86" screen for display (the environment). The value to be measured is the elapsed time required to navigate through the display space, find specific items within the space, and identify the item's location. The specification then shows that the Benchmark requires the user to perform the operations on 4 distinct items in each test (pass). This means that the user executes the Benchmark 3 times and the tester records elapsed times for each distinct item (part). Upon completion of the testing, the elapsed times for each part and pass are summed for later analysis. It should also be noted that the specification presents some target levels of results. These are, in this case, elapsed times determined (estimated) by the developers and users that are acceptable, planned, and "best possible"

Hix and Hartson [78] define the usability attribute as: “the general usability characteristic to be measured for an interface” and suggest the following as common, measurable usability attributes:

- 1<sup>st</sup> impression – user’s reaction to the presentation of the system
- Initial performance – user’s performance during the very first use
- Long-term performance – user’s performance during more constant use
- Learnability – how quickly and easily users learn to use the system
- Retainability – how well users retain what they have learned over time
- Advanced feature usage – helps determine use of complicated actions.
- Long-term user satisfaction – user’s opinion of the system after protracted use.
- 

**Table 3 – Sample Usability Specification Table [88]**

Usability Specification Table: Benchmark 1 - Navigation												
Attributes - Measures			Levels (Targets)				Observed (in Seconds)					
Usability Attribute	Measuring Instrument	Value to be Measured	Part	Accept	Plan	Best	Pass	User 1	User 2	:::	User N	Avg
Traditional CAD Navigation	Benchmark 1 86" Non-Stereo ImmersaDesk Navigation	Time req'd to locate, identify, and catalog parts	1	120	30	10	1	45	45	:::	70	95.56
							2	50	35	:::	80	62.78
							3	35	50	:::	65	43.33
			2	120	30	10	1	55	90	:::	85	90.00
							2	45	50	:::	40	68.89
							3	40	60	:::	80	49.44
			3	120	30	10	1	35	25	:::	20	34.44
							2	25	20	:::	20	31.67
							3	20	25	:::	15	23.89
			4	120	30	10	1	100	100	:::	45	112.78
							2	100	80	:::	50	82.22
							3	65	70	:::	85	63.33
			Sum	480	120	40	1	200	235	:::	200	298.33
							2	195	165	:::	170	213.89
							3	140	180	:::	230	156.11

Skill decay after periods of skill disuse is well known and has substantial implications when relatively long periods of time separate training from the application of learned skills. In a 1998 study of VR interface training, Hall et al. [89] examined the differential effects of virtual reality versus conventional computer-based media on skill retention. The results reported were consistent with earlier research showing that VR may not be superior to conventional electronic media for training certain intellectual skills.

However, usability attributes should be chosen according to the identity of the intended users and what representative tasks are to be utilized across these users and user groups.

The measuring instrument is a description of the method of determining the values for the usability attribute and may be broadly classified as objective (observable quantitative) measures or subjective (quantifiable opinion) measures of the attribute. The measuring instrument is normally a description of a scenario presented to the user describing the system conditions prior to the action(s), the user action(s), and the desired system conditions after the user action(s). The performance of the tasks provides a required objective usability metric. Often these scenarios are documented as Benchmarks for repeatable observation.

Questionnaires can be used for collecting subjective measures from users and works well for quantifying user satisfaction with an interface. For this project, a variation of the Questionnaire for User Interface Satisfaction (QUIS) developed at the University of Maryland by Chu et al. [20], is used as a baseline for measuring user satisfaction.

In the October 2001 Society for Technical Communication newsletter, Lund [90] discusses a number of issues that tend to recur in the life of an interface design and suggests that the USE Questionnaire (Usefulness, Satisfaction, and Ease of use and learning) appears to help interface evaluation by providing a mechanism for subjectively measuring these characteristics that tend to drive user satisfaction. His work describes the kinds of questions that tend to provide “a good sense of what is usable and what is not.” Sample questions reflect on:

- 1) Usefulness
  - 1.1) It helps me be more effective.
  - 1.2) It helps me be more productive.
  - 1.3) It is useful.
  - 1.4) It gives me more control over the activities in my life.
  - 1.5) It makes the things I want to accomplish easier to get done.
  - 1.6) It saves time when I use it.
  - 1.7) It meets my needs.
  - 1.8) It does everything I would expect it to do.

2) Ease of Use

- 2.1) It is easy to use.
- 2.2) It is simple to use.
- 2.3) It is user friendly.
- 2.4) It requires the fewest steps possible to accomplish what I want.
- 2.5) It is flexible.
- 2.6) Using it is effortless.
- 2.7) I can use it without written instructions.
- 2.8) I notice no inconsistencies as I use it.
- 2.9) Both occasional and regular users would like it.
- 2.10) I can recover from mistakes quickly and easily.
- 2.11) I can use it successfully every time.

3) Ease of Learning

- 3.1) I learned to use it quickly.
- 3.2) I easily remember how to use it.
- 3.3) It is easy to learn to use.
- 3.4) I quickly became skillful with it.

4) Satisfaction

- 4.1) I am satisfied with it.
- 4.2) I would recommend it to a friend.
- 4.3) It is fun to use.
- 4.4) It works the way I want it to work.
- 4.5) It is wonderful.
- 4.6) It is pleasant to use.
- 4.7) I feel I need to have it.

The effectiveness of virtual environments (VE's) has been linked to the sense of presence reported by those users of VE's; where presence is defined as the subjective experience of being in one environment, even when one is physically in another. This can



be described as a normal awareness phenomenon that requires directed attention and is based on the interaction between the sensory stimulation, environmental factors that encourage involvement and enable immersion, and internal tendencies to become involved. Witmer and Singer [91] use these factors and others as the basis for a presence questionnaire (PQ) to assist in making these subjective measurements in VE's. They also developed an immersive tendencies questionnaire (ITQ) to measure the differences in the tendencies of individuals to experience presence. The PQ and ITQ are noted to be internally consistent measures showing a weak but consistently positive relationship between presence and task performance in VE's.

The value to be measured provides the metric for which the data are collected under the specific conditions outlined in the scenario description. Common objective measures are: time to complete a task and the number of errors encountered during the performance of the task. An important part of the design of the usability assessment is the determination of what constitutes an error. In assessing the usability of virtual environment activities such errors might include: invalid commands initiated via the user interface or inaccurate/incomplete results of the operations.

Scenario/Benchmark identification, generation and documentation is an important part of this project. Working with novice and experienced CAD designers the investigators identify common 3D design activities and prepare detailed descriptions of these activities. The documentation of these processes is carried out using the user notation formats described in the work of Chin et al. [87] to produce consistent, repeatable scenarios in the virtual environment.

## **3 Competitive Usability Study of Shipbuilding VE's**

### ***3.1 Need***

As described in the literature, many studies of VR and VE equipment, environments, and systems are documented that span a vast array of applications from meta-data display to medical training and aircraft design. However, there is very little work reported on the use of VR and VE in shipbuilding applications and even fewer studies of the appropriateness of their user interfaces. This project addresses that void via the application of Usability Analysis (UA) techniques to compete several VE's and interfaces that are currently available to the industry. This study is performed to define appropriate measures and mechanisms for collecting these data so that a statistically rigorous analysis can be performed to identify and quantify the "best" environment/interface for use in a shipbuilding context.

### ***3.2 Usability Tests for Systems and Interfaces***

In order to assure consistent results across a broad user spectrum, the principles of Usability Analysis are applied to this virtual design environment project. The aim throughout the project is to enforce a well-developed user interface development process. Standard configurations, metrics, measures, benchmarks, and scenarios are developed, implemented, and the results are analyzed.

#### **3.2.1 Specify Standard User Configurations**

A standard user configuration for the ImmersaDesk system was specified to include hardware, software, interfaces, and network requirements. Such configurations define the environments for local and distance users. Metrics and traceability were used to help collect accurate data for usability analysis.

### **3.2.2 Usability Tests of Systems and Interfaces**

During the design and development stages of the project, investigators identify metrics that provide measures of many attributes of the systems. In this phase, suitable standards, metrics, and acceptable values are defined for reporting system attributes such as:

- Effectiveness
- Ease of use
- Performance
- Maintainability
- Learnability
- Retainability.

### **3.2.3 Identify and Document Basic User Functions as Benchmark Scenarios**

Operational sequences of basic functions were developed into standard Benchmarks also including summative usability metrics. As described by Nielsen and Phillips [92], estimating values for absolute user performance shows very high variability, but choosing the fastest of the alternative designs has “a net present value more than 1,000 times the cost of getting the estimates.” Using this information as a basis, a library of basic functional scenarios with associated acceptable measurement values (elapsed times for navigation/object manipulation and object placement offsets) was created from which user or operations specific Benchmarking tests were created for evaluating both systems and users in a consistent, repeatable manner.

## ***3.3 Environments/Interfaces***

This study attempts to quantifiably identify the “best” of four user-centered design environments by performing a set of project specific Benchmarks to determine the usability of each environment. The environments tested were:

- **W/S:** Standard 19” screen, CAD workstation (W/S) for rendering perspective representations of a 3D ship space to a 2D screen. This environment uses conventional non-stereoscopic/mouse-driven interface design review and evaluation tools and is relatively inexpensive. The environment is consistent with CAD and engineering software commonly employed in practice, and therefore represents scenarios consistent with conventional practice.

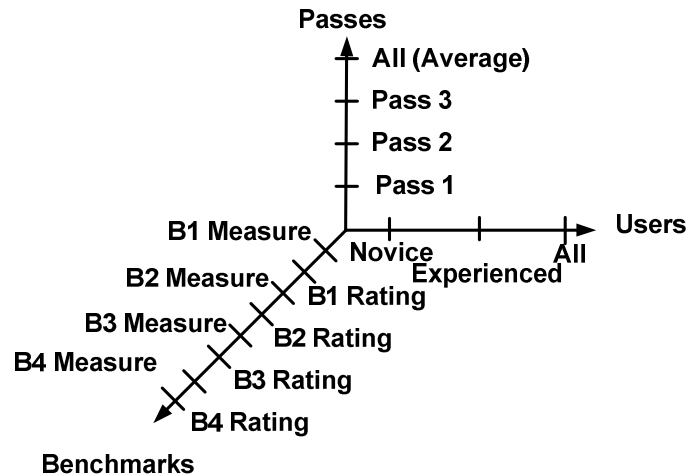
- **86" Non:** A similar CAD workstation utilizing the same non-stereoscopic/mouse-driven interface design review and evaluation tools but projected, without stereoscopic viewing, on the 86" ImmersaDesk screen. Essentially, 86" Non is the same as a W/S with a large screen. The use of the 86" screen allows for evaluation of screen size as a factor in system testing.
- **Wand:** A more complicated, more expensive, fully tracked implementation of a stereoscopic, wand-driven interface projected in the 86" stereoscopic ImmersaDesk environment. The wand is used to direct activity in the virtual environment as a gestural interface.
- **Glove:** A similar, more extensive, multi-modal, gestural interface with a fully tracked implementation in the stereoscopic environment using a voice-command and right-hand data-glove interface and design review tools.

### ***3.4 Scenarios/Passes/Groups/Data***

Figure 2 diagrams the dimensions of testing performed in each of the four study environments. The horizontal axis of the diagram shows that within each environment the testing is performed by each member of 2 user groups (15 novice and 15 experienced designers). The vertical axis shows that each test is performed 3 times (passes). The third axis shows that 3 distinct test scenarios (Benchmarks) are executed. With each Benchmark, a quantitative measurement of both user interface performance is recorded. Additionally, at the completion of each pass, a subjective user's evaluation (rating) is recorded via a questionnaire/survey reflecting of the user's opinion of the interface.

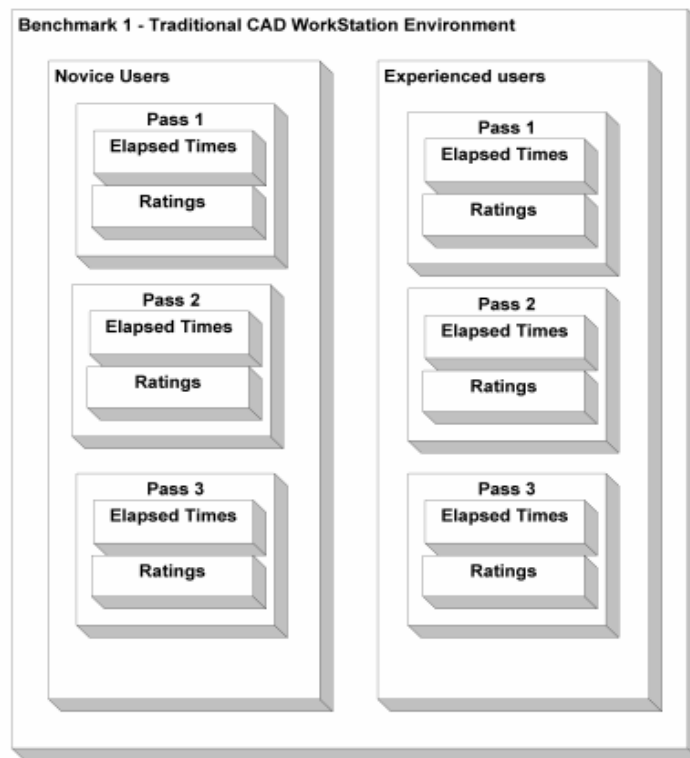
Each Benchmark is comprised of a set of objective measures and subjective measures. For Benchmark 1 (navigation) and Benchmark 2 (error identification and object manipulation) the measure is the elapsed time to perform each portion of the test. For Benchmark 3 (spatial awareness) the objective measure is a distance measurement of perceived placement of an object vs actual placement (offset, in mm).

The subjective portion of each Benchmark, in the form of a questionnaire, was administered at the completion of each pass. The survey was a means for the users to provide their impressions of the environment/interface during that portion of the testing sequence. Figure 4 provides an example of the survey form showing each of the 22 specific items rated (scale: 1-5, poor – very good). Altogether, more than 10,580 distinct points were collected during the study.



**Figure 2 – Test/Data Dimensions**

As another means of representing the dimensions of the usability tests, Figure 3 shows the typical groupings, levels, and layers of the tests performed. This figure is shown for Benchmark 1, but the other 4 Benchmarks are similar.



**Figure 3 - Benchmark Test Groupings and Layers**

This study was performed using two groups of testers. Each group was composed of at least 15 novice or 15 experienced users. The Novice User group consisted of primarily undergraduate engineering students with little actual ship design experience. Novice users

were selected to help define the intuitiveness and learnability of the interface. The Experienced user group was comprised of a cross-section of designers experienced in a variety of ship design related specialties including: electrical, layout, piping, etc. The competence level of this group ranged from two to eighteen years of direct ship design experience.

This study developed three Benchmark scenarios to be performed as the basis of the investigation. Each was designed to test specific attributes of the interfaces within each of the environments. Each hands-on test was followed with the questionnaire/survey designed to elicit the subjective evaluation of the interface/environment from each user. Figure 4 shows the questionnaire/survey used.

Usability Survey						
User ID: _____						
Environment:    19" CAD    86" CAD    Stereo    Glove						
Pass:    1    2    3						
		Very Good 5	Good 4	Neutral 3	Poor 2	Very Poor 1
<b>Navigation</b>						
1	Initial impression of navigational modes					
2	Gross control movement					
3	Speed of cursor/pointer movement					
4	Ability to make fine adjustments to the placement of the cursor/pointer					
5	Ability to recover cursor/pointer movements					
6	Ease of use					
7	After-test impression of the navigational modes					
<b>Locating</b>						
1	Initial impression of the interface in locating specific parts/equipment					
2	Ease of identification of selected part/equipment					
3	Ability to make fine adjustments in selecting specific parts/equipment					
4	Ease of use					
5	After-test impression of the location/selection mechanism					
<b>Movement</b>						
1	Initial impression of the interface for relocating parts/equipment					
2	Ease of movement across the three axis					
3	Ability to make fine part/equipment movement adjustments					
4	Ease of use					
5	After-test impression of the movement mechanism					
<b>General</b>						
1	Initial impression of the overall system					
2	Ability to relate a 2D planform to the space as presented					
3	"Intuitiveness" of the interface - do the controls follow expected use?					
4	Overall ease of use					
5	After-test impression of the overall system.					
<b>Comments</b>						

**Figure 4 – Usability Questionnaire**

Each user executed each Benchmark three times (passes) with the starting environments randomized as to sequence of presentation for each user:

- **1<sup>st</sup> pass** user's initial exposure to the space and to the interface environment. Results help gauge the intuitiveness of the environment/interface and provide a baseline for analysis of ease of use, learnability and retainability.
- **2<sup>nd</sup> pass** – normally run later the same day or early the next day providing the user some time to assess his/her experience within the environment. Results help extend knowledge of the characteristics of the environment/interface.
- **3<sup>rd</sup> pass** – normally run one or two days after pass 2 to help further assess the retainability of the interface and provide trend data for analysis.

A highly cluttered, Air Conditioning and Refrigeration space (AC&R) of the USS San Antonio class Amphibious Transport Dock (LPD-17), was selected for the Benchmark 1 and 2 test scenarios (see Figure 3). The navigation scenario (Benchmark 1) required that users start at the entryway into the AC&R space then locate pre-defined equipment within the space and then return to the entryway utilizing the interface tools of each of the environments under test. The measure was elapsed time (in seconds) to perform each task. Each user located, identified, and catalogued (noted the equipment and position) four distinct parts within the space.

Benchmark 2 is similar to Benchmark 1 in that elapsed time from entry into the AC&R space, movement within the space, and return to the starting point (entry hatch) is the prime measurement. However, with Benchmark 2, errors in construction of the space are made (two errors per pass), and the identification and correction of those errors are required by the user.

Benchmark 3 was designed to measure the ability of the environments to aid user spatial awareness. Users are presented with a totally foreign space into which two readily recognizable icons are randomly placed. Using each of the interfaces/environments, users are asked to navigate through the space to locate each icon. While the elapsed time required to locate each icon is recorded, the primary measure provided by this Benchmark is the user's placement offset from the actual location of the icon. The user's perceived placement of each icon is recorded on a 2D plan view of the space and the offset from the actual placement is measured in mm.

The environment execution sequence for each user within each test group (Novice and Experienced) was randomized so as to equalize the benefit of navigating within a specific environment using one interface paradigm before another paradigm. This randomization avoids statistical bias in the test results.

As can be seen in Figure 2, large volumes of objective and subjective test data were collected during the course of the study allowing detailed analysis. Each portion of the overall study allows for a detailed analysis with comparison of search times or other measurements for each specific item/part in a specific Benchmark/environment by class of user or specific user. The following sections of this document discuss higher-level test results and provide some detailed analysis of these results by Benchmark. Results and analysis of the testing performed in this study are presented here using the nomenclature described in Figure 5.

As an example, B1Np2Tw/s refers to the Benchmark 1 (Navigation), Novice user, 2<sup>nd</sup> iteration, timing test within the WorkStation interface.

By Benchmark	
1 - Navigation	B1
2 - Manipulation (Find/Repair)	B2
3 - Spatial Awareness	B3
By Environment	
Traditional 19" CAD WorkStation	w/s
36" Non-Stereoscopic ImmersaDesk	non
Stereoscopic Wand	wand
Stereoscopic Voice/Glove	glove
By User Group	
Novice	N
Experienced	E
All	A
By Pass	
Pass 1 - Initial	p1
Pass 2 - Intermediate	p2
Pass 3 - Final	p3
3 Pass Average	avg
By Metric	
Timing or Offset	T/O
Rating	R
Navigation	nav
Locating (Find)	loc
Manipulation (Move)	mov
General Impressions	gen
Overall Impressions	ovr

**Figure 5 – Test/Data/File Naming Conventions**

### ***3.5 Role of Statistical Analysis***

Macnaughton [93] describes empirical research as “any activity in which data are gathered from some area of experience and then conclusions are drawn from the data about



the area of experience” and is “a crucial step in the scientific method, which is central to many areas of human endeavor, such as science, education, business, industry, law, and government.” Furthermore, Macnaughton describes statistics as “a set of optimal techniques used to help study the variables and relationships between variables samples as a means to accurately predict and control the values of variables (properties) in entities in populations.”

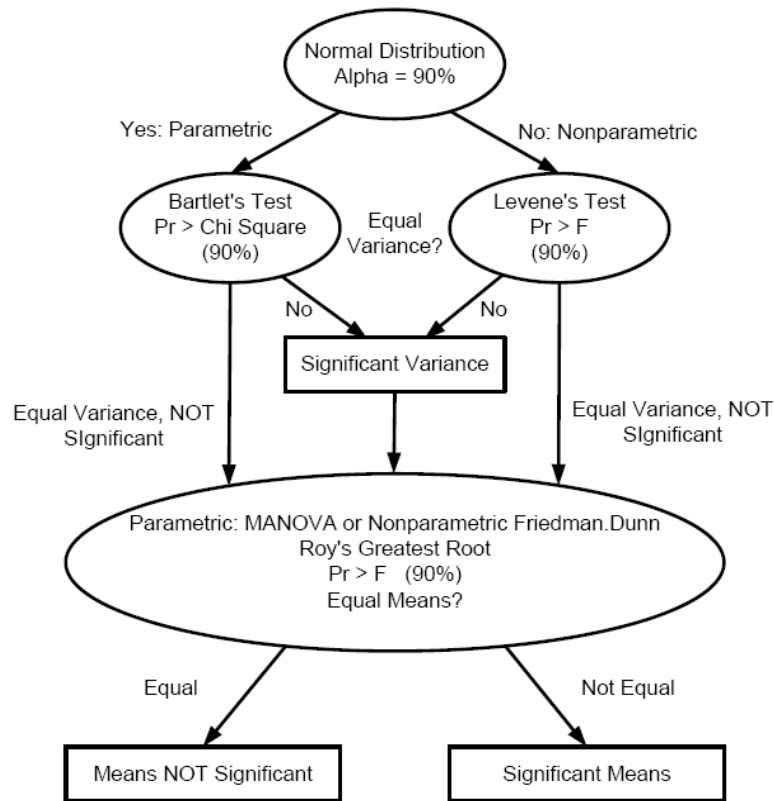
The empirical research reported herein applies several standard statistical analysis techniques to both the objective (timings, distance offsets, and fault counts) and subjective (user evaluation ratings) data collected during the study. This study uses two advanced software packages for selecting and applying the appropriate post-ANOVA tests for comparing the measures of the environments:

- GraphPad Prism version 4.00 for Windows [94]
- SAS for Windows version 8.0 [95].

Two basic statistical approaches for making conclusions about the measure of the study’s test environments are applied:

1. The first is to assume that the parameter values (elapsed times, offsets, or rating values) for the population follow a Gaussian (normal, bell-shaped) distribution. A normal distribution allows statistical tests providing inferences about the mean (and other properties) of the population. Such tests are called parametric tests.
2. The second method is to rank all values in the population from low to high and then to compare the distribution of ranks. This is the principle underlying most commonly used nonparametric tests, which are used to analyze data from non-Gaussian distributions. Prism uses the Friedman-Dunn test which compares the measures from paired environments. It calculates the difference between each set of pairs and analyzes that list of differences.

Figure 6 shows the statistical test selection criteria for the analysis of the data. The selection process diagrammed in Figure 6 provides a sequential (top-down) view of the analysis regimen.



**Figure 6 – Statistical Analysis Test Selection**

### 3.5.1 Kolmogorov-Smirnov Normality Testing

Both the InStat/Prisim and SAS software test for normality using the Kolmogorov-Smirnov test. The KS statistic (reported as the KS-statistic in InStat/Prisim and the D-statistic in SAS) quantifies the discrepancy between the distribution of the data and an ideal Gaussian (normal or bell-shaped) distribution - a larger value denotes a larger discrepancy. It is not informative by itself, but is used to compute a P value.

The method of Kolmogorov and Smirnov originally published cannot be used to calculate the P value because their method assumes that the mean and standard deviation of the overall population is known. However, in analysis, rarely is the overall population mean and standard deviation known; more often only the mean and standard deviation of the sample is known. To compute the P value, therefore, the software uses the Dallal and Wilkinson approximation to Lilliefors' method [96]. Since that method is only accurate with small P values, the software often simply reports "P>0.10" for large P values.

The P value from the Kolmogorov-Smirnov normality test answers this question: Randomly sampling from a Gaussian population, what is the probability of obtaining a sample that deviates as much from a Gaussian distribution (or more so) as this sample does. More precisely, the P value answers this question: If the population was really Gaussian, what is the chance that a randomly selected sample of this size would have a KS distance as large, or larger, as observed [97]?

### **3.5.2 Homogeneity of Variance (Homoscedasticity)**

ANOVA is based on the assumption the populations (environments) all have the same variance (homoscedasticity). However, early test results showed large variations and standard deviations in Benchmark 3 which is a measure of spatial awareness (distance offsets). Several statistical measures are available to help determine the relative “sameness” of the variance reported values across the user groups within the benchmarks. These include: the coefficient of variance, Bartlett’s test (for normally distributed results), and Levene’s test (for non-gaussian distributions).

#### **3.5.2.1 Coefficient of Variance**

In order to use ANOVA, tests must be completed to ensure homogeneity or relative sameness of population variances. As one means of assessing variance, the coefficient of variance (CV) is the degree to which a set of data points varies and is sometimes called the relative standard deviation since it takes into account the mean (average). The CV may also be described as the measure of significance of the sigma, the standard deviation, in relation to the mean and may be reported as a percentage value using the formula:

$$CV = (\sigma / \text{mean}) * 100. \quad (3-1)$$

The larger the CV the greater the variability in the data or the more significant the sigma relative to the mean; thus reporting the CV aids in interpreting the characteristics of the distribution of the test results. For example, simply reporting a standard deviation of 10 says nothing of the significance of this value. If the average test value (mean) is 1000, a sigma of 10 is not very significant whereas if the mean is 15, a sigma of 10 is very significant.

### 3.5.2.2 Bartlett's Test/Levene's Test

Bartlett's test provides a mechanism for determining the homogeneity of the variances of test data sets by calculating a P value for reporting the significance of the comparison. Small P values show a high significance and indicate that the variances differ and that nonparametric methods for comparing the means of the environments are required. The test P value answers the question: "If the populations really have the same variance, what is the chance that you'd randomly select samples whose variances are different (or more different) as observed in your experiment?" [97].

Low Bartlett's test P values may be due to data that are non Gaussian, rather than due to unequal variances. For this reason another test is required for data that fail the Kolmogorov-Smirnov normality test. Levene's test for homoscedasticity is less robust than Bartlett's test but is less sensitive to data that is non Gaussian in distribution and is therefore indicated when data for either (or both) of the environments fails the normality test. Levene's test P values are interpreted in the same manner as for Bartlett's P values.

### 3.5.3 F-Statistic

In ANOVA, the sampling distribution is based on a ratio called the F-statistic which is the ratio of the population variance as estimated between groups versus within groups. If the test values (samples) come from populations with identical means and variances, then each of the sample variances is an estimate of the same quantity (i.e., the population variance). The term *within* is used because the population variance is being estimated separately within each sample. Estimating the variance in the population can also be made by using the variation in sample means across or *between* conditions (environments). Thus the F-statistic can be reported as the ratio the mean squares of these values [98]:

$$F = (MS_{\text{between}}) / (MS_{\text{within}}). \quad (3-2)$$

In the context of a one-way ANOVA, several methods are used to report the importance of variables within a test (the F-statistic): Wilks' Lambda, Pillai's Trace, Hotelling-Lawley Trace, and Roy's Greatest Root are all reported by the SAS software. Each shows a reflectance of a variables importance and produce much the same F-statistic [99].

Roy's Greatest Root is a powerful multivariate generalization of the univariate F-statistic and is reported by the SAS software as:

$$F(\text{groups}, df) = (MS_{\text{between}}) / (MS_{\text{within}}) \quad (3-3)$$

where the F-statistic is reported for the noted degrees of freedom.

### **3.5.4 Analysis of Variance (ANOVA)**

Analysis of variance (ANOVA) is a statistical procedure where the means of two or more groups are compared. When more than two groups are compared, a post-hoc test is done to determine which of the groups differed. ANOVA only indicates that differences exist and is a test of the significance of the differences between the compared means as it analyzes the variation between and within each group. However, since the goal of the study is to identify the “best” environment by competing the usability attributes (elapsed times, distance offsets, fault counts, and user ratings), a repeated measures analysis of variance is indicated. Repeated measures occur when the measurements can be thought of as responses to levels of an experimental factor of interest (such as time or distance). Thus, a repeated measures analysis of variance (ANOVA) approach is applied for the comparison of the means of environments.

#### **3.5.4.1 Hypothesis Testing**

Hypothesis testing is the use of statistics to determine the probability that a given hypothesis is true. Testing to statistically determine whether two samples are from different populations is to test hypotheses about each sample (environment). To do so, a set of random samples is taken from each environment to determine the mean and standard deviation of each set. The mean of each sample is then the summary of the characteristic under observation. With this data the test is now whether there is a difference between the sample means. This may be stated more formally as the probability that any difference between sample means is simply due to the effects of random sampling of the same population rather than random sampling of two different populations. Statistically this is expressed as “Null” and “Alternative” Hypotheses. Null Hypothesis ( $H_0$ : environment 1 means = environment 2 means) true indicates that the sample means are drawn from the same or identical populations. Alternative Hypothesis ( $H_a$ : environment 1 means  $\neq$

environment 2 means) true indicates that the sample means are drawn from different populations.

### 3.5.4.2 One-tailed and Two-tailed Hypothesis Testing

The discussion of hypothesis testing above presents the  $H_0$  and  $H_a$  for two-tailed tests of hypothesis. These same tests also may be crafted to examine a different set of hypothesis that show whether one population mean is greater than (or less than) another. One-tailed tests allow results to be interpreted as showing one environment to be “better” (faster, higher, closer, lower) than another.

The alternative hypothesis ( $H_a$ ) under test changes with the interpretation of “better” for the measure. More specifically, for elapsed time and distance offsets measures a “better” environment is indicated by shorter elapsed times or smaller distance offsets; thus:

$$H_a: (1^{\text{st}} \text{ environment mean}) \leq (2^{\text{nd}} \text{ environment mean}). \quad (3-4)$$

Similarly, higher user subjective ratings of an environment or larger count of faults identified indicate the more “preferred” interface. Thus, the alternative hypothesis for comparing environments is:

$$H_a: (1^{\text{st}} \text{ environment rating mean}) \geq (2^{\text{nd}} \text{ environment rating mean}). \quad (3-5)$$

The SAS software uses the GLM (General Linear Model) process for performing MANOVA test reported for this project. The purpose of the MANOVA (Multivariate ANOVA) is to use several environments simultaneously to discern significant differences among the experimental environments. For the analysis of the results of this study, the SAS outputs compare Roy’s F-statistic with the F-statistics of the pair-wise comparisons of the GLM to determine the statistical significance of the comparisons. Specifically, if the results of the comparison are significant, the null hypothesis  $H_0$  is rejected and the alternative hypothesis ( $H_a$ ) is accepted. Therefore, the environment with the lower mean value (for timings and offsets) or higher mean value (for user ratings) can be said to be statistically “better” than the other environment considered.

Similarly for data sets that are not normally distributed or exhibit a high coefficient of variance, Friedman-Dunn testing is indicated. The Friedman test compares groups by ranking the values in each matched set from high to low. The method then sums the ranks in each environment. If the sums are very different, the P value is small. The P value

answers the question: If the different environments really are identical, what is the chance that random sampling would result in sums of ranks as far apart (or more so) as observed in the testing [97]? Dunn's method is an extension of the Friedman method that provides for the pair-wise comparison of multiple environments.

In Benchmark 4 testing the data are unpaired thus negating the use of the Friedman-Dunn method for data sets that are not normally distributed. A suitable nonparametric comparison method (Mann-Whitney) is employed to make the test of the hypothesis of equal means ( $H_0$ ).

The Mann-Whitney test [97] was selected because:

- (1) the difference between each value and population median is independent
- (2) the data are unpaired so that comparison of the mean rank in the two populations may be used
- (3) the populations are not normally distributed.

The key result of the test is a P value that answers the question: If the populations really have the same median, what is the chance that random sampling would result in means as far apart (or more so) as observed? Thus, if the P value is small, one can reject the idea that the difference is a coincidence, and conclude that the populations have different medians. However, if the P value is large (as is the Benchmark 3 comparisons), the data do not provide any reason to conclude that the overall medians differ, and  $H_0$  is accepted.

### **3.5.5 Reliability - Cronbach's $\alpha$**

Cronbach's  $\alpha$  (alpha) measures how well a set of items measures a single unidimensional latent construct (user environment ratings, elapsed times, offset distances, or fault counts). Technically speaking, Cronbach's  $\alpha$  is not a statistical test – it is a coefficient of reliability or consistency. Cronbach's  $\alpha$  measures the reliability of tests, observations, experiments, or measurements by estimating the extent to which they provide the same results on repeated trials [100]. It has an important use as a measure of the reliability of a psychometric instrument, since it assesses the extent to which a set of test items can be treated as measuring a single latent variable.

The benchmark scenarios of this study are the psychometric instruments used to measure the objective user applications of the interfaces (elapsed times, distance offsets, or fault counts) and the user's subjective ratings/impressions of the environment.

Cronbach's  $\alpha$  is defined as a function of the number of test items and the average inter-correlation among the items [101]:

$$\alpha = (N * r\text{-bar}) / 1 + ((N-1) * r\text{-bar}), \quad (3-6)$$

where  $N$  is the number of items (user measures) and  $r\text{-bar}$  is the average inter-item correlation among the items. It is important to note that as the number of items (users) increases,  $\alpha$  increases, and if the average inter-item correlation is low,  $\alpha$  is low. Thus, if the inter-item correlations are high, there is evidence that the items (user ratings, etc.) are measuring the same underlying construct, i.e., the reliability of the measuring instrument is high referring to how well the values measure a single condition or property [100].

$\alpha$  can take values between minus infinity and 1 (although only positive values make sense). As a rule of thumb, the measuring instrument (benchmark) should only be used if an  $\alpha$  value of 0.70 or higher is obtained. This value is sometimes reported as a percentage and is termed the reliability of the measuring instrument [101].

As is reported in the following sections, Cronbach's  $\alpha$  for the benchmarks of this study range from 0.79 to 0.97. This indicates that the study is based upon a set of reliable objective and subjective measuring instruments.



## **4 Benchmark 1 (Navigation)**

### ***4.1 Description***

As previously discussed, the Benchmark 1 scenario was designed to test the user's ability to utilize the four environments/interfaces to navigate through the study space locating each of four distinct items/parts within the space. The common measure recorded was simply the elapsed time to navigate the space (from a common starting point), locate each required item/part, and return to the entry hatchway. Each novice and experienced user performed this Benchmark 3 times in each of the four environments. What follows is an analysis of the final pass results of these Benchmark 1 tests presented by the two classes of users (Novice and Experienced). Pass 3 results represent each user's final exposure to each environment within each scenario (Benchmark). Therefore, pass 3 results show the user's ability to perform the required tasks. Each environment/interface (W/S, 86" Non, Wand, and Glove) is represented in a distinct chart.

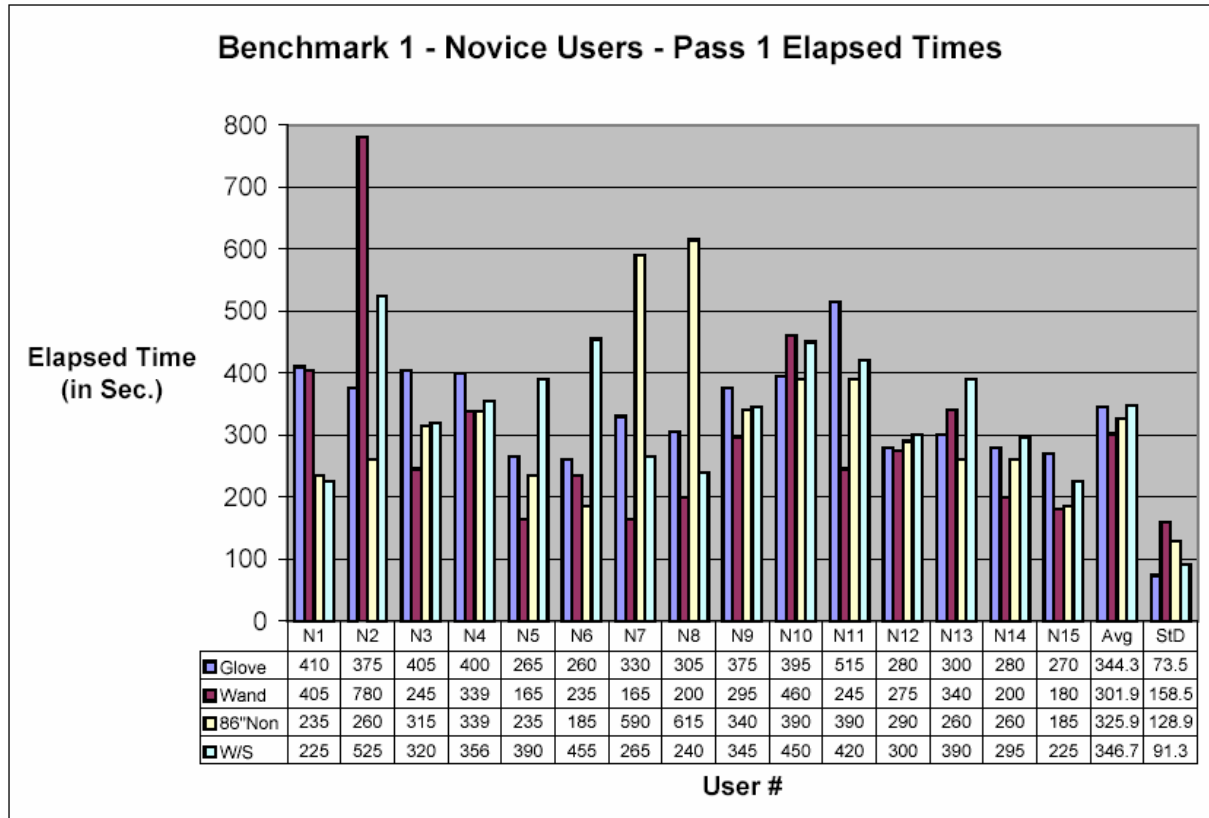
### ***4.2 Pass-to-Pass Improvements in Elapsed Times***

#### **4.2.1 Novice Users**

Figure 7 shows novice user elapsed times for pass 3 of the navigation Benchmark tests in all four environments. A preliminary investigation of the chart data shows that the novice users performed navigation tasks approximately 10.2% faster using the stereoscopic (wand) interface over either of the non-stereoscopic environments. However, overall non-stereoscopic interface methods resulted in elapsed navigation times that were only 3.9% faster than the stereoscopic methods. This result is attributed to the higher navigation times posted for the novice user's first exposure to the more complex stereoscopic voice/glove interface.

Table 4 presents the improvements in navigation times for novice users with each successive exposure to each of the four test environments. Note that there appears to be only a 3.1% difference in the increase between stereoscopic environments versus non-

stereoscopic environments for novice users in comparing pass 1 to pass 3 speeds. Thus, for novice users, initial exposure to navigation methods shows a slight increase in navigation speeds for any of the environments.



**Figure 7 – B1Np3Tim – Novice User Pass 3 Navigation Times**

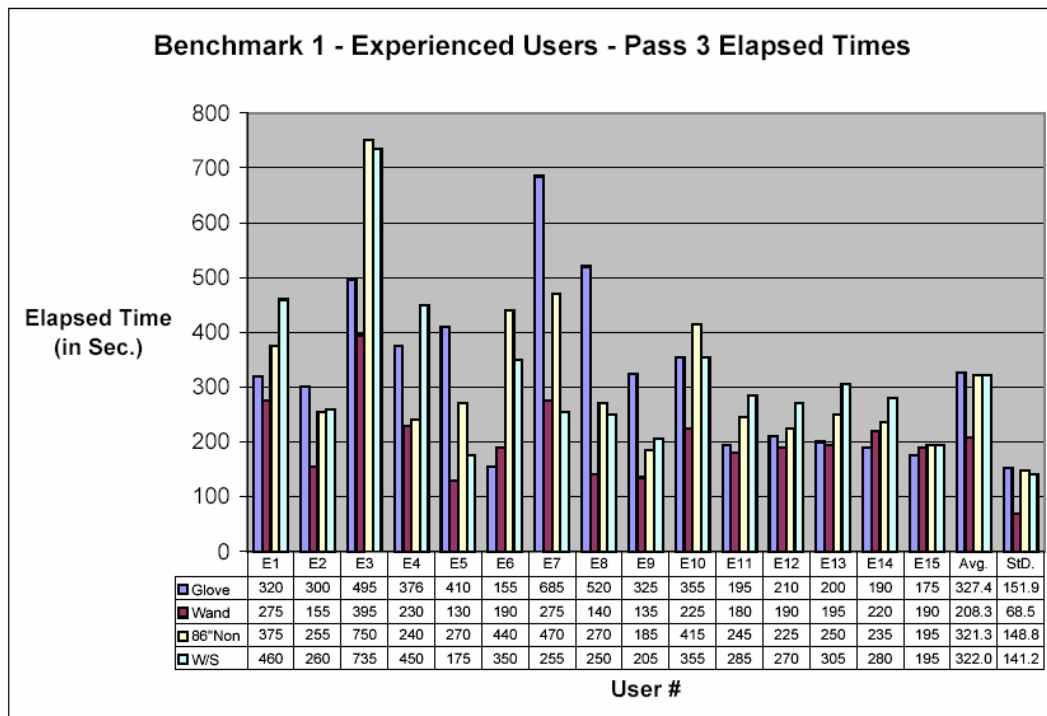
**Table 4 – B1N Pass-to-Pass Elapsed Time Changes**

Benchmark 1 - Novice Users - Pass-to-Pass Elapsed Time Changes								
B1N	Pass 1 to Pass 2		Pass 2 to Pass 3		Pass 1 to Pass 3		Average to Pass 3	
Δ Improve	Sec.	%age	Sec.	%age	Sec.	%age	Sec.	%age
Glove	77.33	22.46%	54.67	25.75%	132.00	38.33%	62.22	29.30%
Wand	110.43	36.58%	19.57	11.38%	130.00	43.06%	49.86	29.00%
86" Non	88.60	27.19%	42.77	21.98%	131.37	40.31%	58.04	29.84%
W/S	83.00	23.94%	68.34	34.98%	151.34	43.65%	73.23	37.48%

#### 4.2.2 Experienced Users

Figure 8 shows experienced user elapsed times for pass 3 of the navigation Benchmark tests in all four environments. Examination of these results shows that

experienced users were approximately 10% faster than novice users across all the environments. The results show that for experienced users the stereoscopic wand interface performs approximately 35% faster than the stereoscopic voice/glove interface or either non-stereoscopic interface.



**Figure 8 – B1Ep3Tim - Experienced User Pass 3 Navigation Times**

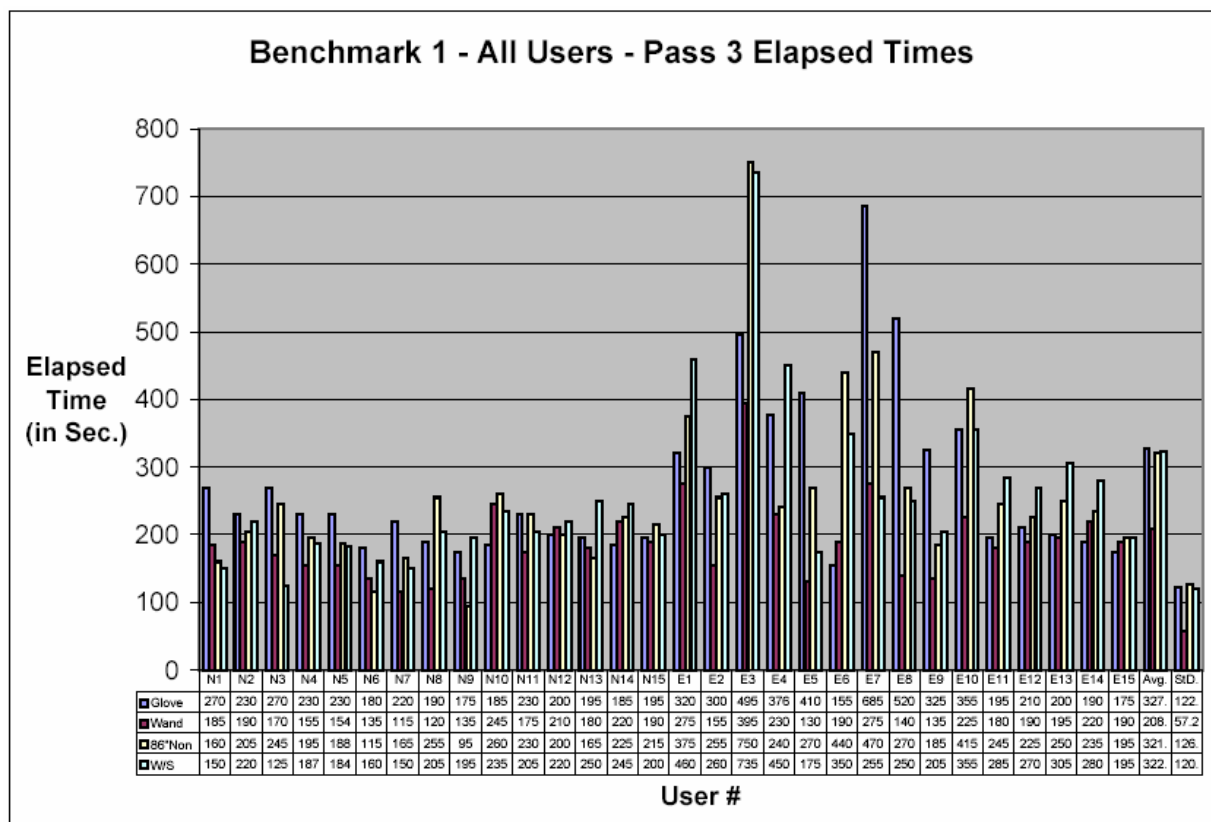
Table 5 presents the improvements in navigation times for experienced users with each successive exposure to each of the four test environments. Note that experienced users show little appreciable improvement in navigation times across the 3 passes with the stereoscopic voice/glove or non-stereoscopic traditional workstation environments. However, experienced users showed improved navigation speeds by 31.54% from pass 1 to pass 3.

**Table 5 – B1E Pass-to-Pass Elapsed Time Changes**

Benchmark 1 - Experienced Users - Pass-to-Pass Elapsed Time Changes								
B1E Δ Improve	Pass 1 to Pass 2		Pass 2 to Pass 3		Pass 1 to Pass 3		Average to Pass 3	
	Sec.	%age	Sec.	%age	Sec.	%age	Sec.	%age
Glove	3.50	1.03%	9.50	2.82%	13.00	3.82%	7.50	2.24%
Wand	53.00	17.42%	43.00	17.11%	96.00	31.54%	46.33	18.19%
86\" Non	25.67	6.37%	55.67	14.77%	81.33	20.20%	45.67	12.44%
W/S	54.67	13.96%	15.00	4.45%	69.67	17.79%	28.22	8.06%

### 4.2.3 All Users

Figure 9 provides a summary of Benchmark 1, pass 3 elapsed navigation times for all users elapsed times in all four environments. As shown in the chart, while the stereoscopic voice/glove environment elapsed times were somewhat higher than all of the other environments and the stereoscopic wand/joystick environment resulted in somewhat lower navigation times. As is shown with the analysis of the Benchmark 2 and 3 tests (to follow), a possible reason for this is that the somewhat more complicated voice/glove interface required longer period of exposure for users to become proficient. It should also be noted all users performed better using the stereoscopic environments (Wand and Glove) as a group over the non-stereoscopic environments.



**Figure 9 – B1Allp3Tim - All User Pass 3 Navigation Times**

Table 6 presents the improvements in navigation times for experienced users with each successive exposure to each of the four test environments. Note that, as a group, all users show little appreciable improvement in navigation times across the 3 passes with the stereoscopic voice/glove or non-stereoscopic traditional workstation environments. However, all users improved navigation speeds by 31.5% from pass 1 to pass 3.

**Table 6 – B1All Pass-to-Pass Elapsed Time Changes**

Benchmark 1 - All Users - Pass-to-Pass Elapsed Time Changes								
B1A Δ Improve	Pass 1 to Pass 2		Pass 2 to Pass 3		Pass 1 to Pass 3		Average to Pass 3	
	Sec.	%age	Sec.	%age	Sec.	%age	Sec.	%age
Glove	5.47	1.60%	9.50	2.82%	14.97	4.37%	7.50	2.24%
Wand	53.00	17.42%	43.00	17.11%	96.00	31.54%	46.33	18.19%
86" Non	25.67	6.37%	55.67	14.77%	81.33	20.20%	45.67	12.44%
W/S	54.67	13.96%	15.00	4.45%	69.67	17.79%	28.22	8.06%

### 4.3 Elapsed Times Detailed Statistical Analysis

As discussed in section 3.5 (Role of Statistics), all statistical analyses of the test data were performed using a combination of two software packages: GraphPad Prism version 4.0 [94] and SAS for Windows version 8.0 [99]. Interpretation of the program results was made with assistance of Motulsky's Prism Guide [102] and the SAS User's Guide [95]. The software reported the results of normality testing (for Gaussian distribution), the coefficient of variance, and one-tailed, repeated measures, ANOVA tests (of means) which were used to do pair-wise comparisons of the environments.

The tests applied to the data are in accordance with the logic presented in Figure 6, starting with descriptive statistic testing performed to determine if the sets of environment data are normally distributed (Gaussian distribution). Such testing (Kolmogorov-Smirnov normality test) quantified and reported the discrepancy between the distribution of the data and the ideal Gaussian distribution. The results presented here were based on the means and standard deviations of each set of Benchmark, environment, and test pass sample results. As discussed in Chapter 3, these conventions have been consistently used in the reporting of the statistical results.

The P value from the normality test answers the question: In a random sample from a Gaussian distribution, what is the probability (P value) of obtaining a sample that deviates as much from a Gaussian distribution (or more so) than the given sample? Stated differently, the P value answers the question: *If the population is Gaussian, what is the chance (as measured by probability) that a randomly selected sample of this size would have a statistic as large, or larger, than observed?*

Since the sample sizes for this study are relatively small (15 novice and 15 experienced users), a large P value only means that the data are not consistent within a

Gaussian (normal) population. This does not exclude the possibility of a non-Gaussian population for which nonparametric tests are applicable. Small sample sizes simply do not provide enough data to accurately discriminate between normal and non-Gaussian distributions. However, the one-way ANOVA method is known to be tolerant of moderate departures from the assumption of normality [103].

SAS and Prism use one-way analysis of variance (ANOVA), and corresponding nonparametric tests (Friedman-Dunn Sum Ranks), to test whether the mean (or median) of a variable differs among the four groups (environments) for this testing. ANOVA tests whether there is statistical significance among the means of more than two groups, i.e. it tests the variability among group means by using computed values (sum of squares, degrees of freedom, and means squares) to produce the F-ratio and the significant value. The F-ratio is a measure of variation between the data for different interfaces (W/S, 86" non, wand, and glove) and variation, within the data for the interfaces. The larger this value, the greater the chance that the differences between the means are due to real effects rather than chance alone. For the F-ratio to be considered statistically significant, a confidence level of 90% is used with a corresponding alpha level of 10%. If the result is significant, the null hypothesis ( $H_0$ ) can be rejected, and the alternative hypothesis ( $H_a$ ) can be accepted ( $H_a$ : 1<sup>st</sup> environment elapsed time < 2<sup>nd</sup> environment elapsed time).

The limitation of ANOVA tests is that it does not make multiple pair-wise comparisons. Therefore, additional testing after the ANOVA calculations (post hoc) is required. The software provides post hoc tests for both normally distributed data sets and for data sets that are non-Gaussian in distribution.

At a standard alpha level of 10% (90% confidence interval) a significant value of less than 0.10 would indicate that the two means are statistically different; the null hypothesis ( $H_0$ , 1<sup>st</sup> environment elapsed time  $\geq$  2<sup>nd</sup> environment elapsed time) is rejected; and the alternative hypothesis ( $H_a$ , 1<sup>st</sup> environment elapsed time < 2<sup>nd</sup> environment elapsed time) is accepted, allowing the conclusion that the environment with the lower elapsed execution times is shown to be statistically different.

Multiple pair-wise comparisons of data sets are performed using either standard ANOVA (parametric) or Friedman-Dunn (nonparametric) post tests to provide the difference of the means or medians, the F value, the P value, and 90% confidence interval for the comparisons.  $H_a$  true indicates that there is a 90% confidence that the means are not equal

and thus navigation in the two environments are statistically different, and by analysis, the environment producing lower elapsed times is “better.” This constitutes a statistically significant proof of different means for the data.

#### 4.3.1 Benchmark 1 - Pass 3 Statistics

##### 4.3.1.1 B1Np3Tim - Novice User Timing Statistics

Table 7 presents the B1Np3Tim (Benchmark 1, novice user, pass 3, elapsed times) statistics and ANOVA comparison of the environments. Review of the data set distribution shows that the results for all of the environments are normally distributed and that both Bartlett’s Test and the coefficients of variance show that the variance of the data sets are relatively equal. Therefore, a one-tailed, paired measures ANOVA parametric test method is applied using a 90% confidence interval. As discussed in Section 3.5, in such tests, if the F statistic calculated for the pair of environments being compared exceeds the F statistic calculated by any of the multivariate generalizations (reported here as Roy’s Greatest Root), then the results of the pair-wise environment comparisons may be tested for significance using standard p-value calculations (Bartlett’s or Levene’s test for variance and ANOVA for equality of means) as described in Figure 6. Statistically significant environment pair-wise comparisons allow a conclusion that the environment recording the lower elapsed time is the statistically “better” environment. Specifically, Table 7 shows that for novice user third pass Benchmark 1 (navigation) tests, the ANOVA comparisons show, with statistical significance, that the stereoscopic wand interface produces faster navigation times by an average of 17% over any of the other environments.

**Table 7 – B1Np3Tstat - Pass 3 Novice User Timing Statistics**

Benchmark 1, Novice Users, Pass 3, Elapsed Time:											
B1Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	212.3	30.64	175.0	270.0	> 0.1000	Yes	14.43%			
Wand	15	171.9	37.11	115.0	245.0	> 0.1000	Yes	21.59%			
86"Non	15	194.5	48.14	95.0	260.0	> 0.1000	Yes	24.75%			
W/S	15	196.4	36.80	125.0	250.0	> 0.1000	Yes	18.74%			
B1Np3		Homogeneity of Variance			Roy's Greatest Root			F(3,12)		Pr > F	
		Bartlett's Test		Levene's Test		Equal Var ?		df		F(1,14)	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					Pr > F	
Glove vs Wand		0.4924	0.4828			Yes	14	10.23	0.0064	No	Means
Glove vs 86"Nor		2.7107	0.0997			Yes	14	1.75	0.2069	Yes	Neither
Glove vs W/S		0.4509	0.5019			Yes	14	1.17	0.2969	Yes	Neither
Wand vs 86"Non		0.9305	0.3347			Yes	14	3.76	0.0729	No	Means
Wand vs W/S		0.0009	0.9757			Yes	14	6.72	0.0213	No	Means
86"Non vs W/S		0.9893	0.3199			Yes	14	6.72	0.0213	No	Means
											86" Non
											0.98%

#### 4.3.1.2 B1Ep3Tim - Experienced User Timing Statistics

Table 8 presents the B1Ep3Tim (Benchmark 1, experienced user, pass 3, elapsed times) statistics and ANOVA comparison of the environments. Review of the data set distribution shows that the results for all but the 86” non-stereoscopic environment are normally distributed. Thus, Bartlett’s Test is used to determine the homoscedasticity of environment comparisons not including the 86” non-stereoscopic interface and Levene’s Test is used for those including the 86” non-stereoscopic interface. The one-tailed, paired measures ANOVA parametric test (using a 90% confidence interval) is used to determine the “better” of each of the paired environments. As discussed in Section 3.5, in such tests, if the p-value reported is means of determining if the elapsed time means of the two environments are equal. As shown in Figure 6, if either the variances are unequal or the means are unequal, the comparison is statistically significant and the environment recording the lower elapsed time is statistically the “better” environment. In this specific test, only the comparisons of the stereoscopic wand interface show statistically significant results that indicate that navigation using the stereoscopic wand interface average 55.3 faster than any of the other environments.

**Table 8 – B1Ep3Tstat - Pass 3 Experienced User Timing Statistics**

Benchmark 1, Experienced Users, Pass 3, Elapsed Times											
B1Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	327.40	151.90	155.00	685.00	> 0.1000	Yes	46.40%			
Wand	15	208.30	68.47	130.00	395.00	> 0.1000	Yes	32.87%			
86"Non	15	321.30	148.60	185.00	750.00	0.0007	No	46.25%			
W/S	15	322.00	141.20	175.00	735.00	0.0615	No	43.85%			
B1Ep3		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	12.74	Pr > F	0.0005
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand		7.8080	0.0052			No	14	10.32	0.0063	No	Var/Mean
Glove vs 86"Non				0.0000	0.9518	Yes	14	0.02	0.8803	Yes	Neither
Glove vs W/S				1.2800	0.2620	Yes	14	0.00	0.9117	Yes	Neither
Wand vs 86"Non				1.8500	0.0064	No	14	19.30	0.0006	No	Var/Mean
Wand vs W/S				4.7000	0.0389	No	14	24.55	0.0002	No	Var/Mean
86"Non vs W/S				0.0200	0.8984	Yes	14	0.00	0.9788	Yes	Neither

#### 4.3.1.3 B1Allp3Tim - All Users Timing Statistics

Table 9 presents the B1Allp3Tim statistics and post-hoc test results. As can be seen in Table 9, none of the environments produce normally distributed datasets. Therefore, nonparametric test methods (Friedman-Dunn) are applied. As discussed in Section 3.5, the Friedman-Dunn tests are nonparametric tests to perform pair-wise comparisons of multiple environments. The test analyzes only the differences between the pared measurements for each environment. The P-value answers the question: *If the median difference really is zero*



overall, what is the chance that random sampling would result in a median difference as far from zero (or more so) as observed in the test results? [97] If the P-value is small, one rejects the idea that the difference is a coincidence, and concludes instead that the populations have different means. If the P-value is large, the data do not give any reason to conclude that the overall medians differ. This is not the same as saying that the medians are the same, only that there is insufficient evidence that they differ.

Examination of the results posted in Table 9 show that the only the stereoscopic wand environment comparisons provide statistically significant results. These results show that the stereoscopic wand interface proves to be statistically better than any of the other environments. Specifically, both user groups found the stereoscopic wand interface to perform the navigation functions an average of 38.25% faster than using the other interfaces.

**Table 9 – B1Allp3Tstat - Pass 3 All Users Timing Statistics**

Benchmark 1, All Users, Pass 3, Elapsed Times											
B1Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	30	269.90	122.60	155.00	685.00	< 0.0001	No	45.42%			
Wand	30	190.10	57.19	115.00	395.00	0.0311	No	30.08%			
86"Non	30	257.90	126.40	95.00	750.00	< 0.0001	No	49.01%			
W/S	30	258.70	120.10	125.00	735.00	0.0012	No	46.42%			
B1Ap3		Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	10.89	Pr > F	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				3.4100	0.0697	Yes	29	15.01	0.0006	No	Means
Glove vs 86"Non				0.0100	0.9278	Yes	29	0.34	0.5659	Yes	Neither
Glove vs W/S				0.0000	0.9519	Yes	29	0.20	0.6552	Yes	Neither
Wand vs 86"Non				2.2400	0.1400	Yes	29	17.84	0.0002	No	Means
Wand vs W/S				1.9700	0.1661	Yes	29	21.67	0.0001	No	Means
86"Non vs W/S				0.0200	0.8926	Yes	28	0.00	0.9563	Yes	Neither

#### 4.4 User Subjective Overall Environment Ratings

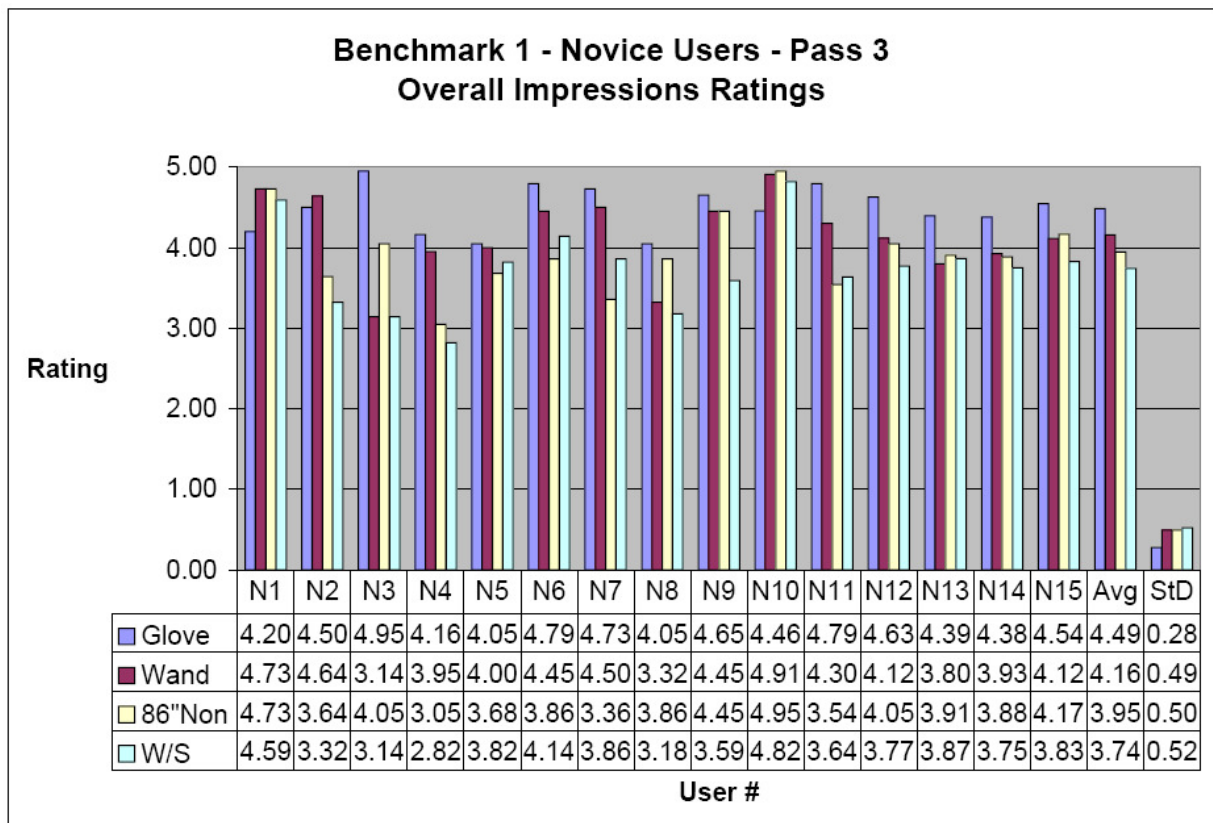
After completion of each pass of each Benchmark test in each environment, users provided their subjective views of their experience by completing the 22-question Usability Survey (see Figure 4) rating the environment on a scale of 1 to 5 (very poor to very good).

The survey questions were grouped in five areas (navigation, locating, manipulation, general, and overall impressions). What follows is a presentation of user overall impressions ratings of the interfaces for performing Benchmark 1 tasks (navigation) at the completion of the 3<sup>rd</sup> pass as a representation of user final evaluations of each interface. Detailed analysis of each of the five areas is presented in the appendices.

#### 4.4.1 Novice Users

As previously discussed, each novice user was asked to rate his/her experience via the Usability Survey at the completion of each pass of each Benchmark test. Figure 10 presents the overall impressions ratings of the novice users at the completion of the 3<sup>rd</sup> pass of the Benchmark 1 scenario. As such, this represents each user's final impression of the navigational capabilities of each environment.

A further examination of the results detailed in Figure 10 shows that at completion of the Benchmark tests, novice users preferred the stereoscopic voice/glove interface over all other environments. Analysis of the results presented in the chart shows that novice users rated the voice/glove interface 5.4% higher than the stereoscopic wand environment; 11% higher than the 86" non-stereoscopic environment; and 16% higher than the traditional CAD workstation interface.



**Figure 10 – B1Np3Ovr – Novice User Pass 3 Overall Impression Ratings**

Table 10 shows the pass-to-pass improvements in novice user overall impression ratings for each of the environments. Note that with each successive exposure (pass-to-pass)

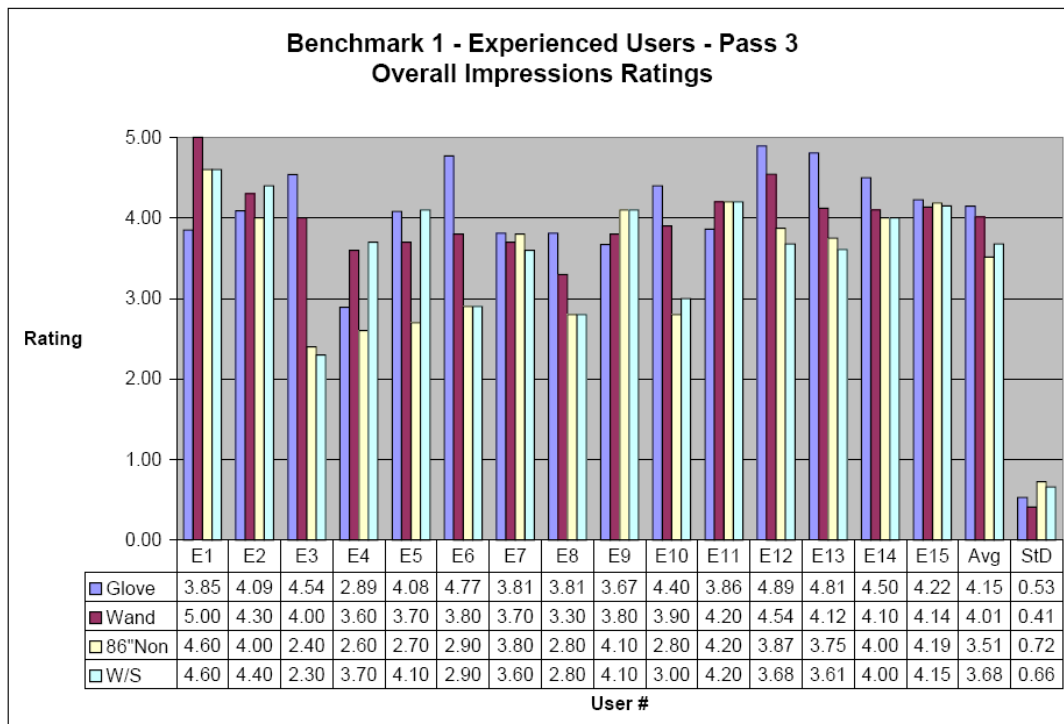
the user's overall impressions of the interfaces improved. Examination of the pass-to-pass analysis of improvements noted in Table 10 shows that novice user impressions of the stereoscopic interfaces improved 9.1% from pass 1 to pass 3 whereas the non-stereoscopic environments showed a 5.8% improvement from pass 1 to pass 3.

**Table 10 – B1N – Novice User Pass-to-Pass Overall Impression Ratings Changes**

Novice User Benchmark 1 Pass-to-Pass Overall Impressions Rating Changes								
B1N Overall	Pass 1 to Pass 2		Pass 2 to Pass 3		Pass 1 to Pass 3		Average to Pass 3	
Δ Improve	Rating	%tage	Rating	%tage	Rating	%tage	Rating	%tage
Glove	0.23	5.82%	0.27	6.46%	0.50	12.66%	0.26	6.21%
Wand	0.03	0.78%	0.19	4.68%	0.22	5.49%	0.14	3.33%
86"Non	0.10	2.55%	0.07	1.72%	0.16	4.32%	0.08	1.98%
W/S	0.12	3.40%	0.14	3.82%	0.26	7.34%	0.13	3.63%

#### 4.4.2 Experienced Users

As with the novice user group, each experienced user was asked to rate his/her experience via the Usability Survey at the completion of each pass of each Benchmark test. Figure 11 presents the overall impressions ratings of the experienced users at the completion of the 3<sup>rd</sup> pass of the Benchmark 1 scenario. As such, this represents each user's final impression of the navigational capabilities of each environment.



**Figure 11 – B1Ep3Ovr – Experienced User Pass 3 Overall Impression Ratings**

Analysis of the results noted in Figure 11 shows that the experienced users preferred the stereoscopic interfaces to the non-stereoscopic environments by 12.2%. Experienced users slightly favoured the voice/glove stereoscopic interface over the wand interface by only 3%.

Table 11 shows the pass-to-pass improvements in experienced user overall impression ratings for each of the environments. Examination of the results presented show that for the experienced user group the pass-to-pass improvements in overall environment impressions averaged 7.35% from pass 1 to pass 3 and that the stereoscopic environments show a 5% better improvement from pass 1 to pass 3 over the non-stereoscopic interfaces.

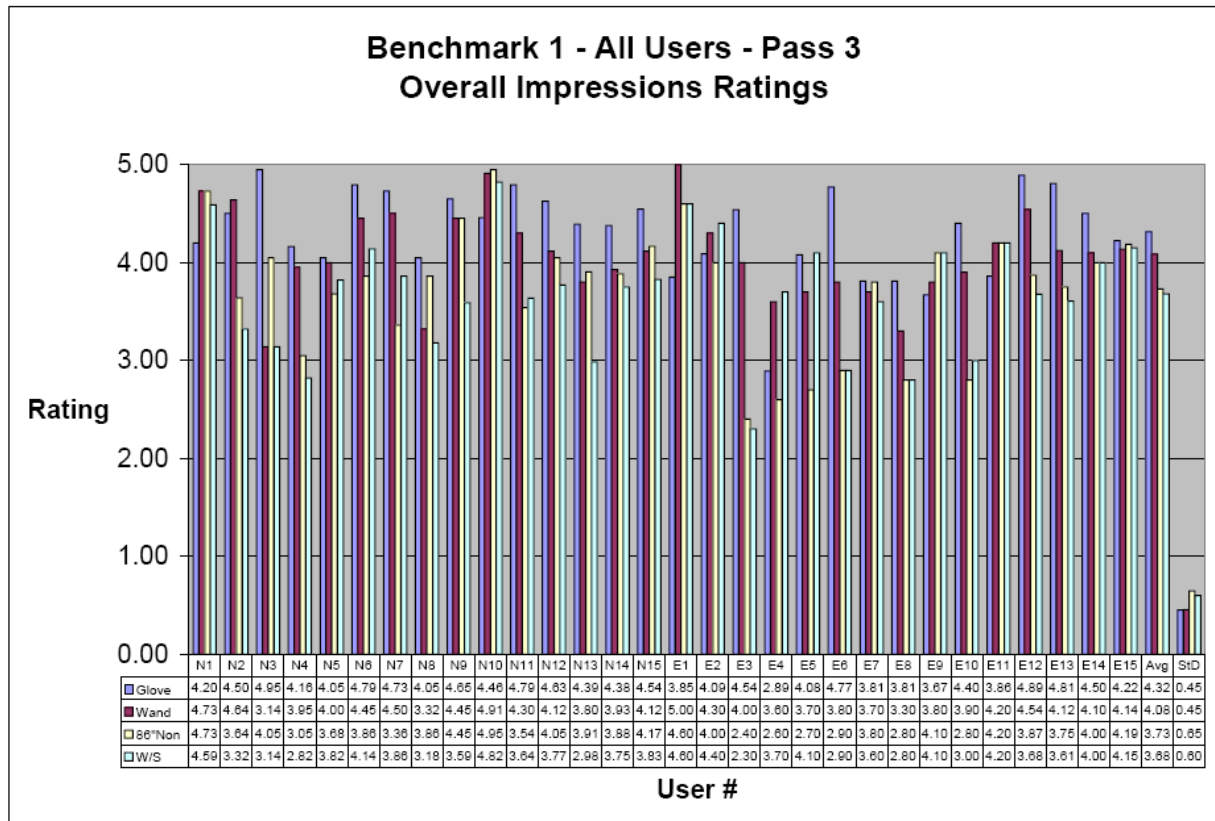
**Table 11 – B1E – Experienced User Pass-to-Pass Overall Impression Ratings Changes**

Benchmark 1 - Experienced Users - Pass-to-Pass Overall Impressions Ratings Changes								
B1E Overall Δ Improve	Pass 1 to Pass 2		Pass 2 to Pass 3		Pass 1 to Pass 3		Average to Pass 3	
	Rating	%tage	Rating	P2 to P3	Rating	P1 to P3	Rating	Avg to P3
Glove	0.06	1.60%	0.22	5.33%	0.28	7.01%	0.16	4.00%
Wand	0.00	0.00%	0.32	8.08%	0.32	8.08%	0.21	5.25%
86"Non	0.06	1.90%	0.18	5.32%	0.25	7.32%	0.14	4.12%
W/S	-0.07	-2.02%	0.33	9.25%	0.25	7.05%	0.19	5.28%

#### 4.4.3 All Users

Figure 12 presents the overall impressions ratings of all users at the completion of the 3<sup>rd</sup> pass of the Benchmark 1 scenario. As such, this represents each user's final impression of the navigational capabilities of each environment. Analysis of the results presented in the chart shows that all users rated the voice/glove interface slightly higher than the stereoscopic wand environment and approximately 1.2% higher than either of the non-stereoscopic environments.

Table 12 shows the pass-to-pass improvements in the overall impression ratings for each of the environments for all the users. Examination of the pass-to-pass analysis of improvements noted in Table 12 shows that user impressions of the stereoscopic voice glove interface improved 9.84% from pass 1 to pass 3 and this improvement rate shows to be about 60% higher than any of the other environments.



**Figure 12 – B1Allp3Ovr – All Users Pass 3 Overall Impression Ratings**

**Table 12 – B1All – Pass-to-Pass Overall Impression Ratings Changes**

All Users Benchmark 1 Pass-to-Pass Overall Impressions Rating Changes								
B1A Overall	Pass 1 to Pass 2		Pass 2 to Pass 3		Pass 1 to Pass 3		Average to Pass 3	
Δ Improve	Rating	%tage	Rating	%tage	Rating	%tage	Rating	%tage
Glove	0.15	3.71%	0.24	5.89%	0.39	9.84%	0.21	5.10%
Wand	0.02	0.39%	0.25	6.38%	0.27	6.79%	0.17	4.29%
86\"Non	0.08	2.23%	0.12	3.52%	0.20	5.82%	0.11	3.05%
W/S	0.02	0.69%	0.23	6.53%	0.25	7.20%	0.16	4.46%

#### 4.5 User Environment Overall Ratings Statistics

The following sections present a statistical analysis of user subjective ratings of each of the four test environments following pass 3. These ratings represent the user’s final overall impressions of the Benchmark 1 (navigational) characteristics of the environments. Details of the results of pass 1, pass 2, and 3 pass average user ratings for each of the separate survey areas (navigation, manipulation, general impressions) that makeup the overall impressions ratings are presented in the appendices.

#### 4.5.1 B1Np3R – Novice User Overall Impressions Statistics

Table 13 presents the statistics for each of the Benchmark 1 novice user overall rating of each environment following their final (pass 3) exposure and as such represents each novice user's overall impression of the interfaces. As noted in the discussion of the statistical methods used to analyze user Benchmark elapsed times, the P value is greater than the alpha of 0.10 thus indicating that each data set is normally distributed. Normal distribution and the relatively equal variances of the data sets (as shown by both the Bartlett's Test results and the coefficients of variance) indicate that parametric, one-tailed, paired samples ANOVA comparisons provide the information needed to determine the statistical significance of the test results.

Examination of the ANOVA results posted in Table 13 shows that there is sufficient statistical evidence to say that the stereoscopic voice/glove environment is preferred by the novice user group by an average of 17.74% over either non-stereoscopic interface and that the stereoscopic wand environment is preferred by the novice user group by an average of 9.2% over the either non-stereoscopic interface. It should also be noted that the stereoscopic voice/glove interface is preferred over the wand interface by 7.89%.

**Table 13 – B1Np3Ovr – Novice Users Overall Impressions Statistics**

Benchmark 1, Novice Users, Pass 3, Overall Impressions Ratings												
B1Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	15	4.485	0.280	4.05	4.95	> 0.1000	Yes	6.25%				
Wand	15	4.157	0.494	3.14	4.91	> 0.1000	Yes	11.89%				
86"Non	15	3.945	0.496	3.05	4.95	> 0.1000	Yes	12.57%				
W/S	15	3.683	0.557	2.82	4.82	> 0.1000	Yes	15.11%				
B1Np3	Homogeneity of Variance					Roy's Greatest Root:			F(3,12)		9.09	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Pr > F	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							0.0021	
Glove vs Wand	4.1398	0.0419			No	14	5.37	0.0361	No	Var/Mean	Glove	7.89%
Glove vs 86"Non	6.2003	0.0450			No	14	13.85	0.0023	No	Var/Mean	Glove	13.69%
Glove vs W/S	5.9180	0.0150			No	14	26.04	0.0002	No	Var/Mean	Glove	21.78%
Wand vs 86"Non	0.0000	0.9907			Yes	14	2.05	0.1743	Yes	Neither	Wand	5.37%
Wand vs W/S	0.1891	0.6636			Yes	14	20.33	0.0213	No	Means	Wand	12.87%
86"Non vs W/S	0.1791	0.6721			Yes	14	5.49	0.0344	No	Means	86" Non	7.11%

#### 4.5.2 B1Ep3Ovr – Experienced User Overall Impressions Statistics

Table 14 presents the results of the statistical tests performed on the results of the experienced user pass 3 Benchmark 1 tests. As expected, these results are similar to those of the novice users. The statistical methods used to analyze user Benchmark elapsed times, for all but the 86" non-stereoscopic environment, the median of each of the environment falls within 90%

level and thus each data set is normally distributed. Thus, Bartlett's test (parametric) is used to determine the homogeneity of variance for all comparisons without the 86" non-stereoscopic environment and Levene's test (non-parametric) is used for comparisons with the 86" non-stereoscopic environment.

**Table 14 – B1Ep3Ovr – Experienced Users Overall Impression Statistics**

Benchmark 1, Experienced Users, Pass 3, Overall Impressions Ratings												
B1Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	15	4.146	0.531	2.89	4.89	> 0.1000	Yes	12.80%				
Wand	15	4.013	0.413	3.30	5.00	> 0.1000	Yes	10.28%				
86"Non	15	3.514	0.723	2.40	4.60	0.0349	No	20.58%				
W/S	15	3.676	0.659	2.30	4.60	> 0.1000	Yes	17.92%				
B1Ep3	Homogeneity of Variance				Roy's Greatest Root: F(3,12)			3.71	Pr > F 0.0425			
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand	0.8466	0.3575			Yes	14	0.84	0.3751	Yes	Neither	Neither	
Glove vs 86"Non			2.6100	0.1171	Yes	14	7.67	0.0151	No	Means	Glove 17.99%	
Glove vs W/S	0.6266	0.4286			Yes	14	3.57	0.0798	No	Means	Glove 12.79%	
Wand vs 86"Non			8.1100	0.0081	No	14	12.97	0.0029	No	Var/Means	Wand 14.20%	
Wand vs W/S	2.8550	0.1000			Yes	14	5.31	0.0371	No	Means	Wand 9.17%	
86"Non vs W/S			0.2700	0.6051	Yes	14	1.79	0.2026	Yes	Neither	Neither	

Examination of the ANOVA results posted in Table 14 shows that there is sufficient statistical evidence to say that the stereoscopic voice/glove environment preferred by the experienced users over either of the non-stereoscopic interfaces by an average of 15.39% and the stereoscopic wand interface is also preferred by the novice user group by an average of 11.69% over either non-stereoscopic environment. However, there is insufficient statistical significance to infer the preference of one stereoscopic interface over the other.

#### 4.5.3 B1Allp3Ovr – All Users Overall Impressions Statistics

Table 15 presents the results of the statistical analysis test of the overall impressions ratings of the entire user group at the completion of the 3<sup>rd</sup> pass of the Benchmark 1 scenario. As might be expected from the novice and experienced user ratings, the aggregate user ratings pass the normality test for all the test environments at the 90% confidence level, and since the coefficients of variability (and Bartlett's Test results) show relatively equal variances, the ANOVA comparisons are made using parametric, one-tailed, 90% confidence interval calculations.

Examination of the ANOVA results posted in Table 15 show that there is sufficient statistical significance to say that the stereoscopic voice/glove environment is preferred by all users by an average of 16.47% over either non-stereoscopic interface and that the stereoscopic wand environment is preferred by the all users by an average of 10.27% over

the either non-stereoscopic interface. It should also be noted that the stereoscopic voice/glove interface is preferred over the wand interface by 5.63%.

**Table 15 – B1Allp3Ovr – All Users Overall Impressions Statistics**

Benchmark 1, All Users, Pass 3, Overall Impressions Ratings												
B1Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	30	4.315	0.451	2.89	4.95	> 0.1000	Yes	10.45%				
Wand	30	4.085	0.453	3.14	5.00	> 0.1000	Yes	11.10%				
86"Non	30	3.730	0.648	2.40	4.59	> 0.1000	Yes	17.36%				
W/S	30	3.680	0.599	2.30	4.82	0.0998	Yes	16.28%				
B1Ap3		Homogeneity of Variance				Roy's Greatest Root:			F(3,27)		7.77	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		0.8466	0.3375			Yes	29	5.19	0.0303	No	Means	Glove 5.63%
Glove vs 86"Non		1.2766	0.2585			Yes	29	19.36	0.0001	No	Means	Glove 15.68%
Glove vs W/S		0.6266	0.4286			Yes	29	18.50	0.0002	No	Means	Glove 17.26%
Wand vs 86"Non		2.1347	0.1440			Yes	29	11.88	0.0018	No	Means	Wand 9.52%
Wand vs W/S		2.6999	0.1004			Yes	29	20.56	0.0001	No	Means	Wand 11.01%
86"Non vs W/S		0.1183	0.7309			Yes	29	0.31	0.5831	Yes	Neither	Neither

#### 4.6 Benchmark 1 Reliability

Table 16 presents the Cronbach's  $\alpha$  values computed for the objective measures (elapsed navigation times) and overall subjective user ratings for the Benchmark 1 scenario. Note that the alphas for the standardized variables (removal of duplicate values) are above 90%, indicating a high reliability for the benchmark. (Appendix M presents the detailed pass/environment results of the reliability calculations.)

**Table 16 – Benchmark 1 Cronbach's  $\alpha$  Reliability Coefficients**

Benchmark 1 - All Users Cronbach Coefficient Alpha			
Variables	Alpha	Variables	Alpha
Standardized	86.12%	Standardized	97.57%



## **5 Benchmark 2 (Find and Repair Manipulation)**

### ***5.1 Description***

Using the same AC&R space as used for Benchmark 1, for Benchmark 2 users were required to navigate through the space looking for “errors” that had been injected into the design. Typical “errors” were a passageway blocked by a misplaced control box or a misaligned section of pipe. Users were then required to “fix” the error. The “fix” required the user to utilize the interface (environment) under test, typically, re-positioning the part to a more suitable location/orientation. Elapsed times were noted for each activity. The elapsed time recorded was the time required to:

1. locate and identify the 1<sup>st</sup> error; plus
2. the time to “fix” the 1<sup>st</sup> error; plus
3. the time to locate and identify the 2<sup>nd</sup> error; plus
4. the time to “fix” the 2<sup>nd</sup> error; plus
5. the time to return to the starting position within the space.

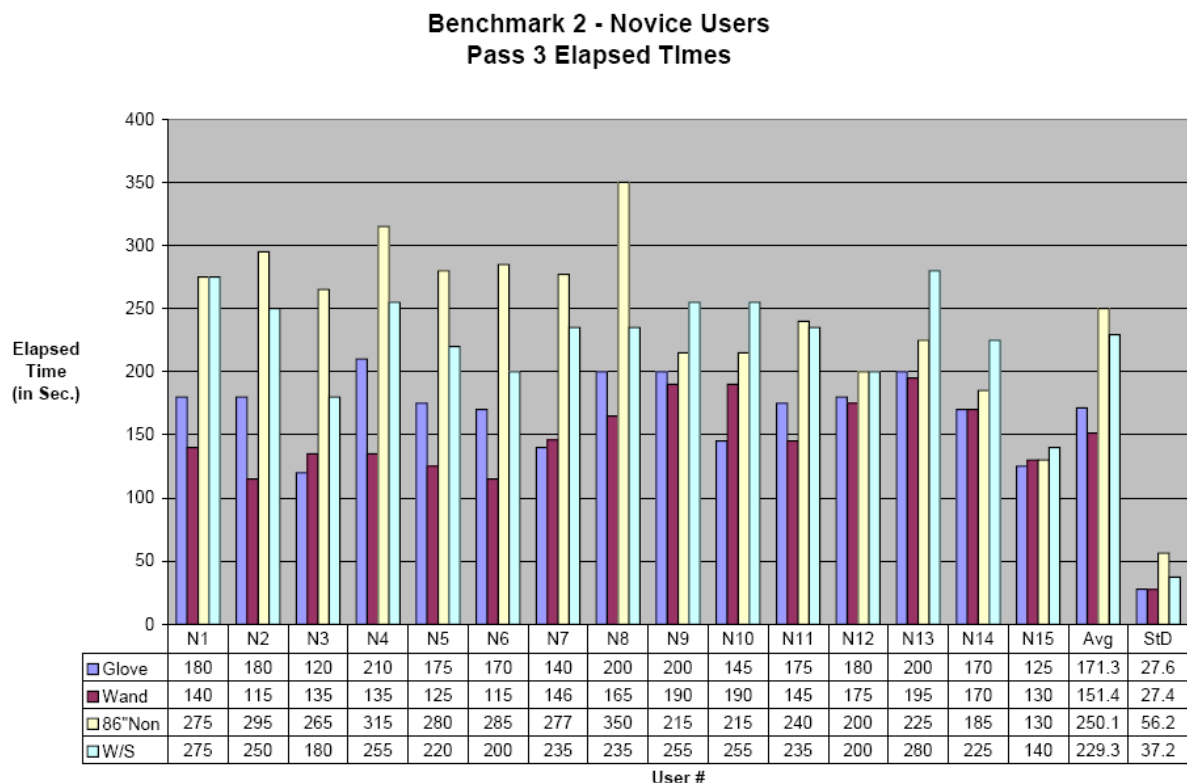
The find/repair exercise (Benchmark 2) was also repeated in each of the 4 environments under test and the User Survey administered to each user after each pass in each environment. As with the Benchmark 1 testing, sequencing of the testers through the four environments was randomized so that not all of the users tested in the same order. This randomization of interface paradigms avoided statistical bias in the results.

As reported for the Benchmark 1 tests results, the following sections detail the results of the final (pass 3) measures of user performance with each interface. Details of each of pass 1, pass 2, and 3 pass average results are reported in the appendices.

## 5.2 Pass-to-Pass Improvements in Elapsed Times

### 5.2.1 Novice Users

Figure 13 presents a representation of the elapsed times required by novice users to perform a typical set of find/repair operations as defined in the Benchmark 2 scenarios. The results presented are for the last (3<sup>rd</sup>) execution of the test. These times should represent the “best/fastest” execution times for the group. It should be noted that while both stereoscopic interfaces resulted in shorter execution times (as compared to the non-stereoscopic interfaces), as a group the novice users performed best using the voice/glove stereoscopic interface. Specifically, Figure 13 shows the wand/joystick environment to be 13% faster than the voice/glove environment; 65% faster than the 86” non-stereoscopic environment; and 51% faster than the traditional, desktop, CAD workstation environment.



**Figure 13 – B2Np3Tim – Novice User Pass 3 Manipulation Times**

Table 17 presents the improvements in find/repair (manipulation) times for novice users with each successive exposure to each of the four test environments. Note that there

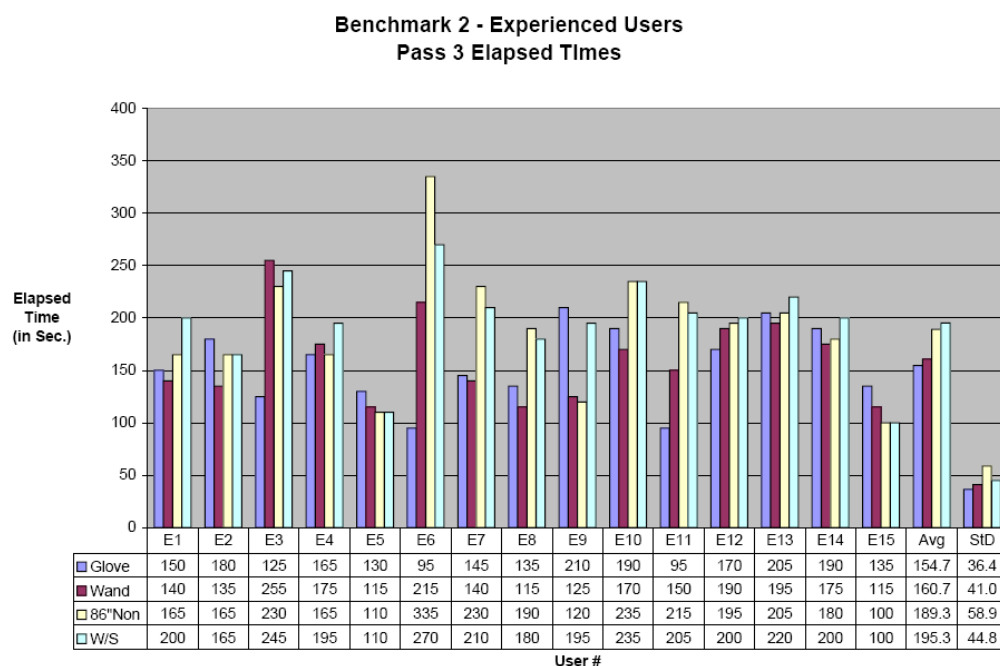
appears to be an average difference of 164% in the increase between stereoscopic environments versus non-stereoscopic environments for novice users. Thus, for novice users, initial exposure to manipulation methods show little improvement for any of the environments.

**Table 17 – B2N – Novice User Pass-to-Pass Elapsed Time Changes**

B2N Times	Pass 1 to Pass 2		Pass 2 to Pass 3		Pass 1 to Pass 3		Average to Pass 3	
$\Delta$ Improve	Rating	%tage	Rating	%tage	Rating	%tage	Rating	%tage
Glove	20.77	9.30%	31.23	15.42%	52.00	28.21%	-27.74	-18.21%
Wand	41.13	16.96%	50.00	24.83%	91.13	31.23%	-47.04	-25.19%
86"Non	2.03	0.75%	19.00	7.06%	21.03	14.45%	-13.34	-8.31%
W/S	41.67	14.64%	13.67	5.62%	55.33	21.84%	-23.00	-10.83%

### 5.2.2 Experienced Users

Figure 14 shows experienced user elapsed times for pass 3 of the manipulation Benchmark tests in all four environments. As with the novice users, the experienced users performed the find/repair tasks of Benchmark 2 faster in the stereoscopic environments. Comparing the stereoscopic interfaces shows them to be almost equal with the voice/glove interface being approximately 3.9% faster than the wand/joystick interface. However, as with the novice users, the stereoscopic environments show significant reductions in find/repair task times (as much as 26.2%).



**Figure 14 – B2Ep3Tim - Experienced User Pass 3 Manipulation Times**

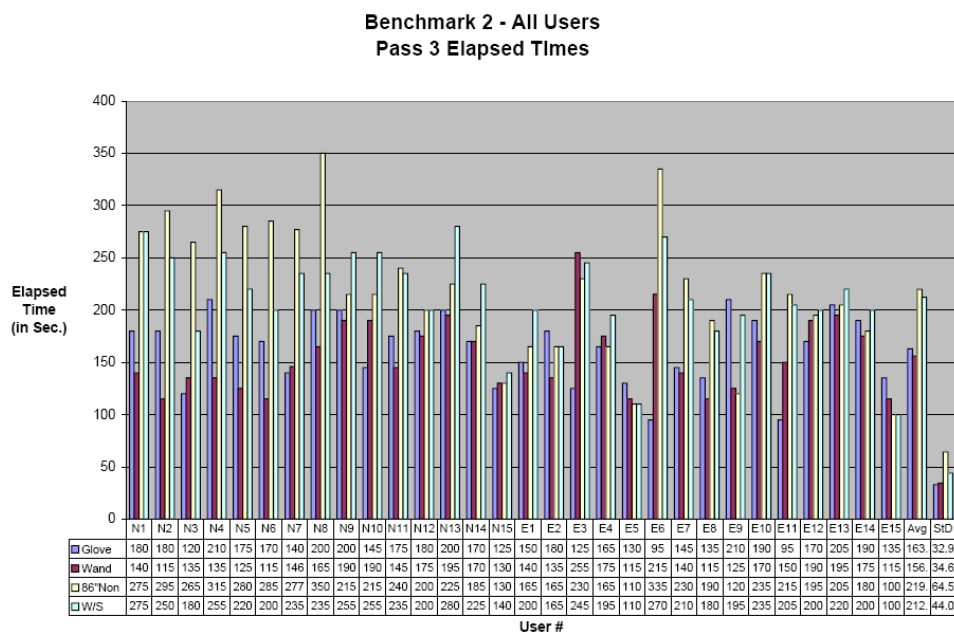
Table 18 presents the improvements in find/repair (manipulation) times for experienced users with each successive exposure to each of the four test environments. Note that experienced users show slightly better improvement rates using the stereoscopic voice/glove interface than with any of the other environments. Note also that, on average, the stereoscopic environments show a 1.2 times higher improvement in find/repair actions (pass 1 to pass 3) than non-stereoscopic interfaces.

**Table 18 – B2E – Experienced User Pass-to-Pass Elapsed Time Changes**

B2E Times	Pass 1 to Pass 2		Pass 2 to Pass 3		Pass 1 to Pass 3		Average to Pass 3	
Δ Improve	Rating	%tage	Rating	%tage	Rating	%tage	Rating	%tage
Glove	44.00	19.24%	30.00	16.25%	74.00	32.36%	-34.67	-22.41%
Wand	33.00	14.93%	27.33	14.54%	60.33	27.30%	-29.22	-18.19%
86"Non	30.00	12.16%	27.33	12.62%	57.33	23.24%	-28.22	-14.91%
W/S	58.00	21.94%	11.00	5.33%	69.00	26.10%	-26.67	-13.65%

### 5.2.3 All Users

Figure 15 provides a summary of Benchmark 2, pass 3 elapsed find/repair elapsed times for all users in all four environments. As shown in the chart, the stereoscopic wand environment elapsed times were slightly faster than the stereoscopic voice/glove interface (4.48%) and that the stereoscopic interfaces produced significantly reduced find/repair elapsed times (35.1%) over the non-stereoscopic interfaces.



**Figure 15 – B2Allp3Tim - All User Pass 3 Manipulation Times**

Table 19 presents the improvements in find/repair (manipulation) times for all users with each successive exposure to each of the four test environments. Note that stereoscopic environments resulted in sharper decreases in find/repair times than for non-stereoscopic environments. On average, for all users, find/repair times in stereoscopic environments proved to be 1.6 times faster (pass 1 to pass 3) than for non-stereoscopic interfaces.

**Table 19 – B2A – All Users Pass-to-Pass Elapsed Time Changes**

<b>B2A Times</b>	<b>Pass 1 to Pass 2</b>		<b>Pass 2 to Pass 3</b>		<b>Pass 1 to Pass 3</b>		<b>Average to Pass 3</b>	
<b>Δ Improve</b>	<b>Rating</b>	<b>%tage</b>	<b>Rating</b>	<b>%tage</b>	<b>Rating</b>	<b>%tage</b>	<b>Rating</b>	<b>%tage</b>
<b>Glove</b>	32.38	14.33%	30.62	16.73%	63.00	27.88%	-31.21	-19.14%
<b>Wand</b>	37.07	15.99%	38.67	19.04%	75.73	32.68%	-38.13	-24.44%
<b>86"Non</b>	16.02	6.19%	23.17	6.59%	39.18	15.13%	-20.78	-9.46%
<b>W/S</b>	49.83	18.15%	12.33	22.18%	62.17	22.65%	-24.83	-11.70%

### ***5.3 Elapsed Times Detailed Statistical Analysis***

As described for the Benchmark 1 testing, all statistical analyses of the test data were performed using two commercial standard statistical analysis software packages. GraphPad's Prism version 4.0 [94] provides an excellent GUI (Graphical User Interface) and an excellent guide to interpreting normality testing, coefficients of variance, and Bartlett's tests results (Motulsky's Prism Guide [102]). Additionally, the SAS software reports the results of standard ANOVA testing for one-way, repeated measures ANOVA [98]. The analyses reported follow the mechanism detailed in the flowchart presented in Figure 6.

#### **5.3.1 Benchmark 2 - Pass 3 Statistics**

In the sections following, the Benchmark 2 (Find/Repair) measures are analyzed using the statistical tests previously discussed. The test results are presented and as an analysis of the results are included in the discussion. As described for Benchmark 1 analysis, only the pass 3 results are detailed as the most current (and final) representation of the user's activities. Detailed results for all previous passes (and averages) are provided in the appendices.

### 5.3.1.1 B2Np3Tim - Novice User Timing Statistics

Table 20 presents the B2Np3Tim (Benchmark 2, novice users, pass 3 timings) statistics as reported by the software discussed in Section 3.5 and Figure 6. Since all four pass 3 novice user environments data sets are normally distributed and the coefficients of variance are relatively low, the repeated measures ANOVA parametric test methods are applied. Examination of these results as presented in Table 20 show that for this benchmark scenario there is sufficient statistical significance in the results to suggest that the novice user stereoscopic interfaces averages 30.25% shorter find/repair times than either of the non-stereoscopic interfaces. There is also sufficient statistical evidence to infer that the stereoscopic wand interface is faster than the stereoscopic glove environment by 9.98%.

**Table 20 – B2Np3Tstat – Novice User Pass 3 Novice Users - Timing Statistics**

Benchmark 2, Novice Users, Pass 3, Elapsed Times											
B2Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	171.3	27.61	120.0	210.0	0.0628	Yes	16.12%			
Wand	15	154.2	27.40	115.0	195.0	>0.1000	Yes	17.77%			
86"Non	15	250.0	56.22	130.0	350.0	>0.1000	Yes	22.49%			
W/S	15	229.3	37.22	140.0	280.0	>0.1000	Yes	16.23%			
B2Np3	Homogeneity of Variance				Roy's Greatest Root:				F(3,12) 27.26 Pr > F <0.0001		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.0008	0.9776			Yes	14	5.35	0.0001	No	Means	Wand 9.98%
Glove vs 86"Non	6.3265	0.0119			No	14	34.96	0.0001	No	Var/Means	Glove 31.48%
Glove vs W/S	1.1887	0.2758			Yes	14	63.99	0.0001	No	Means	Glove 25.29%
Wand vs 86"Non	6.4531	0.0111			No	14	29.23	0.0001	No	Var/Means	Wand 31.48%
Wand vs W/S	1.2482	0.2639			Yes	14	69.44	0.0001	No	Means	Wand 32.75%
86"Non vs W/S	2.2377	0.1347			Yes	14	2.26	0.1546	Yes	Neither	Neither

### 5.3.1.2 B2Ep3Tim - Experienced User Timing Statistics

Table 21 presents the B2Ep3Tim (Benchmark 2, experienced users, pass 3 timings) statistical analysis results. Examination of the normality test data presented in the table shows that all but the 19", non-stereoscopic traditional CAD timings data sets are normally distributed. Therefore, parametric methods (Bartlett's test) are used to determine the homoscedasticity for all of the comparisons except those including the non-stereoscopic, traditional CAD environment. The pair-wise comparisons including the non-stereoscopic, traditional CAD interface require non-parametric methods (Levene's test).

Examination of these results as presented in Table 21 shows that for there is sufficient statistical significance in the results to suggest that the experienced user stereoscopic wand interface averages 16.42% shorter find/repair times than either of the non-stereoscopic interfaces. However, it should be noted that there is only additional

sufficient statistical evidence to infer any that the stereoscopic voice/glove interface is faster than the 19” non-stereoscopic traditional CAD interface.

**Table 21 – B2Ep3Tstat – Experienced Users Pass 3 Experienced Users Timing Statistics**

Benchmark 2, Experienced Users, Pass 3, Elapsed Times												
B2Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	154.7	36.42	95.0	210.0	> 0.1000	Yes	23.54%				
Wand	15	160.7	41.01	115.0	255.0	> 0.1000	Yes	25.52%				
86"Non	15	189.3	58.85	100.0	335.0	> 0.1000	Yes	31.09%				
W/S	15	195.3	44.82	100.0	270.0	0.0313	No	22.95%				
B2Ep3	Homogeneity of Variance				Roy's Greatest Root: F(3,12)				7.60	Pr > F	0.0041	
	Bartlett's Test		Levene's Test		Equal Var ?		df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand	0.1896	0.6633			Yes	29	0.17	0.6894	Yes	Neither	Neither	
Glove vs 86"Non	2.9988	0.0833			No	29	2.86	0.1128	Yes	Var	Glove	18.28%
Glove vs W/S			0.5800	0.4531	Yes	29	7.34	0.0169	No	Means	Glove	20.79%
Wand vs 86"Non	1.7273	0.1888			Yes	29	6.41	0.0239	No	Means	Wand	15.11%
Wand vs W/S			0.1000	0.7522	Yes	29	19.60	0.0006	No	Means	Wand	17.72%
86"Non vs W/S			0.7000	0.4106	Yes	29	0.59	0.4568	Yes	Neither	Neither	

### 5.3.1.3 B2Allp3Tim – All Users Timing Statistics

Results presented in Table 22 show that the elapsed times values for the aggregated novice and experienced users are normally distributed for only the 86” non-stereoscopic environment. Therefore, non-parametric methods (Levene’s test) are used to determine the homoscedasticity of the pair-wise comparisons. As reported in Table 22, the Friedman-Dunn tests indicate that there is sufficient statistical significance to infer that for novice users the stereoscopic interfaces average 25.89% faster find/repair times than the non-stereoscopic environments.

**Table 22 – B2Allp3Tstat – All Users Pass 3 All Users Timing Statistics**

Benchmark 2, All Users, Pass 3, Elapsed Times												
B2Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	163.0	32.87	95.0	210.0	0.07901	No	20.17%				
Wand	30	156.0	34.59	115.0	255.0	0.0948	No	22.17%				
86"Non	30	219.7	64.45	100.0	350.0	> 0.1000	Yes	29.34%				
W/S	30	212.3	44.02	100.0	280.0	0.0975	No	20.73%				
B2Ap3	Homogeneity of Variance				Roy's Greatest Root: F(3,27)				22.93	Pr > F	<0.0001	
	Bartlett's Test		Levene's Test		Equal Var ?		df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			1.3289	0.2590	Yes	29	0.06	0.4305	Yes	Neither	Neither	
Glove vs 86"Non			6.7750	0.0062	No	29	19.99	0.0001	No	Var/Means	Glove	25.81%
Glove vs W/S			1.5500	0.2240	Yes	29	34.95	0.0001	No	Means	Glove	23.22%
Wand vs 86"Non			4.8379	0.0319	No	29	26.39	0.0001	No	Var/Means	Wand	28.99%
Wand vs W/S			1.6550	0.2092	Yes	29	60.90	0.0001	No	Means	Wand	26.52%
86"Non vs W/S			4.7200	0.0384	No	29	0.82	0.3739	Yes	Var	W/S	3.37%

## **5.4 User Subjective Overall Environment Ratings**

As described in Section 1.4, after completion of each pass of each Benchmark test in each environment users provided their subjective views of their experience by completing the 22-question Usability Survey (see Figure 3) rating the environment on a scale of 1 to 5 (very poor to very good). The questions were grouped into 4 areas (navigation, locating, movement, and general). What follows is a presentation of user overall impressions ratings of the interfaces for performing Benchmark 2 tasks (find/repair) at the completion of the 3<sup>rd</sup> pass as a representation of user final evaluations of each interface.

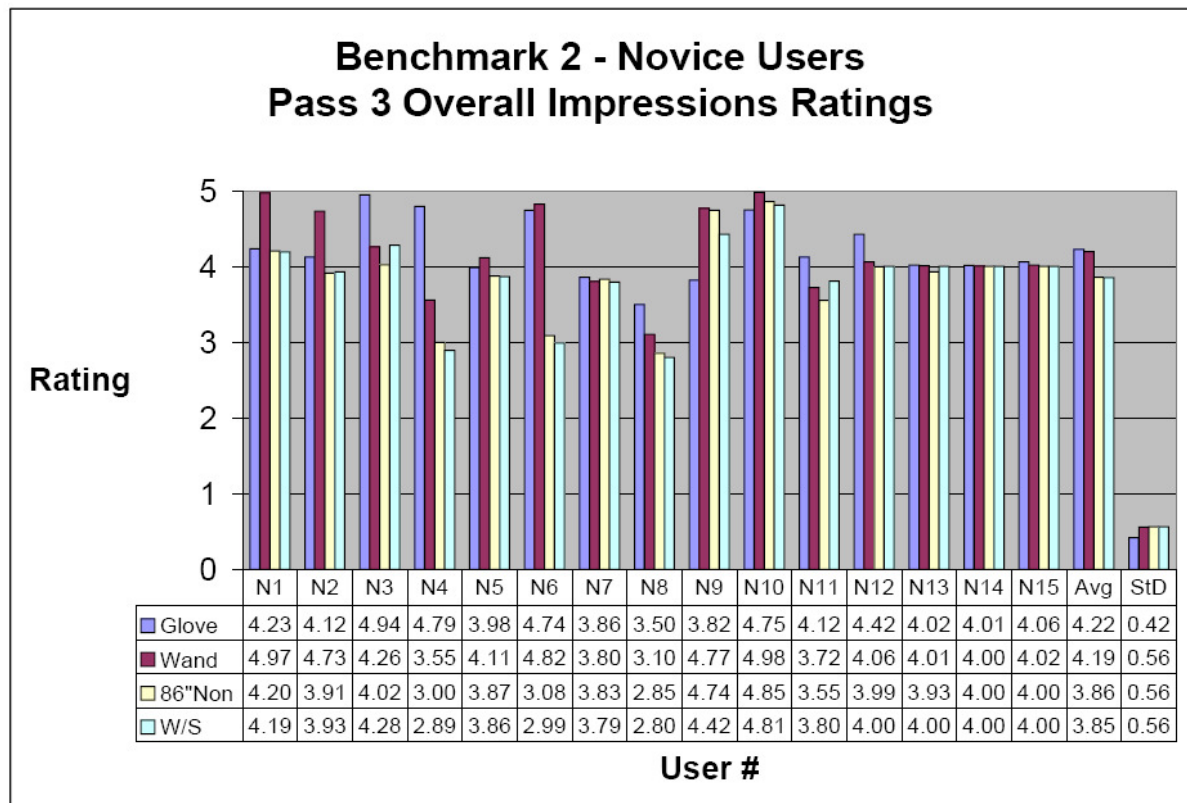
### **5.4.1 Novice Users**

As in Benchmark 1, each novice user was asked to rate his/her experience via the Usability Survey at the completion of each pass of each Benchmark test. Figure 16 presents the overall impressions ratings of the novice users at the completion of the 3<sup>rd</sup> pass of the Benchmark 2 scenario. As such, this represents each user's final impression of the navigational capabilities of each environment.

A further examination of the results detailed in Figure 16 shows that at completion of the Benchmark tests, novice users preferred the stereoscopic voice/glove interface over all other environments. Analysis of the results presented in the chart shows that novice users rated the stereoscopic interfaces slightly higher (8.3%) than the non-stereoscopic environments.

Table 23 shows the pass-to-pass improvements in novice user overall impression ratings for each of the environments. Note that with each successive exposure (pass-to-pass), the user's overall impressions of the interfaces improved. Examination of the pass-to-pass analysis of improvements noted in Table 23 shows that, on average, novice user impressions of the stereoscopic interfaces improved 4.75 times higher from pass 1 to pass 3 over non-stereoscopic interface environment improvements.





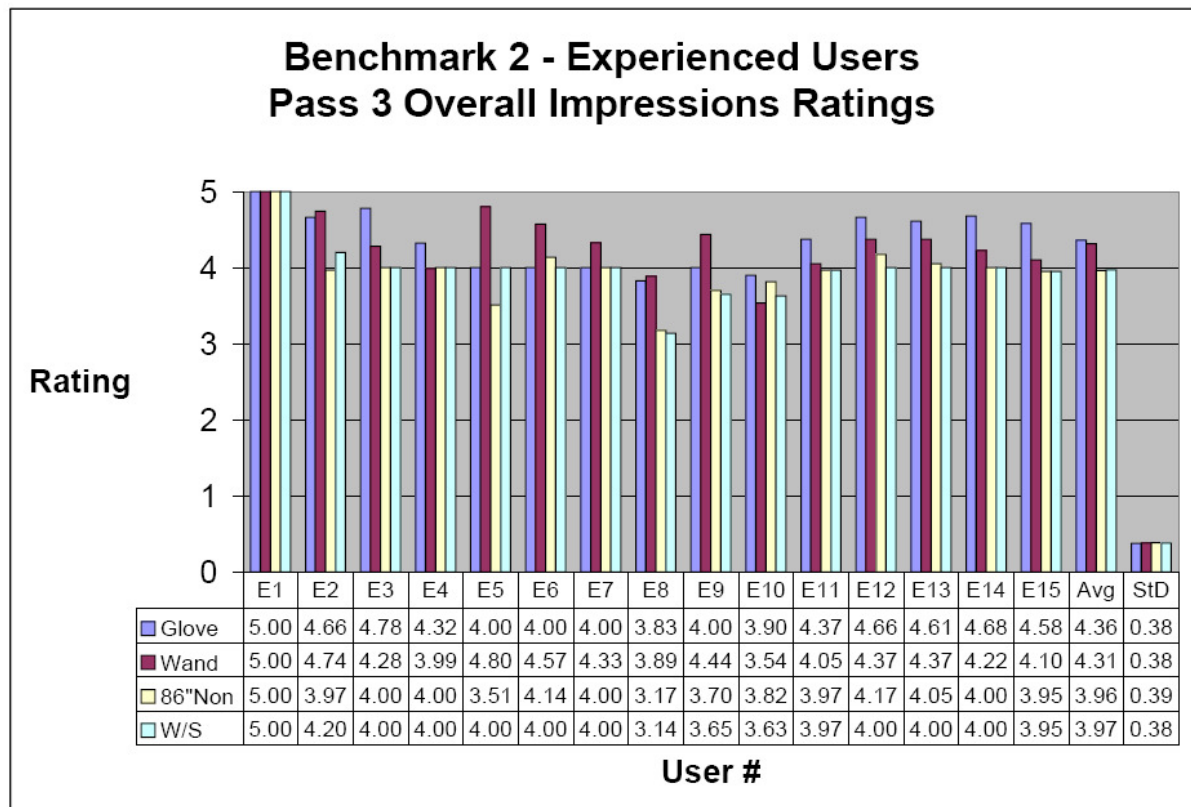
**Figure 16 – B2Np3Ovr – Novice User Pass 3 Overall Impression Ratings**

**Table 23 – B2N – Novice Users Pass-to-Pass Overall Impression Ratings Changes**

B2N Overall	Pass 1 to Pass 2		Pass 2 to Pass 3		Pass 1 to Pass 3		Average to Pass 3	
Δ Improve	Rating	%tage	Rating	%tage	Rating	%tage	Rating	%tage
Glove	0.05	1.14%	-0.01	-0.34%	0.03	0.79%	-0.01	-0.15%
Wand	0.10	2.34%	-0.03	-0.59%	0.07	1.73%	-0.02	-0.37%
86\"Non	0.01	0.33%	0.00	-0.07%	0.01	0.26%	0.00	-0.06%
W/S	0.00	0.12%	0.01	0.15%	0.01	0.27%	-0.01	-0.14%

#### 5.4.2 Experienced Users

As with the novice user group, each experienced user was asked to rate his/her experience through the Usability Survey at the completion of each pass of each Benchmark test. Figure 17 presents the overall impressions ratings of the experienced users at the completion of the 3<sup>rd</sup> pass of the Benchmark 2 scenario. As such, this survey represents each user's final impression of the navigational capabilities of each environment.



**Figure 17 – B2Ep3Ovr – Experienced User Pass 3 Overall Impression Ratings**

Analysis of the results noted in Figure 17 shows that the experienced users showed a 9.3% higher preference for the stereoscopic interfaces over the non-stereoscopic environments.

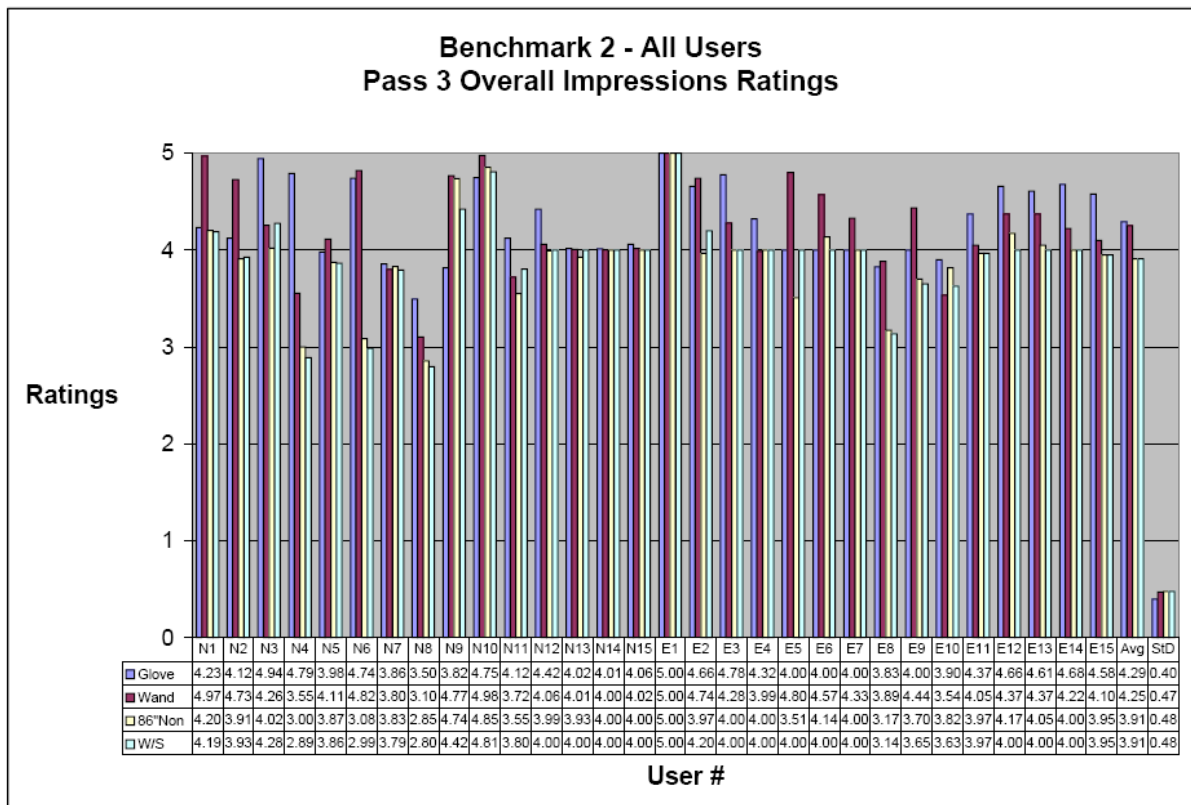
Table 24 shows the pass-to-pass improvements in experienced user overall impression ratings for each of the environments. Examination of the results presented show that for the experienced user group, on average, the pass-to-pass improvements in overall stereoscopic environments impressions was 9.4% higher from pass 1 to pass 3 than for non-stereoscopic environments.

**Table 24 – B2E – Experienced User Pass-to-Pass Overall Impression Ratings Changes**

B2E Overall	Pass 1 to Pass 2		Pass 2 to Pass 3		Pass 1 to Pass 3		Average to Pass 3	
Δ Improve	Rating	%tage	Rating	%tage	Rating	%tage	Rating	%tage
Glove	0.19	4.50%	0.38	9.50%	0.57	14.94%	-0.31	-7.22%
Wand	0.25	5.98%	0.32	7.96%	0.56	15.07%	-0.29	-6.82%
86\"Non	0.11	2.93%	0.06	1.52%	0.17	4.54%	-0.08	-1.95%
W/S	0.15	3.81%	0.10	2.58%	0.25	6.61%	-0.12	-2.90%

### 5.4.3 All Users

Figure 18 presents the overall impressions ratings of all users at the completion of the 3<sup>rd</sup> pass of the Benchmark 2 scenario. As such, this represents each user's final impression of the navigational capabilities of each environment. Analysis of the results presented in the chart shows that all users rated the voice/glove interface slightly higher than the stereoscopic wand environment (0.9%) and approximately 35% higher than either of the non-stereoscopic environments.



**Figure 18 – B2Allp3Ovr – All Users Pass 3 Overall Impression Ratings**

Table 25 shows the pass-to-pass improvements in the overall impression ratings for each of the environments for all the users. Examination of the pass-to-pass analysis of improvements noted in Table 25 shows that user impressions of the stereoscopic interfaces improved 7.8% (on average) from pass 1 to pass 3 and this improvement rate shows to be about 3.1 times the improvements shown using the non-stereoscopic interfaces.

**Table 25 – B2All – All Users Pass-to-Pass Overall Impression Ratings Changes**

B2A Overall	Pass 1 to Pass 2		Pass 2 to Pass 3		Pass 1 to Pass 3		Average to Pass 3	
$\Delta$ Improve	Rating	%tage	Rating	%tage	Rating	%tage	Rating	%tage
Glove	0.12	2.96%	0.18	4.42%	0.30	7.51%	-0.16	-3.74%
Wand	0.17	4.36%	0.15	3.57%	0.32	8.08%	-0.15	-3.64%
86"Non	0.06	1.64%	0.03	0.73%	0.09	2.39%	-0.04	-1.02%
W/S	0.08	1.99%	0.05	1.37%	0.13	3.39%	-0.06	-1.54%

## 5.5 User Environment Overall Ratings Statistics

The following sections present a statistical analysis of user subjective ratings of each of the four test environments following pass 3. These ratings represent the user's final overall impressions of the Benchmark 2 (find/repair) characteristics of the environments.

### 5.5.1 B2Np3Ovr – Novice User Overall Impressions Statistics

Table 26 presents the statistical analysis test results for the overall impressions ratings provided by the novice user group following completion of the 3<sup>rd</sup> pass of the Benchmark 2 (find/repair manipulation) tests. As described in Section 3.5 and Figure 6, the data sets for the environments are not all seen to be normally distributed, indicating that both parametric (Bartlett's) and non-parametric (Levene's) tests must be applied as described in the flowchart of Figure 6. These tests show all of the pair-wise environment comparisons to be of equal means. Thus, standard ANOVA tests are sufficient to determine the significance of the hypothesis of equal means for the environments.

**Table 26 – B2Np3Ovr – Novice Users Overall Impressions Statistics**

Benchmark 2, Novice Users, Pass 3, Overall Impressions Ratings												
B2Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.22	0.42	3.50	4.94	> 0.1000	Yes	9.86%				
Wand	15	4.19	0.56	3.10	4.98	>0.1000	Yes	13.30%				
86"Non	15	3.86	0.56	2.85	4.85	0.0584	No	14.58%				
W/S	15	3.85	0.56	2.80	4.81	0.0087	No	14.58%				
B2Np3	Homogeneity of Variance					Roy's Greatest Root: F(3,12)			2.53	Pr > F		0.1069
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand		1.1334	0.2870			Yes	14	0.05	0.8341	Yes	Neither	Neither
Glove vs 86"Non				1.2200	0.2780	Yes	14	4.33	0.0563	No	Means	Glove 9.57%
Glove vs W/S				1.3300	0.2593	Yes	14	4.71	0.0476	No	Means	Glove 9.69%
Wand vs 86"Non				1.9500	0.1736	Yes	14	7.74	0.0147	No	Means	Wand 8.77%
Wand vs W/S				0.0000	0.9767	Yes	14	6.88	0.0201	No	Means	Wand 8.88%
86"Non vs W/S				0.0000	0.9962	Yes	14	0.01	0.9117	Yes	Neither	Neither

Examination of these results as presented in Table 26 shows that for there is sufficient statistical significance in the results to suggest that the novice user voice/glove

stereoscopic interface overall ratings averages 9.63% higher than the non-stereoscopic interfaces and that the stereoscopic wand interface overall ratings average 8.83% higher than non-stereoscopic interfaces. However, it should be noted that there is insufficient statistical evidence to infer any other novice user overall environment preferences.

### 5.5.2 B2Ep3Ovr – Experienced User Overall Impressions Statistics

Table 27 presents the results of the statistical test performed on the results of the experienced user pass 3 Benchmark 2 tests that are similar to those of the novice users. Table 27 includes the results of the Kolmogorov-Smirnov test (KS statistic) indicating that the pass 3 results are not all normally distributed, and therefore, the comparisons for the environments must be performed using one-way, repeated measures, nonparametric ANOVA tests (Friedman-Dunn). Examination of these comparisons shows that there is sufficient statistical evidence in the by the experienced user Benchmark 2 (find/repair manipulation) comparisons results to suggest that the experienced user voice/glove stereoscopic interface overall ratings averages 9.91% higher than the non-stereoscopic interfaces and that the stereoscopic wand interface overall ratings average 8.75% higher than non-stereoscopic interfaces. However, it should also be noted that there is insufficient statistical evidence to infer any other novice user overall environment preferences.

**Table 27 – B2Ep3Ovr – Experienced Users Overall Impressions Statistics**

Benchmark 2, Experienced Users, Pass 3, Overall Impressions Ratings											
B2Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.36	0.38	3.83	5.00	0.0324	No	14.43%			
Wand	15	4.31	0.38	3.54	5.00	> 0.1000	Yes	21.58%			
86"Non	15	3.96	0.39	3.17	5.00	0.0324	No	24.75%			
W/S	15	3.97	0.38	3.14	5.00	0.0087	No	18.83%			
B2Ep3	Homogeneity of Variance					Roy's Greatest Root: F(3,12)			14.21	Pr > F	0.0003
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.0000	0.9938	Yes	29	0.19	0.6703	Yes	Neither	Neither
Glove vs 86"Non			0.0100	0.9404	Yes	29	27.47	0.0001	No	Means	Glove 9.99%
Glove vs W/S			0.0000	0.9774	Yes	29	28.09	0.0001	No	Means	Glove 9.83%
Wand vs 86"Non			0.0100	0.9412	Yes	29	11.85	0.0040	No	Means	Wand 8.83%
Wand vs W/S			0.0000	0.9757	Yes	29	20.26	0.0005	No	Means	Wand 8.67%
86"Non vs W/S			0.0000	0.9727	Yes	29	0.02	0.8913	Yes	Neither	Neither

### 5.5.3 B2Allp3Ovr – All Users Overall Impressions Statistics

Table 27 presents the statistical analysis test results of the overall impressions ratings of the entire user group at the completion of the 3<sup>rd</sup> pass of the Benchmark 2 scenario. As expected the aggregate user ratings pass the KS statistics indicate that not all of the environment data sets are normally distributed. Thus, as discussed in Section 3.5 and the flowchart presented in Figure 6, Levene's is required to determine the homogeneity of variance for the pair-wise comparisons of the environments.

As reported in Table 27, the Friedman-Dunn tests indicate that there is sufficient statistical significance in the results to suggest that the combined user group voice/glove stereoscopic interface averages 9.78% higher overall ratings than either of the non-stereoscopic interfaces. Additionally, it should be noted that there is sufficient statistical evidence to infer any that the stereoscopic wand interface is rated higher than both non-stereoscopic interfaces by an average of 8.78%.

**Table 28 – B2Allp3Ovr – All Users Overall Impressions Statistics**

Benchmark 2, All Users, Pass 3, Overall Impressions Ratings											
B2Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.29	0.40	3.50	5.00	0.0313	No	9.24%			
Wand	30	4.25	0.47	3.10	5.00	> 0.1000	Yes	11.09%			
86"Non	30	3.91	0.48	2.85	5.00	0.006	No	12.20%			
W/S	30	3.91	0.48	2.80	5.00	0.0005	No	12.15%			
B2Ap3	Homogeneity of Variance					Roy's Greatest Root:		F(3,27)	8.29	Pr > F	0.0005
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			1.1900	0.2803	Yes	29	0.19	0.6642	Yes	Neither	Neither
Glove vs 86"Non			0.8700	0.3543	Yes	29	16.29	0.0004	No	Means	Glove 9.80%
Glove vs W/S			0.7600	0.3873	Yes	29	17.24	0.0003	No	Means	Glove 9.77%
Wand vs 86"Non			0.0000	0.9523	Yes	29	19.51	0.0001	No	Means	Wand 8.80%
Wand vs W/S			0.0000	0.9694	Yes	29	21.29	0.0001	No	Means	Wand 8.77%
86"Non vs W/S			0.0000	0.9860	Yes	29	0.00	0.9710	Yes	Neither	Neither

## 5.6 Benchmark 2 Reliability

Table 29 presents the Cronbach's  $\alpha$  values computed for the objective measures (elapsed navigation times) and overall subjective user ratings for the Benchmark 2 scenario. Note that the alphas for the standardized variables (removal of duplicate values) are above 85%, indicating a high reliability for the benchmark. (Appendix M presents the detailed pass/environment results of the reliability calculations.)

**Table 29 – Benchmark 2 Cronbach’s  $\alpha$  Reliability Coefficients**

Benchmark 2 - All Users Cronbach Coefficient Alpha			
Variables	Alpha	Variables	Alpha
Standardized	86.12%	Standardized	97.57%

## **6 Benchmark 3 (Spatial Awareness)**

### ***6.1 Description***

In order to evaluate the ability of each environment/interface to aid users in their awareness of a design space, a unique space, totally unknown to the users, was created. For the test, the space created was a melding (at right angles) of a typical berthing space and a shipboard machine shop. Into this space the test administrators were able to inject an obelisk icon (an elongated, grey-white, pyramid topped by a sphere) that is not normally found in any shipboard space. Two such icons were randomly placed into the new space for each pass of the test. From a common starting point, users were required to navigate through the space looking for the icons within the space. The time required each user to locate each icon was recorded and the users were asked to note the location for each (placement within the space).

Upon completion of the test each user was shown a 2-dimensional, 8.5” x 11,” plan-view of the space and asked to note the placement of each of the two icons. The test administrators then recorded the offset (in mm) between user placement and the actual location of the icons.

This exercise (Benchmark 3) was repeated in each of the 4 environments under test and the User Survey was administered to each user after each pass in each environment. As with the other Benchmark testing, sequencing of the testers through the four environments was randomized so that not all of the users were testing the same interface in the same order. This randomization was used to eliminate bias in the testing.

### ***6.2 Pass-to-Pass Placement Improvements in Locating Icons***

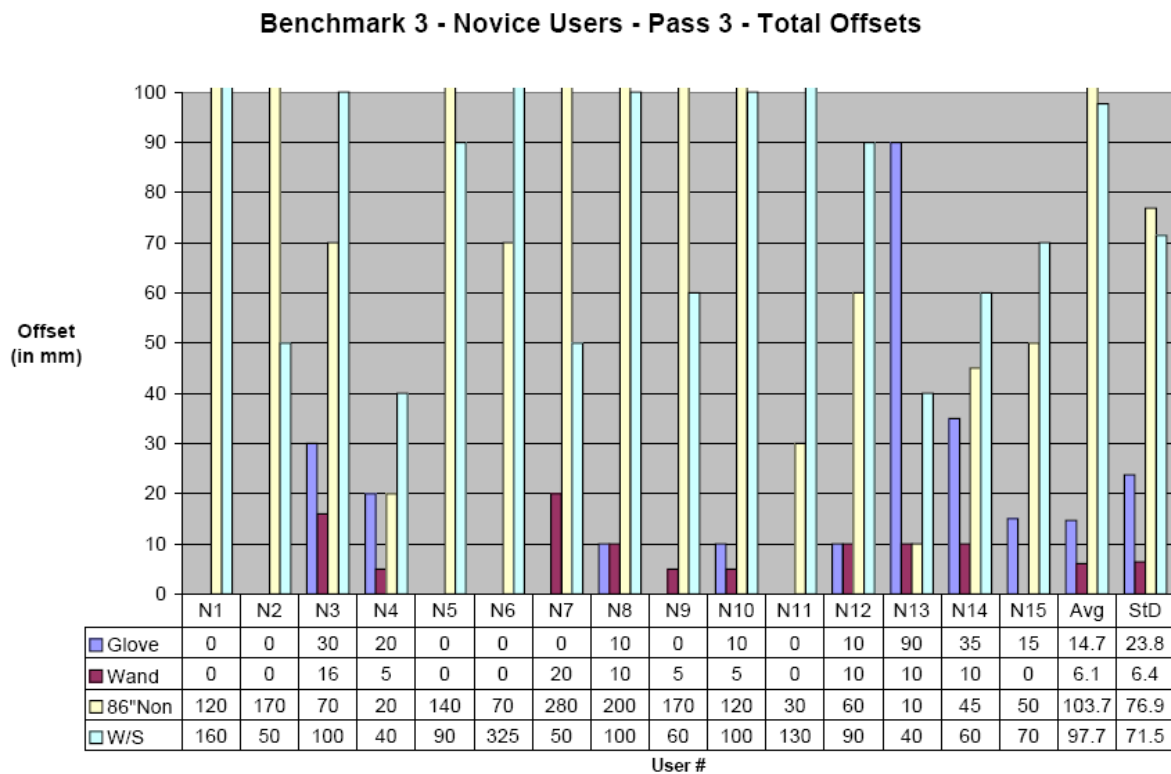
What follows is a presentation of the Benchmark 3, pass 3, part 1 and part 2 placement offsets for novice and experienced users. Since the spatial awareness test involves locating a first icon (part 1) and a second icon (part 2) the total of these two offsets is used as the basis for this evaluation. Pass 3 results are presented here as representative of user best-final case results.



### 6.2.1 Novice Users

Figure 19 presents novice user average offset distance for the placement of the two Benchmark 3 icons within the new space. The results clearly indicate a higher spatial awareness using the stereoscopic environment. Using the stereoscopic interfaces novice users, on average, located the each of the two icons within 5.2 mm of its actual location.

Inspection of the standard deviation values for the average offsets for the two icons show a very high variance in offset for the non-stereoscopic interfaces and shows low variance for the stereoscopic interfaces. This is an indication of the consistency of the stereoscopic method in spatial recognition efforts. As can be determined from the data presented in Figure 19, novice users demonstrated a markedly better spatial awareness using the stereoscopic interfaces (more 868% better).

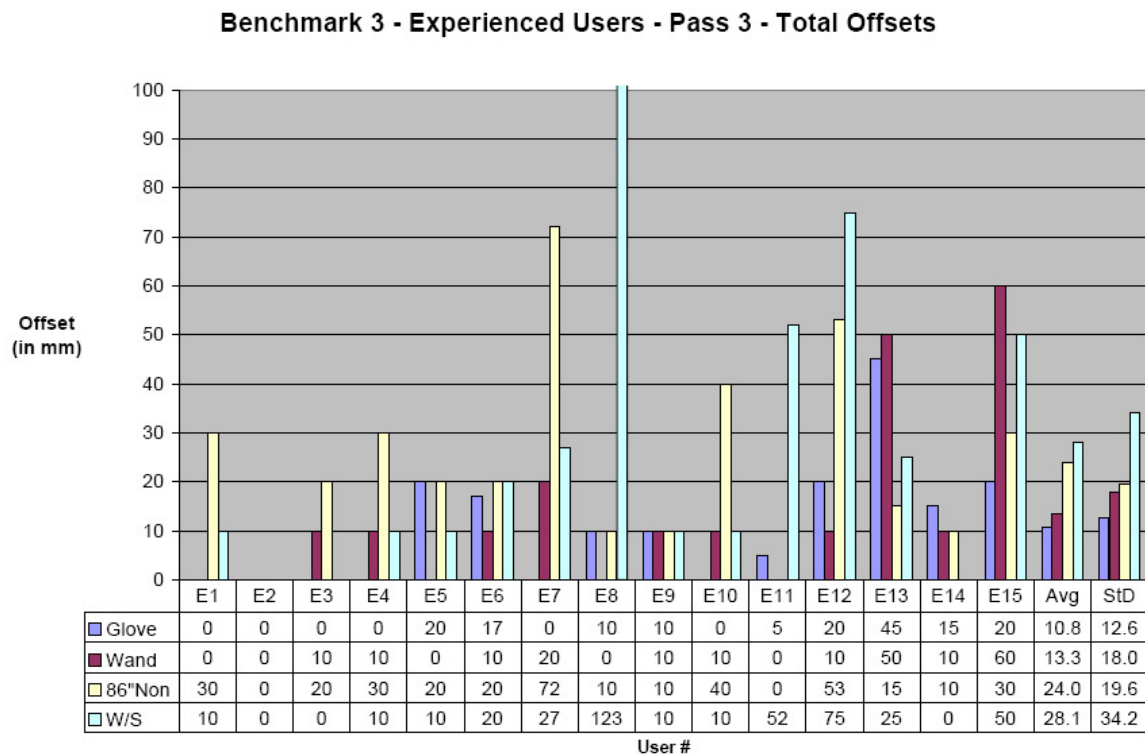


**Figure 19 – B3Np3-Toff – Pass 3 Novice User Total Placement Offset**

### 6.2.2 Experienced Users

Figure 20 presents experienced user average offset distance for the placement of the two Benchmark 3 icons within the new space. The results clearly indicate a higher spatial

awareness using the stereoscopic environments. The experienced users, on average, located each icon within 6.25 mm of its actual location. By comparison, experienced user placement in the non-stereoscopic environments averaged 13.02 mm from the actual location. Experienced user stereoscopic interface identification of actual placement of the icon was 2.08 times better than with non-stereoscopic methods.

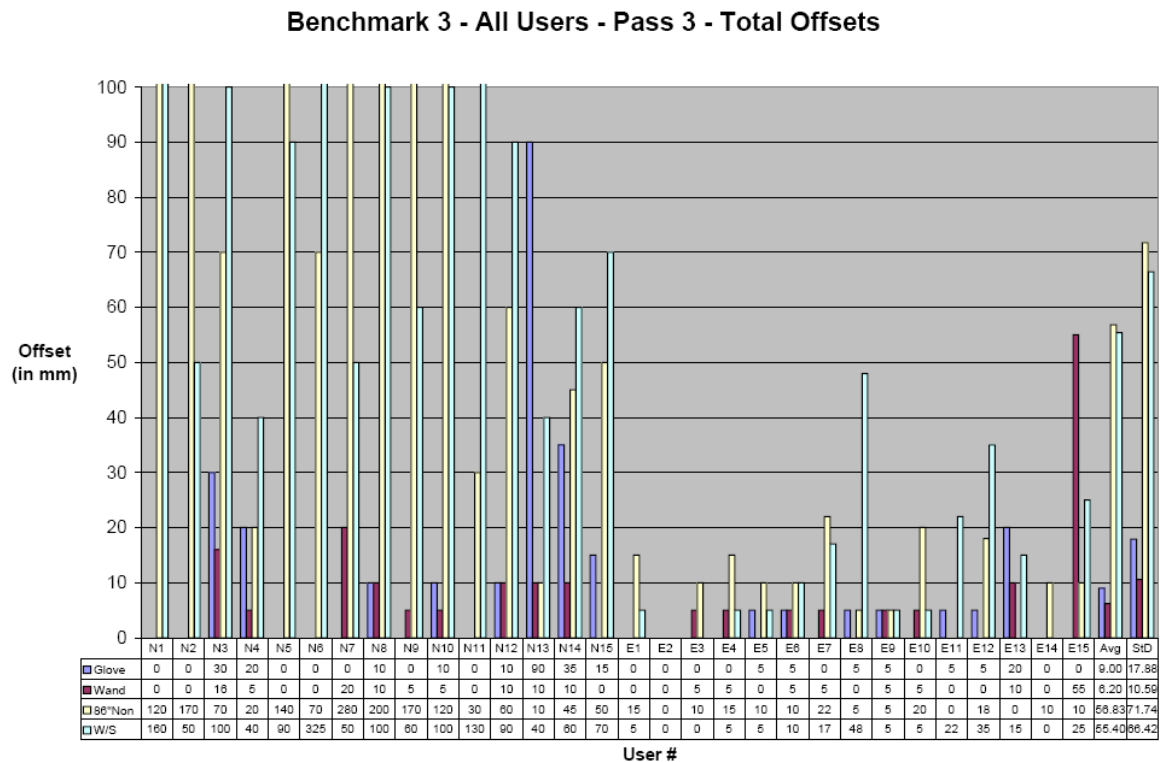


**Figure 20 – B3Ep3-Toff – Pass 3 Experienced User Total Icon Placement Offsets**

### 6.2.3 All Users

Figure 21 presents the aggregated results of all users placement of the 2 icons of Benchmark 3 within the new space. As would be expected, the results show that all users demonstrate consistently better spatial awareness with the use of stereoscopic over non-stereoscopic interfaces (average 111% better).

Inspection of the data also shows that the stereoscopic wand interface provides better spatial awareness for the user groups resulting in offsets 45% closer to actual placement than in the stereoscopic voice/glove environment. The variance or standard deviation is also significantly reduced for the stereoscopic environments.



**Figure 21 – B3Allp3-Toff – Pass 3 All Users Total Icon Placement Offsets**

### 6.3 Spatial Awareness Detailed Statistical Analysis

The following sections present a detailed statistical analysis of the Benchmark 3 results of the user groups in a manner similar to the previous Benchmarks. As discussed in Section 3.5, the GraphPad Prism [98] and SAS for Windows [99] software packages were used to perform each analysis. Each set of user icon 2 placement offsets is first examined to determine if the data is normally distributed (Gaussian distribution) using the KS statistic. The descriptive statistics test results are presented in tabular form followed by the results of an ANOVA pair-wise comparison of the offsets for each environment using standard parametric tests for normally distributed data or Friedman-Dunn nonparametric post-tests for sets of offsets data that are not normally distributed. As described in the Benchmark 1 and 2 discussions, the analyses were performed in accordance with the flowchart presented in Figure 6.

### 6.3.1 Benchmark 3 - Pass 3 Statistics

Benchmark 3, pass 3, total offsets represent each user's view of the placement of the required device in a foreign space. As such, the results of this pass/icon placement represent a reasonable characterization of the user's spatial awareness within each environment.

#### 6.3.1.1 B3Np3-Toff – Novice User Offsets Statistics

Table 30 presents the results of the statistical analysis of novice user pass location of icons in the test environment. Note that none of the environment data sets prove to be normally distributed. Thus, the non-parametric Levene's test is required to test the homogeneity of the pair-wise environment comparisons. The ANOVA analyses for determining equality of means combined with the Levene's test results provide the information required to determine the statistical significance of the comparisons results as described in Figure 6.

Examination of the results posted in Table 30 show that all but the stereoscopic voice/glove vs non-stereoscopic workstation interface parings produce statistically significant results providing sufficient statistical evidence to suggest that the stereoscopic environments produce smaller offset distances (averaging 91.26% smaller). Also, since the stereoscopic wand environment shows offset distances that are 58.64% smaller than the stereoscopic voice/glove interface, it can be said, with statistical significance, that the stereoscopic wand environment produces more reliable spatial awareness for novice users than does either non-stereoscopic environment.

**Table 30 – B3Np3-Toff – Novice User Pass 3 – Total Offsets Statistics**

Benchmark 3, Novice Users, Pass 3, Total Offsets											
B3Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	14.67	23.79	0.00	90.00	0.0046	No	22.53%			
Wand	15	6.07	6.43	0.00	20.00	0.0359	No	24.73%			
86"Non	15	103.70	76.87	10.00	280.00	0.0985	Yes	29.82%			
W/S	15	97.67	71.49	40.00	325.00	0.0017	No	27.39%			
B3Np3	Homogeneity of Variance				Roy's Greatest Root:				F(3,12) 34.02 Pr > F 0.0026		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			2.5900	0.1185	Yes	14	10.52	0.0179	No	Means	Wand 58.64%
Glove vs 86"Non			5.8900	0.0219	No	14	19.70	0.0044	No	Var/Mean	Glove 85.85%
Glove vs W/S			1.4900	0.2329	Yes	14	2.53	0.1631	Yes	Neither	Neither
Wand vs 86"Non			7.6900	0.0098	No	14	0.41	0.5463	Yes	Var	Wand 94.15%
Wand vs W/S			1.9700	0.1712	Yes	14	5.31	0.0607	No	Means	Wand 93.79%
86"Non vs W/S			0.0400	0.8500	Yes	14	7.71	0.0322	No	Means	W/S 5.81%

### 6.3.1.2 B3Ep3-2off – Experienced User Offsets Statistics

Table 31 presents the results of ANOVA and post-tests calculations of experienced user spatial awareness testing for pass 3 total icon placements. Interpretation of the test results for the experienced user group follow the methods described above for novice users and shows very similar results.

None of the experienced users distance offsets data sets shown in Table 31 prove to be normally distributed. Thus, as with the novice user calculations, ANOVA and nonparametric (Friedman-Dunn) post-tests were performed on the data sets. However, examination of Table 31 shows that the results for all of the environment comparisons can be shown to be statistically significant with the exception of the stereoscopic wand vs non-stereoscopic environment comparisons. Therefore, the results posted in Table 31 show that with statistical significance the stereoscopic voice/glove interface produces “better” spatial awareness than the non-stereoscopic environments (placements that are more than 6 times closer to the actual placement).

**Table 31 – B3Ep3-Toff – Experienced Users Pass 3 – Total Offsets Statistics**

Benchmark 3, Experienced Users, Pass 3, Total Offsets												
B3Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	9.00	17.88	0.00	90.00	< 0.0001	No	198.67%				
Wand	15	6.20	10.59	0.00	55.00	< 0.0001	No	170.81%				
86"Non	15	56.83	71.74	0.00	280.00	< 0.0001	No	126.24%				
W/S	15	55.40	66.42	0.00	325.00	0.0030	No	119.89%				
B3Ep3	Homogeneity of Variance					Roy's Greatest Root:			F(3,12)	11	Pr > F	0.001
	Bartlett's Test		Levene's Test		Equal	df	F(1,14)	Pr > F	Equal Means	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F	Var ?							
Glove vs Wand			1.0400	0.3175	Yes	14	6.41	0.0239	No	Means	Wand	31.11%
Glove vs 86"Non			1.1700	0.2892	Yes	14	14.59	0.0019	No	Means	Glove	531.44%
Glove vs W/S			1.4200	0.2427	Yes	14	13.67	0.0022	No	Means	Glove	515.56%
Wand vs 86"Non			0.1100	0.7432	Yes	14	0.00	0.9759	Yes	Neither		Neither
Wand vs W/S			0.9200	0.3445	Yes	14	1.84	0.1367	Yes	Neither		Neither
86"Non vs W/S			0.6700	0.4189	Yes	14	3.22	0.0942	No	Means	W/S	2.58%

### 6.3.1.3 B3Allp3-2off – All Users Offsets Statistics

Table 32 reports the statistical analysis of all users placement of the 3<sup>rd</sup> pass, total icon placement offsets data in Benchmark 3. Examination of the KS-statistic shows that none of the environments produced data set results that were normally distributed. Thus, nonparametric (Friedman-Dunn) pair-wise environment ANOVA post-tests calculations were performed with the results presented in Table 32.

Examination of the results posted in Table 32 show that there is evidence to infer only that the stereoscopic voice/glove environment is “better” than either non-stereoscopic

environment (by an average of over 776%), and there is insufficient evidence to a claim that the stereoscopic voice/glove interface produces distance offsets that are 4.27 % smaller than the stereoscopic wand environment.

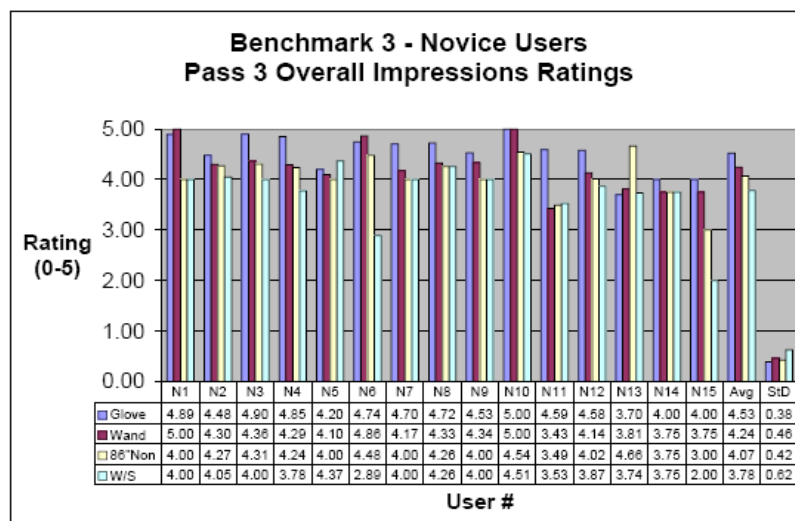
**Table 32 – B3Allp3-Toff – All Users Pass 3 – Total Offsets Statistics**

Benchmark 3, All Users, Pass 3, Total Offsets												
B3Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	6.73	17.88	0.00	90.00	< 0.0001	No	265.56%				
Wand	30	7.03	10.90	0.00	55.00	< 0.0001	No	154.98%				
86"Non	30	59.17	71.02	0.00	280.00	< 0.0001	No	120.03%				
W/S	30	58.73	66.99	0.00	325.00	0.0071	No	114.06%				
B3Ap3	Homogeneity of Variance					Roy's Greatest Root: F(3,27) 10.9 Pr > F 0.001						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			1.0400	0.3175	Yes	29	6.41	0.0239	No	Means	Glove	4.27%
Glove vs 86"Non			1.1700	0.2892	Yes	29	14.59	0.0019	No	Means	Glove	778.81%
Glove vs W/S			1.4200	0.2427	Yes	29	13.67	0.0022	No	Means	Glove	772.27%
Wand vs 86"Non			0.1100	0.7432	Yes	29	0.00	0.9759	Yes	Neither	Neither	
Wand vs W/S			0.9200	0.3445	Yes	29	1.84	0.1367	Yes	Neither	Neither	
86"Non vs W/S			0.6700	0.4189	Yes	29	3.22	0.0942	No	Means	W/S	0.75%

## 6.4 User Subjective Overall Environment Ratings

### 6.4.1 Novice Users

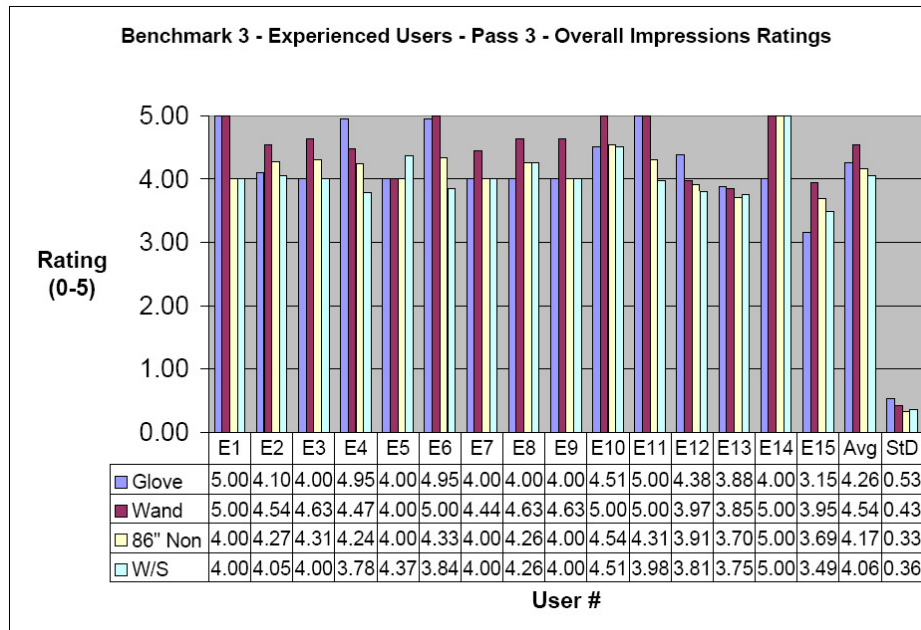
Figure 22 graphically presents a comparison of the Benchmark 3 (spatial awareness) pass 3 overall ratings of the four environments. Inspection of the average ratings shows that novice users preferred the stereoscopic environments over the non-stereoscopic environments by 11.6% and showed, for stereoscopic interfaces, they preferred the voice/glove interface over the wand interface by 6.8%.



**Figure 22 – B3Np3Ovr – Novice User Pass 3 Overall Impressions Ratings**

## 6.4.2 Experienced Users

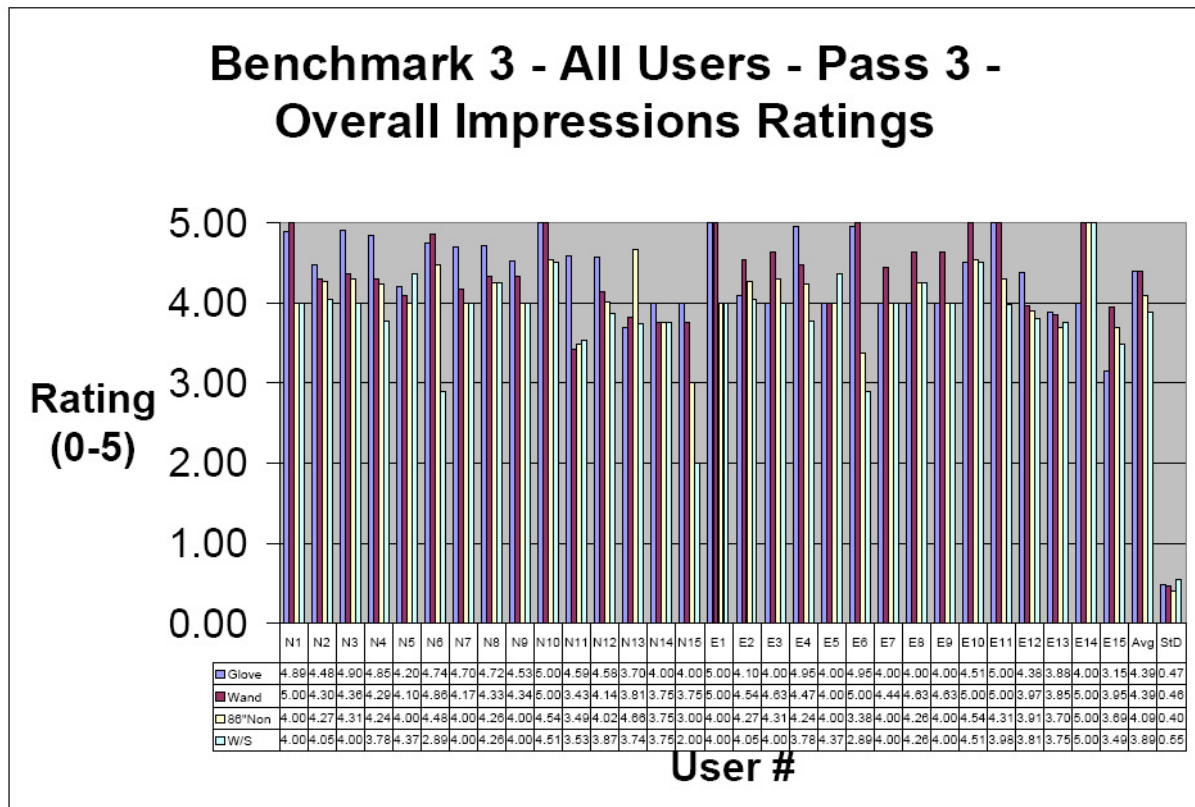
Figure 23 shows that experienced users have slightly different preferences. While the experienced users still preferred the stereoscopic environments by over 6.9% over the non-stereoscopic environments, they preferred the stereoscopic wand interface by 6.6% over the voice/glove interface.



**Figure 23 – B3Ep3Ovr – Experienced User Pass 3 Overall Impressions Ratings**

## 6.4.3 All Users

Figure 24 provides a comparison of all users environment preferences. It shows that, as a combined group, all users reported a preference for the stereoscopic environments by slightly better than 22.8% over the non-stereoscopic interfaces and that there was essentially no preference between either of the stereoscopic interfaces.



**Figure 24 – B3AllP3Ovr – All Users Pass3 Overall Impressions Ratings+**

## 6.5 User Environment Overall Ratings Statistics

The following sections present a detailed statistical analysis of user overall impressions ratings of the four test environments following their 3<sup>rd</sup> and final pass of the Benchmark 3 scenario. The statistical analysis of these ratings provides insight into the final opinions of the users. As discussed in Section 3.5, the [98] Graphpad Prism and SAS for Windows [99] software packages were used to perform each analysis. Each set of user overall impressions ratings is first examined to determine if the data is normally distributed (Gaussian distribution) using the KS statistic. The descriptive statistics test results are presented in tabular form followed by the results of an ANOVA pair-wise comparison of the overall impressions ratings for each environment using standard, multiple-comparison ANOVA parametric post-tests for normally distributed data or Friedman-Dunn nonparametric post-tests for sets of ratings data that are not normally distributed. As discussed in Section 6.3, all analyses were performed in accordance with the logic presented in the flowchart provided in Figure 6.



## 6.5.1 Pass 3 Statistics

As noted, Benchmark 3, pass 3, overall impressions ratings represent each user's view of the placement of the required device in a foreign space. As such, these ratings represent a reasonable characterization of the user's overall impressions of the interfaces after the use of each to determine his/her spatial awareness of a previously unknown environment.

### 6.5.1.1 B3Np3Ovr – Novice User Overall Impressions Statistics

Table 33 presents the results of the statistical analysis of novice user pass 3 overall impressions of the interface. Note that while the stereoscopic environments produce data sets that are normally distributed and the non-stereoscopic environments produce data sets that are nongaussian. Thus, Levene's test is required to determine the homogeneity of the variances of the pair-wise comparisons of environments for all but the comparison of the two stereoscopic interfaces (where Bartlett's test is applied).

**Table 33 – B3Np3Ovr – Novice User Pass 3 Overall Impressions Statistics**

Benchmark 3, Novice Users, Pass 3, Overall Impressions Ratings											
B3Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.53	0.3846	3.70	5.00	> 0.1000	Yes	22.53%			
Wand	15	4.24	0.4582	3.43	5.00	> 0.1000	Yes	24.73%			
86"Non	15	4.09	0.4231	3.00	4.66	0.0237	No	29.82%			
W/S	15	3.78	0.6211	2.00	4.51	0.0043	No	27.39%			
B3Np3	Homogeneity of Variance				Roy's Greatest Root: F(3,12)				8.43 Pr > F 0.0028		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.4128	0.5206	0.1100	0.7446	Yes	14	10.39	0.0061	No	Means	Glove 6.27%
Glove vs 86"Non			1.0700	0.3087	Yes	14	13.57	0.0025	No	Means	Glove 9.72%
Glove vs W/S			0.0800	0.7738	Yes	14	23.03	0.0003	No	Means	Glove 16.44%
Wand vs 86"Non			0.5700	0.4573	Yes	14	2.70	0.1223	Yes	Neither	Neither
Wand vs W/S			0.7500	0.3932	Yes	14	7.67	0.0150	No	Means	Wand 10.84%
86"Non vs W/S					Yes	14	4.65	0.0488	No	Means	86"Non 7.44%

As can be seen in Table 33, the results of ANOVA nonparametric post-tests pair-wise comparison of the four environments show that at the 90% confidence level, there is sufficient statistical significance to indicate that the stereoscopic voice/glove environment is preferred by novice users over the stereoscopic wand environment by 6.27% and over the non-stereoscopic interfaces by an average of 13.08%. It should also be noted that the novice user group rated the stereoscopic wand interface 10.84% higher than the traditional non-stereoscopic workstation environment.

### 6.5.1.2 B3Ep3Ovr – Experienced User Overall Impressions Statistics

Table 34 presents the results of ANOVA and post-tests calculations of experienced user overall impressions ratings of the four environments. Interpretation of the test results follow the methods described in the section above for novice user tests which are quite similar.

Each environment data was found to be normally distributed. Thus, as shown in the flowchart of Figure 6, Bartlett's test is applied to determine the equality of means of the pair-wise environment comparisons as well as parametric post-ANOVA tests to determine the equality of environment means. Further, examination of Table 34 shows that the results of the experienced user post-tests produced results of sufficient statistical significance to infer meaningful comparisons of only the stereoscopic wand environment with the stereoscopic voice/glove interface and with either of the non-stereoscopic environments. Specifically, the results support the conclusion that for the experienced user group the stereoscopic wand interface is preferred by 23.5% over the stereoscopic voice/glove interface and is rated higher than either non-stereoscopic interface by 13.15%.

**Table 34 – B3Ep3Ovr – Experienced Users Pass 3 Overall Impressions Statistics**

Benchmark 3, Experienced Users, Pass 3, Overall Impressions Ratings											
B3Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	212.3	30.64	175.0	270.0	> 0.1000	Yes	14.43%			
Wand	15	171.9	37.11	115.0	245.0	>0.1000	Yes	21.58%			
86"Non	15	194.5	48.14	95.0	260.0	>0.1000	Yes	24.75%			
W/S	15	196.4	36.80	125.0	250.0	> 0.1000	Yes	18.83%			
B3Ep3	Homogeneity of Variance					Roy's Greatest Root:		F(3,12)	4.4	Pr > F	0.0186
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.4924	0.4828			Yes	14	10.23	0.0064	No	Means	Wand 23.50%
Glove vs 86"Non	2.7107	0.0997			Yes	14	1.75	0.2069	Yes	Neither	Neither
Glove vs W/S	0.4509	0.5019			Yes	14	1.17	0.2969	Yes	Neither	Neither
Wand vs 86"Non	0.9305	0.3347			Yes	14	3.76	0.0729	No	Means	Wand 13.15%
Wand vs W/S	0.0009	0.9757			Yes	14	6.72	0.0213	No	Means	Wand 13.15%
86"Non vs W/S	0.9893	0.3199			Yes	14	6.72	0.0213	No	Means	86"Non 0.98%

### 6.5.1.3 B3Allp3Ovr – All Users Overall Impressions Statistics

Table 35 presents the Benchmark 3 environment overall impressions ratings descriptive statistics for all users, both novice and experienced. Examination of the KS-statistic shows that not all of the environments produced data set results that were normally distributed. Thus, nonparametric (Friedman-Dunn) pair-wise environment ANOVA post-tests calculations were performed with the results presented in Table 35.

Examination of the mean ratings listed in Table 35 shows that there is sufficient evidence to support the overall user preference of both stereoscopic interfaces as compared to either non-stereoscopic environment. However, there is sufficient evidence to infer that users rate the stereoscopic environments an average of 9.84% higher than the non-stereoscopic environments.

**Table 35 – B3Allp3Ovr – All Users Pass 3 Overall Impressions Statistics**

Benchmark 3, All Users, Pass 3, Overall Impressions Ratings												
B3Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	4.31	0.56	3.06	5.00	> 0.1000	Yes	12.96%				
Wand	30	4.11	0.52	3.36	5.00	< 0.0001	No	12.73%				
86"Non	30	3.86	0.32	3.00	4.74	0.0080	No	8.32%				
W/S	30	3.72	0.41	2.87	5.00	0.0778	No	10.95%				
B3Ap3						Roy's Greatest Root: F(3,27) 3.97 Pr > F 0.0183						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.2900	0.6296	Yes	29	0.07	0.8033	Yes	Neither	Neither
Glove vs 86"Non				9.1700	0.0037	No	29	6.13	0.0194	No	Var/Mean	Glove 10.37%
Glove vs W/S				3.4000	0.0760	No	29	9.46	0.0046	No	Var/Mean	Glove 13.64%
Wand vs 86"Non				6.6900	0.0122	No	29	9.59	0.0043	No	Var/Mean	Wand 5.96%
Wand vs W/S				1.6900	0.1993	Yes	29	11.60	0.0019	No	Mean	Wand 9.40%
86"Non vs W/S				0.7500	0.3902	Yes	29	7.79	0.0092	No	Mean	86"Non 3.65%

## **7 Collaboration**

Upon completion of testing baseline (Benchmark 1, 2, and 3) testing, Benchmarks 2 and 3 were repeated using 5 groups of 3 individuals simultaneously within the 4 test environments. The purpose of these tests was to investigate the collaborative capabilities of the preferred stereoscopic voice/glove environment. The testing was performed using 3 groups of experienced users (9 individuals) and 2 groups of novice users (6 individuals).

As with the individual testing, upon completion of each of the 3 passes, each user completed the environment survey rating his/her impressions of the collaborative characteristics of the environment.

### ***7.1 Enhanced Stereoscopic Voice/Glove Environment***

Initially the collaborative tests were to include a comparison of the previously reported stereoscopic voice/glove environment and a voice/glove environment enhanced to include a user requested “you-are-here” tracking map or interactive plan view inserted into a corner of the stereoscopic display. This enhancement would provide the user with a quick visual reference of his/her location on a “floor-plan” of the space.

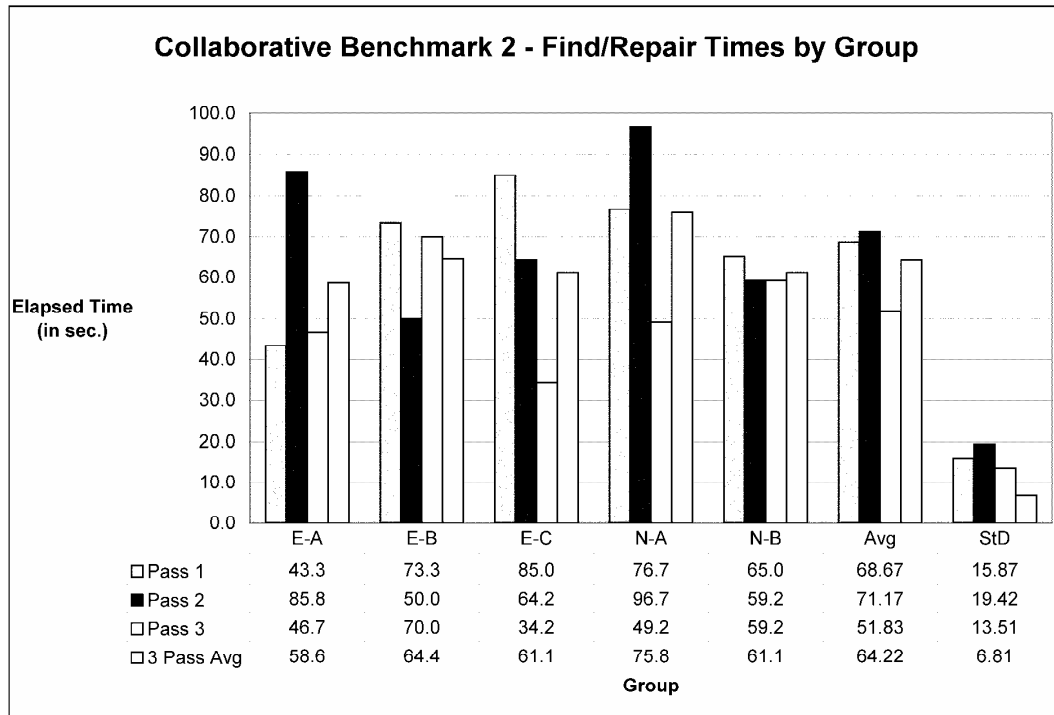
This enhancement was added to the voice/glove environment as a 3D display but quickly proved to render the interface too slow to be a viable solution. The enhancement was removed and collaborative testing was continued with the existing voice/glove environment. However, after discussions with users and the development team, alternative approaches have been developed and are discussed in Section 9.3 of this report.

### ***7.2 Collaborative Benchmark 2 Timings***

As with previous Benchmark 2 (find/repair) tests, users were timed while locating and repairing 2 preset errors within the test space. However, with the collaborative tests, groups of 3 users were each subjected to the exercise with each individual having the opportunity to lead (“drive”) the test.

Figure 25 provides a comparison of the improvements in performance by collaborative group for each pass of the find/repair scenarios. Examination of the chart

shows a 3.7% decrease in performance from pass 1 to pass 3, due to the particularly poor performance of one novice group, thereby influencing a 27.2% increase in performance from pass 2 to pass 3.



**Figure 25 – Collaborative Benchmark 2 Find/Repair Times by Group**

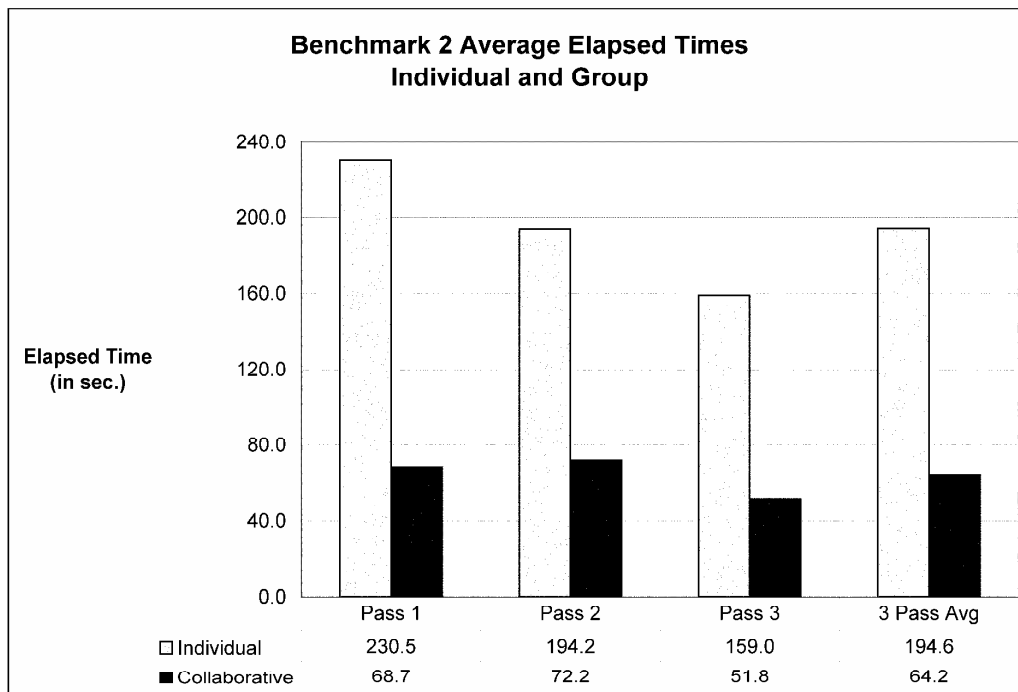
Table 37 provides the statistical comparison of the individual and collaborative Benchmark 2 results. It shows that the data sets are normally distributed and that the individual and collaborative data t-tests show extremely low p values indicating that the results are significant, and thus the data sets have unequal means (reject  $H_0$ ). This indicates that since the collaborative stereoscopic voice/glove environment produces significantly faster elapsed times than do individual users (~207% faster), the collaborative environment is “better” using this measure of performance.

Figure 26 provides a comparison of Benchmark 2 find/repair times for individual users and for collaborative groups. Note that there is a significant improvement in overall locating (find) and manipulation (repair) times with collaboration between the users within a group. Figure 26 shows pass 1, find/repair improvements of more than 235%; pass 2 improvements of almost 170%; pass 3 improvements of almost 207%; and 3 pass average

improvements of more than 203%. Such results indicate that multiple designers within a space find and repair defects approximately twice as fast as individuals.

**Table 37 – Benchmark 2 Timings Comparisons - Individual vs Collaborative**

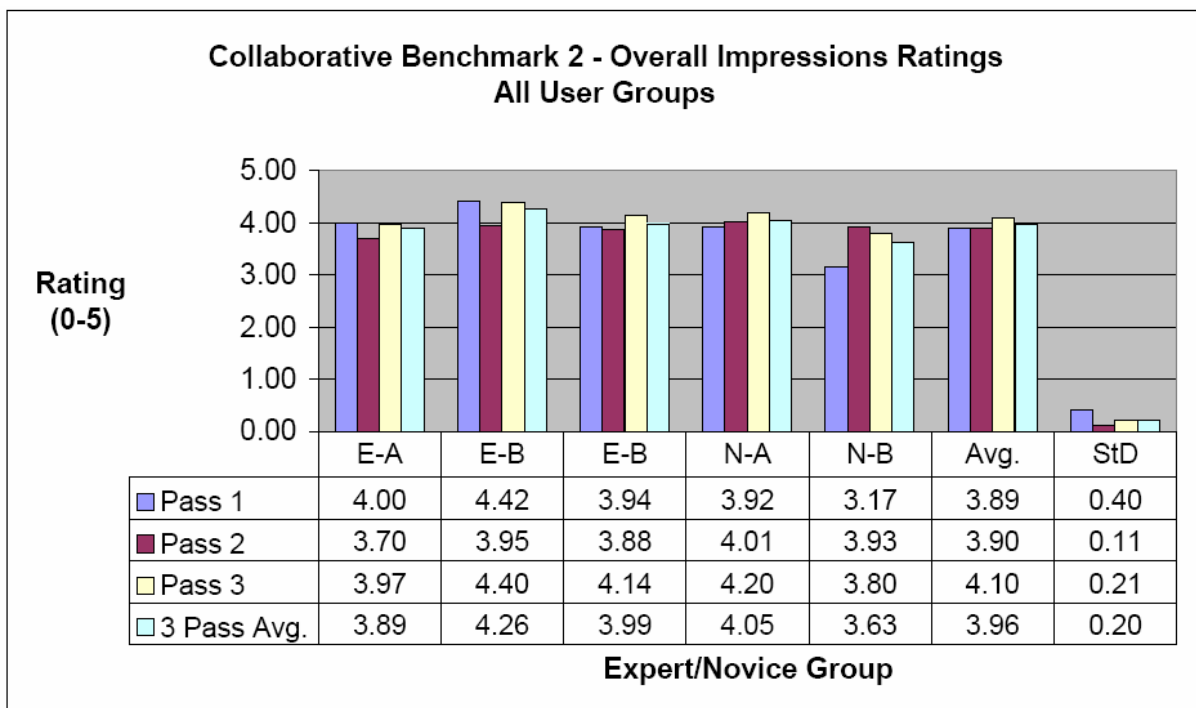
<b>Stereoscopic Glove Benchmark 2 Comparison Individual User vs Collaborative Group</b>		
	<b>Individual</b>	<b>Collaborative</b>
<b>Mean</b>	159.05	51.83
<b>Standard deviation (SD)</b>	34.92	18.21
<b>Sample Size</b>	21	5 teams of 3
<b>Std. error of means (SEM)</b>	7.62	4.70
<b>Lower 95% conf. limit</b>	143.15	41.75
<b>Upper 95% conf. limit</b>	174.94	61.92
<b>Minimum</b>	95.00	22.50
<b>Median (50th percentile)</b>	165.00	55.00
<b>Maximum</b>	210.00	95.00
<b>Normality test KS</b>	0.1067	0.1282
<b>Normality test P value</b>	>0.10	>0.10
<b>Passed normality test?</b>	Yes	Yes
<b>Unpaired t test with Welch Correction</b>	P<0.0001	P<0.0001
<b>Significant?</b>	Extremely	Extremely



**Figure 26 – Comparison of Individual vs Collaborative Find/Repair Times**

### 7.3 Collaborative Benchmark 2 Ratings

As with the individual Benchmark tests, the user groups were asked, after each pass of the test, to rate the collaborative environment via the survey questionnaire. Figure 27 presents a graphical view of each group's overall impressions of the collaborative operations of the stereoscopic glove/voice environment following each pass of the test. Note that the collaborative user groups show a 5.4% increase in approval rating from pass 1 to pass 3.

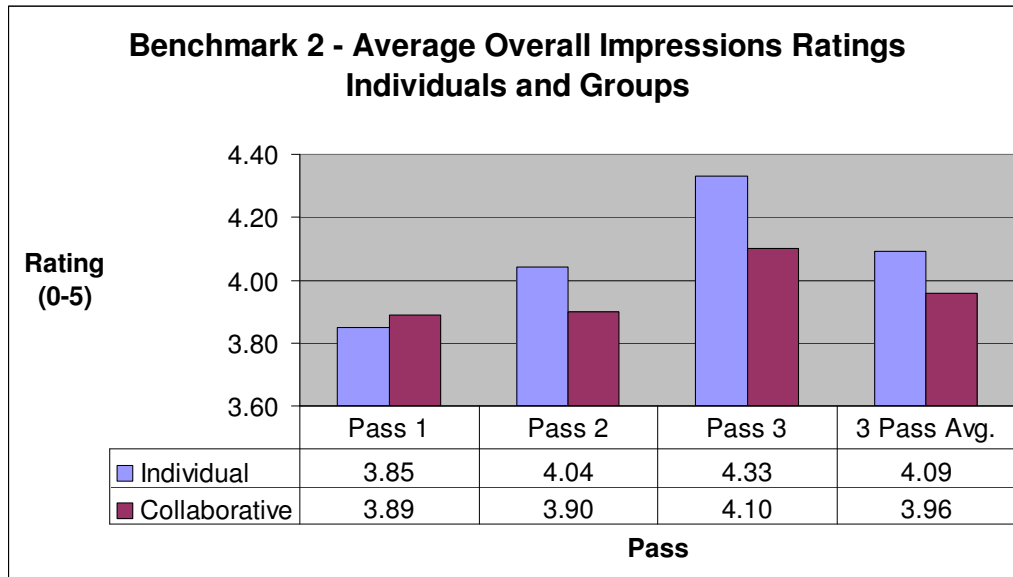


**Figure 27 – Collaborative Benchmark 2 Group Overall Impressions Ratings**

Figure 28 presents a graphic of the find/repair overall impressions ratings of individual users as compared with collaborative group ratings. It should be noted that the overall impressions of the collaborative environment rated somewhat lower (3.2%) than its capability by individual users. As discussed in the conclusions, lower evaluations of the collaborative environment have been attributed to the group discussion of the need for enhancements to the interface as tested.

However, a cursory examination of the group vs individual ratings statistics does seem to indicate that individually and in collaborative groups, users rate the stereoscopic

voice/glove environment somewhat better than 4.0, which is good (4.235/5.0). Further, the testers rated the environments within 3.2% between individual ratings and collaborative group ratings, as shown in Figure 28.



**Figure 28 – Comparison of Individual vs Collaborative Find/Repair Overall Ratings**

The collaborative find/repair rating is found to be normally distributed, but the data for individual user's rating is not normally distributed. To evaluate statistical significance an unpaired, two-tailed, Mann-Whitney test is used, as shown in Table 38. Due to a high P value, no statistically significant difference can be established.

**Table 38 – Collaborative Benchmark 2 Overall Impressions Statistics**

Stereoscopic Glove Benchmark 2 Comparison Overall Impressions Ratings Comparisons Individuals vs Collaborative Groups		
	Individual	Collaborative
Mean	4.33	4.23
Standard deviation (SD)	0.40	0.50
Sample Size	21	5 teams of 3
Std. error of means (SEM)	0.09	0.13
Lower 95% conf. limit	4.15	3.96
Upper 95% conf. limit	4.52	4.51
Minimum	3.80	3.59
Median (50th percentile)	4.30	4.01
Maximum	5.00	5.00
Normality test KS	0.2245	0.2054
Normality test P value	0.007	0.0885
Passed normality test?	No	Yes
Unpaired, two-tailed Mann-Whitney Test P value = 0.5206 is NOT significant.		

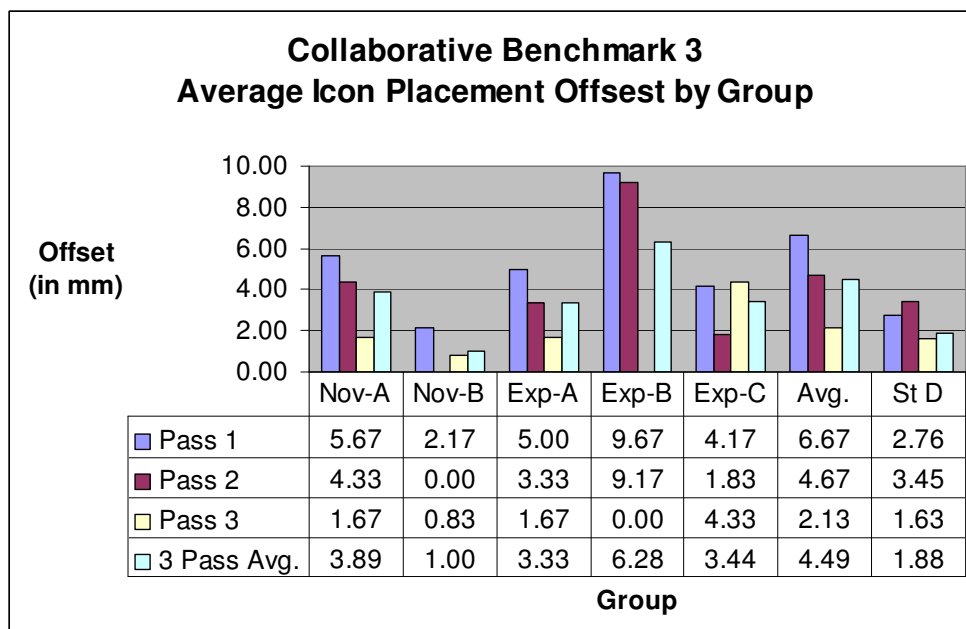


## 7.4 Collaborative Benchmark 3 Offsets

As with the individual Benchmark 3 (spatial awareness), users were timed while navigating through a previously unseen space searching for the placement of two icons. However, with the collaborative tests, groups of 3 users were each subjected to the exercise with each individual having the opportunity to lead (“drive”) the test. During the course of each test, all members of the group collaborated on the search for the icons.

Figure 29 provides a comparison of the offsets from the user perception of the actual placement of the icons. Examination of the chart shows that each user group exhibited a marked improvement in accurately assessing the placement of the icons with each successive execution of the test. Pass 2 placements were 30% more accurate than pass 1; pass 3 placements were 54% more accurate than pass 2; and 68% more accurate than pass 1.

Examination of the data also shows that these results might have been even more dramatic except for some problems experienced by the second group of experienced users during passes 1 and 2. It should also be noted that this same group produced perfect results by pass 3 execution. Although in other circumstances this group’s data might be removed as an outlier, due to the very small test population, all data was retained.



**Figure 29 – Collaborative Benchmark 3 Spatial Awareness Offsets by Group**

Table 39 presents the results of a comparison of the spatial awareness (Benchmark 3) test results for individual vs collaborative users. Inspection of the table shows that the offsets reported by both populations are not normally distributed. Therefore, a suitable nonparametric comparison method (Mann-Whitney) is employed to make the test of the hypothesis of equal medians ( $H_0$ ) as a measure of central tendency [97].

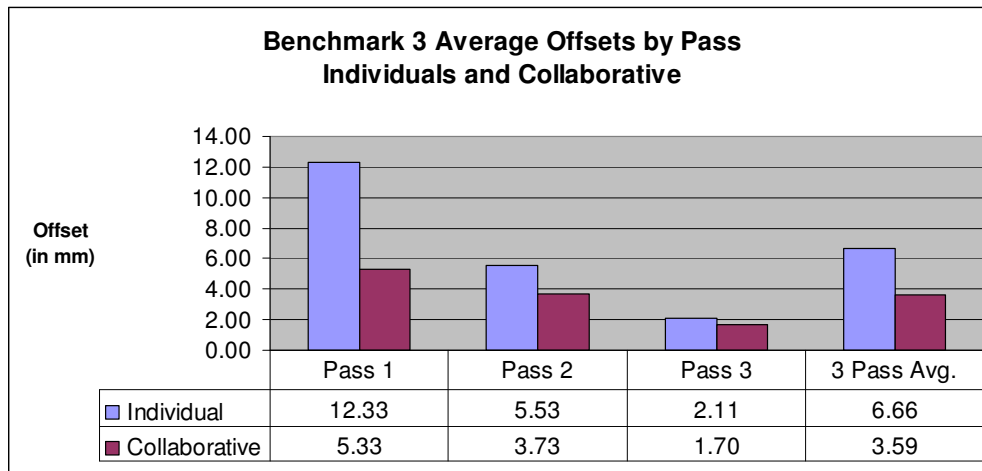
As discussed earlier, the key result of the test is a P value that answers the question: If the populations really have the same median, what is the chance that random sampling would result in means as far apart (or more so) as observed? Thus, if the P value is small, one can reject the idea that the difference is a coincidence, and conclude that the populations have different medians. However, if the P value is large (as is the Benchmark 3 comparison), the data do not provide any reason to conclude that the overall medians differ and  $H_0$  is accepted [97].

**Table 39 – Benchmark 3 Offsets Comparisons – Individuals vs Collaborative**

<b>Stereoscopic Voice/Glove Benchmark 3 Comparison Individual User vs Collaborative Group</b>		
	<b>Individual</b>	<b>Collaborative</b>
Mean	2.86	1.93
Standard deviation (SD)	4.05	3.29
Sample Size	21	5 teams of 3
Std. error of means (SEM)	0.89	0.85
Lower 90% conf. limit	1.33	0.44
Upper 90% conf. limit	4.38	3.43
Minimum	0.00	0.00
Median (50th percentile)	0.00	0.00
Maximum	15.00	10.00
Normality test KS	0.3310	0.3885
Normality test P value	<0.0001	<0.0001
Passed normality test?	No	No
Unpaired t test with Welch Correction	0.4563	0.4563
Significant?	No	No

Figure 30 provides a comparison of Benchmark 3 spatial awareness offsets by pass for individual users and for collaborative groups. Note that there is a consistent improvement in locating actual icon placements from pass to pass for both individuals and in collaborative groups. On average, collaboration produced placements that were 46.1% closer to actual than placements by individuals. It is also worth noting that the pass 3 placements for individuals improved by 46% from pass 1 to pass 3 and improved by more

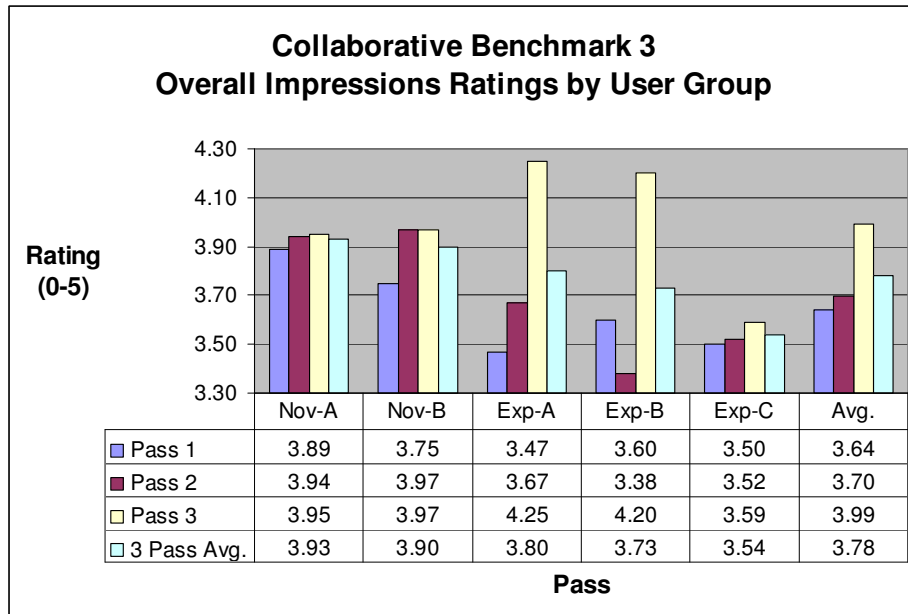
than 68 % as users collaborated on locating the placements. Unfortunately, the small test population and high variance does not allow conclusions of statistical significance.



**Figure 30 – Comparison of Individual vs Collaborative Spatial Awareness Offsets**

### ***7.5 Collaborative Benchmark 3 Ratings***

As with the individual Benchmark tests, the user groups were asked to rate the collaborative environment, via the survey questionnaire, after each pass of the test. Figure 31 presents a graphical view of each group's overall impressions of the collaborative operations of the stereoscopic voice/glove environment following each pass of the test. The descriptive statistics show that both populations fail tests for normality. For this reason, the Mann-Whitney test is employed to determine statistical significance. Table 40 provides the statistical analysis of the comparison showing that due to a high P value, no statistically significant difference in overall impressions of the environments can be established.

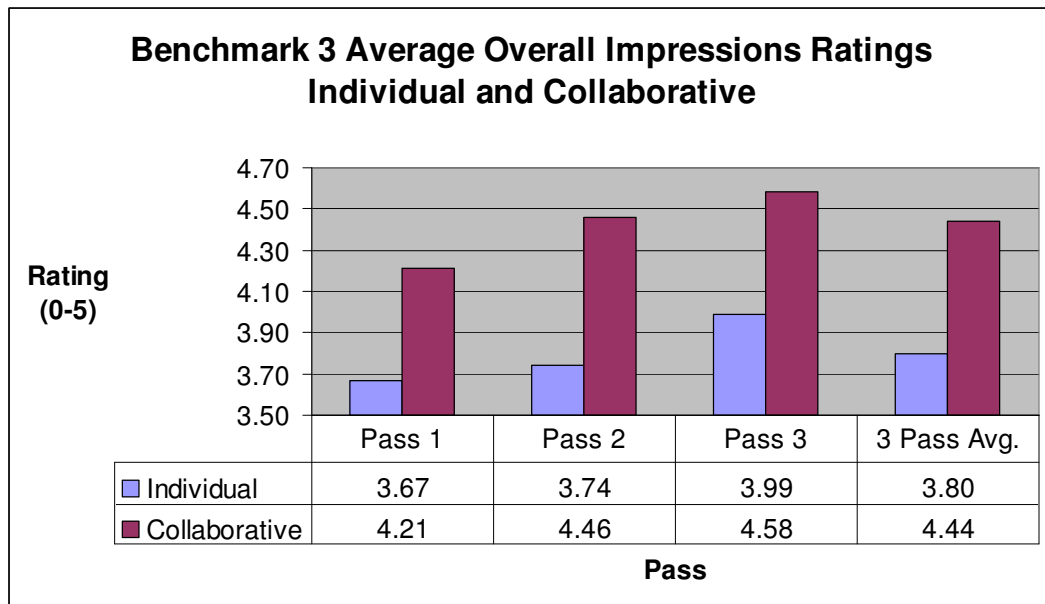


**Figure 31 – Collaboration Benchmark 3 Group Overall Impressions Ratings**

**Table 40 – Collaborative Benchmark 3 Overall Impressions Statistics**

Stereoscopic Glove Benchmark 3 Overall Impresions Ratings Comparisons Individual User vs Collaborative Group		
	Individual	Collaborative
Mean	4.58	4.23
Standard deviation (SD)	0.42	0.50
Sample Size	21	5 teams of 3
Std. error of means (SEM)	0.91	0.13
Lower 95% conf. limit	4.39	3.96
Upper 95% conf. limit	4.77	4.51
Minimum	4.00	3.59
Median (50th percentile)	4.60	4.01
Maximum	5.00	5.00
Normality test KS	0.2060	0.2054
Normality test P value	0.0200	0.0885
Passed normality test?	No	Yes
Unpaired, two-tailed Mann-Whitney Test P value = 0.5206 is NOT significant.		

Figure 32 provides a comparison of user overall impression ratings of spatial awareness for individuals vs collaborative groups



**Figure 32 – Comparison of Individual vs Collaborative Spatial Awareness Overall Impressions Ratings**

## **8 Benchmark 4 (Fault Identification)**

### ***8.1 Description***

In a typical design review process, a design space is presented to the reviewer(s) who examine the space for design flaws (faults). The purpose of this study is to help determine the applicability/usability of various user interfaces (both stereoscopic and non-stereoscopic) in improving this process. Based on the preliminary results of the previous Benchmark testing, a fourth Benchmark scenario was prepared to help provide some initial economic indicators to quantify possible cost savings associated with the use of a stereoscopic CAVE environment in the location and identification of faults within a design space. The scenario implemented and reported here is built upon the operations and scenarios developed for Benchmarks 1, 2, and 3.

Six separate versions of the AC&R space were developed. Each version contains ten distinct design faults similar to those prepared for Benchmark 2 (find/repair). However, the Benchmark 4 testing requires only that the users utilize the interface to locate and identify as many of these faults as possible in four minutes. As with the previous testing, each user searches each of three separate scenarios utilizing the 86" non-stereoscopic interface and the stereoscopic wand interface in the CAVE environment. The various scenario sequences were randomized (non-stereoscopic vs CAVE) and users were randomly assigned to start with either the non-stereoscopic interface or in the CAVE environment.

As each user progressed through the active scenario/environment locating and identifying faults, the specific fault and the elapsed time was recorded for the analysis. Although this method provides a significant quantity of data, for Benchmark 4, the key metric for comparison was the total number of faults found in each environment.

## **8.2 Benchmark 4 - Statistics**

### **8.2.1.1 Two Environment, Unpaired, Nonparametric Data**

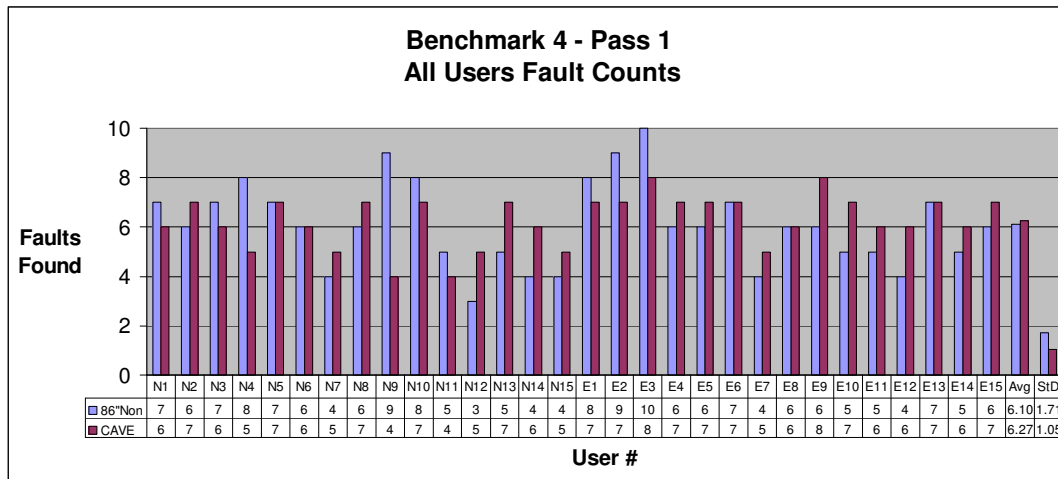
Since the users were presented six separate sets of faults within the design space, the statistical analyses performed are not the same as for the previous Benchmark tests. Benchmark 4 tests are not paired (i.e., the user sees different, albeit very similar, sets of faults in each environment). Thus unpaired comparison analysis methods are employed. Also, since only two environments are being compared and since very few of the resulting user data sets are normally distributed, the Mann-Whitney (nonparametric sum rank) test is used for the comparison of fault counts in the 86" non-stereoscopic and CAVE environments.

The Man-Whitney test is a nonparametric test that compares two unpaired groups. The method first ranks all of the fault counts from low to high, paying no attention to which group each value belongs. If two values are the same, each gets the average of the two ranks for which they tie. The smallest value gets the rank of 1 and the largest gets the rank of N where N is the total number of values in the two groups. The method then sums the ranks in each group. If the sums of the ranks are very different, the P-value should be small.

As with all of the previous tests, the P-value answers the question (within a 90% confidence level): *If the populations really have the same media, what is the chance that random sampling would result in a sum of the ranks as far apart (or more so) as observed? [97]*"

### **8.3 Benchmark 4 - Pass 1**

Figure 33 graphically presents a comparison of the pass 1 Benchmark 4 count of faults identified by each user groups. It should be noted that there is little difference between either the non-stereoscopic or CAVE environments in the quantity of faults identified upon user first exposure (pass 1) to the environment/scenario.



**Figure 33: Benchmark 4 Pass 1 Fault Counts**

The descriptive statistics data and Mann-Whitney test results presented in Table 41 show that at a 90% confidence level for pass 1, neither environment produces a statistically significant improvement in the facilitation of fault identification. However, it is interesting to note that the means of the two environments (86'' non-stereoscopic: 6.100 and CAVE: 6.267) differ by only 0.167 (of a possible 10). Also of interest is that the CAVE environment presents a more uniform range of fault counts (ie., a standard deviation almost 0.7 less than that for the non-stereoscopic interface) and shows 11.3% lower coefficient of variation in reported counts.

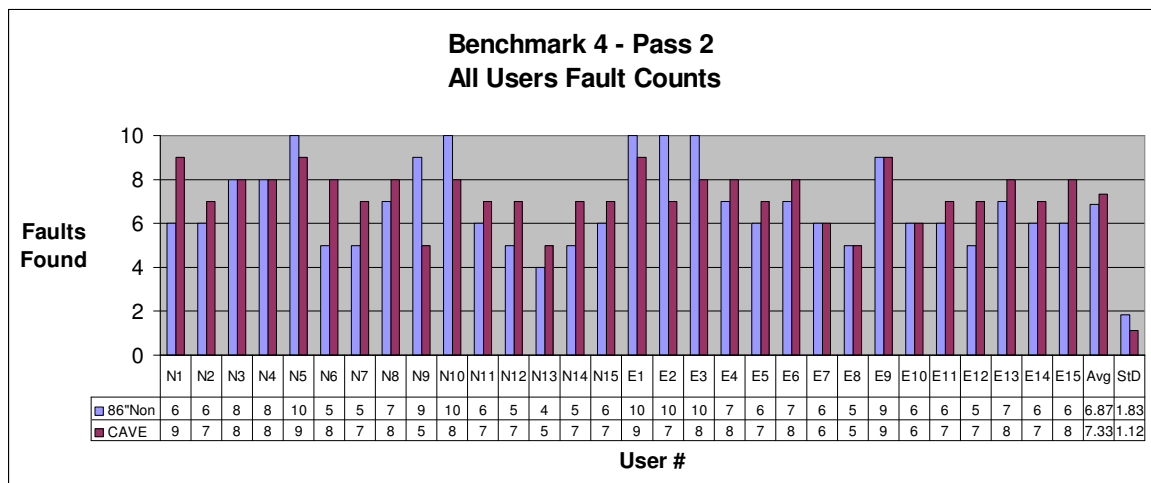
**Table 41 – Benchmark 4 Pass1 Fault Counts Statistics**

Benchmark 4 - Pass 1 Environments Comparison - Fault Counts		
B4p1-Counts	B4p1-86" Non	B4p1-CAVE
Mean	6.100	6.267
Standard deviation (SD)	1.709	1.048
Sample size (N)	30	30
Std. error of mean(SEM)	0.3120	0.1914
Lower 90% conf. limit	5.570	5.942
Upper 90% conf. limit	6.630	6.592
Minimum	3.00	4.00
Median (50th percentile)	6.00	6.50
Maximum	10.00	8.00
Normality test KS	0.1567	0.2579
Normality test P value	0.0583	P>0.0001
Passed normality test?	Yes	No
Coefficient of variation	28.02%	16.73%
Mann-Whitney test one-tailed p-value: 0.2312		
p-Value summary: Means are NOT significantly different		



## 8.4 Benchmark 4 - Pass 2

Figure 34 graphically presents a comparison of the pass 2 Benchmark 4 count of faults identified by the user groups. It should be noted that there is little difference between either the non-stereoscopic (mean: 6.867) or CAVE (mean: 7.333) environments in the quantity of faults identified upon user's 2nd exposure to the environment/scenario.



**Figure 34 – Benchmark 4 Pass 2 Fault Counts**

Table 42 shows that for pass 2 at a 90% confidence level, the CAVE environment produces a statistically significant improvement in the facilitation of fault identification. The statistics show results somewhat similar to those of pass 1 in that the Mann-Whitney test value indicates that the means are significantly different. Thus, for the pass 2 analysis, it is interesting to note that the means of the two environments (86'' non-stereoscopic: 6.867 and CAVE: 7.333) differ by only 0.466 (of a possible 10). Also of interest is that, as for pass 1, the CAVE environment presents a more uniform range of fault counts (ie., a standard deviation of 0.71 less than that for the non-stereoscopic interface) and also shows 11.3% lower coefficient of variation in reported counts.

**Table 42 – Benchmark 4 Pass 2 Fault Counts Statistics**

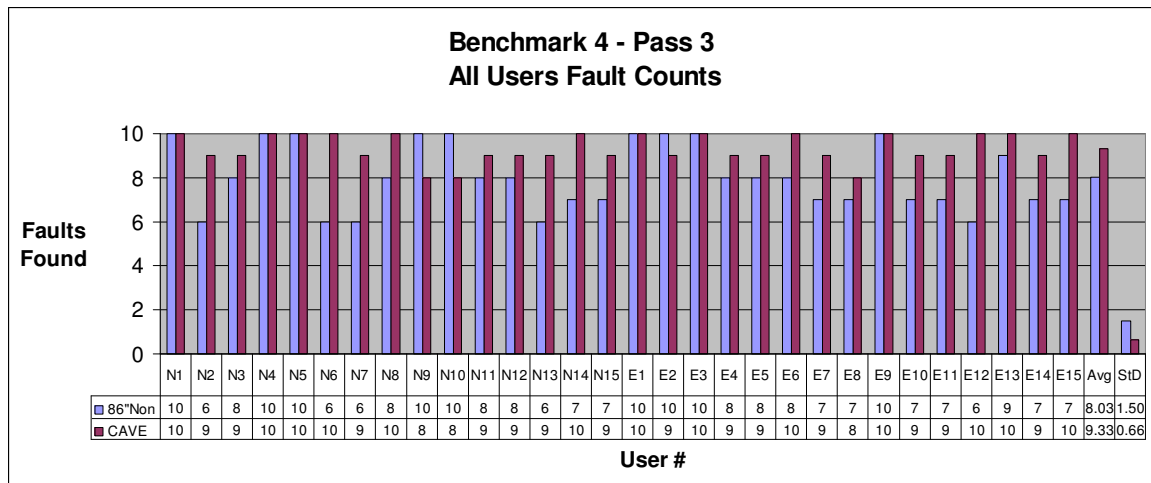
Benchmark 4 - Pass 2 Environments Comparison - Fault Counts		
B4p2-Counts	B4p2-86" Non	B4p2-CAVE
Mean	6.867	7.333
Standard deviation (SD)	1.833	1.124
Sample size (N)	30	30
Std. error of mean(SEM)	0.3347	0.2053
Lower 90% conf. limit	6.298	6.985
Upper 90% conf. limit	7.435	7.682
Minimum	4.00	5.00
Median (50th percentile)	6.00	7.00
Maximum	10.00	10.00
Normality test KS	0.2485	0.2168
Normality test P value	P>0.0001	0.0009
Passed normality test?	No	No
Coefficient of variation	26.70%	15.33%
Mann-Whitney test one-tailed p-value: 0.0508		
p-Value summary: Means ARE significantly different		

### **8.5 Benchmark 4 - Pass 3**

Figure 33 graphically presents a comparison of the pass 3 Benchmark 4 count of faults identified by the user groups. It should be noted that in the allotted time, users identified more than one additional fault using the CAVE environment than with the 86" non-stereoscopic interface and that, as might be expected, the identified fault counts increased by almost 2.5 from pass 1 to pass 3.

As seen in Table 43, for pass 3 neither the descriptive statistics show that while fault counts for the non-stereoscopic interface are normally distributed, those for the CAVE environment are not normally distributed, and therefore, nonparametric methods are used to complete the analysis.

The Mann-Whitney one-tailed test does show that the means of the two sets of test results are significantly different. Thus it can be said, with statistical significance at the 95% level, that the interface providing the higher number of faults identified in the allotted time is the "better" environment. In this case, the mean of the CAVE environment identified faults proves to be 1.3 higher than for the 86" non-stereoscopic interface, indicating that the CAVE environment is "better."



**Figure 35 – Benchmark 4 Pass 3 Fault Counts**

**Table 43 – Benchmark 4 Pass 3 Fault Counts Statistics**

Benchmark 4 - Pass 3 Environments Comparison - Fault Counts		
B4p3-Counts	B4p3-86" Non	B4p3-CAVE
Mean	8.033	9.333
Standard deviation (SD)	1.497	0.6609
Sample size (N)	30	30
Std. error of mean(SEM)	0.2733	0.1207
Lower 90% conf. limit	7.569	9.128
Upper 90% conf. limit	8.498	9.538
Minimum	6.00	8.00
Median (50th percentile)	8.00	9.00
Maximum	10.00	10.00
Normality test KS	0.2055	0.2768
Normality test P value	0.0023	P>0.0001
Passed normality test?	No	No
Coefficient of variation	18.63%	7.08%
Mann-Whitney test one-tailed p-value: 0.0006		
p-Value summary: Means ARE significantly different		

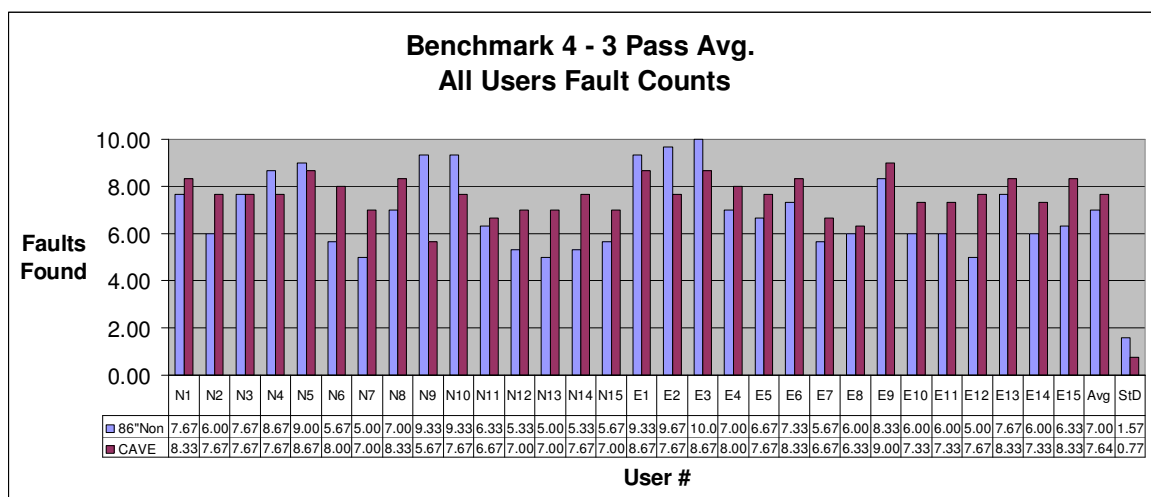
As with pass 1 and 2, the pass 3 results show that there is less variation in the data taken for the CAVE environment. This consistency of results represents another potential advantage of the CAVE environment.

## 8.6 Benchmark 4 - 3 Pass Average

Figure 36 graphically presents a comparison of the 3 pass average of Benchmark 4 count of faults identified by the user groups. It should be noted that use of the CAVE environment resulted in the identification of an average of 7.645

(of the possible 10) faults located in the design space and that use of the 86" non-stereoscopic interface resulted in the identification of an average of 7.0.

Table 44 shows that for the average of the 3 passes only the CAVE environment data set is normally distributed. However, since the 86" non-stereoscopic interface shows a nongaussian distribution, the Mann-Whitney method for comparing the two environments is required. The results of this unpaired, one-tailed test show a P-value of 0.0192 indicating that the means are significantly different and thus the interface producing the higher number of faults identified to be the "better" environment. Since the mean for the CAVE environment is 0.645 higher than that of the 86" non-stereoscopic interface, with statistical significance the CAVE environment can be said to be "better" for fault location and identification.



**Figure 36 – Benchmark 4 – 3 Pass Average Fault Counts**

**Table 44 – Benchmark 4 – 3 Pass Average Fault Counts Statistics**

Benchmark 4 - 3 Pass Avg. Environments Comparison - Fault Counts		
B43pA-Counts	B43pA-86" Non	B43pA-CAVE
Mean	7.000	7.645
Standard deviation (SD)	1.568	0.7675
Sample size (N)	30	30
Std. error of mean(SEM)	0.2863	0.1401
Lower 90% conf. limit	6.514	7.407
Upper 90% conf. limit	7.486	7.883
Minimum	5.00	5.67
Median (50th percentile)	6.50	7.67
Maximum	10.00	9.00
Normality test KS	0.1715	0.2768
Normality test P value	0.0246	P>0.1000
Passed normality test?	No	Yes
Coefficient of variation	22.40%	10.04%
Mann-Whitney test one-tailed p-value: 0.0192		
p-Value summary: Means ARE significantly different		

## 8.7 Benchmark 4 - All Trials

Table 45 presents the descriptive and comparative statistics of all trials (novice and experienced users across all passes). The table shows that the Mann-Whitney comparison of the fault counts across all 90 trials produces a one-tailed p-value of 0.0059. This indicates a statistically significant result at the 99% confidence level, showing that the CAVE environment allows users to locate and identify faults within a design space “faster” than the 86” non-stereoscopic environment using a traditional (keyboard/mouse) CAD interface. Specifically, these overall results show that users can locate/identify 0.611 additional faults (on average) during the 4 minute test durations.

**Table 45 – Benchmark 4 – All Trials Fault Counts Statistics**

Benchmark 4 - All Trials Environments Comparison - Fault Counts		
B4All-Counts	B4All-86" Non	B4All-CAVE
Mean	7.000	7.611
Standard deviation (SD)	1.848	1.584
Sample size (N)	90	90
Std. error of mean(SEM)	0.1948	0.1670
Lower 90% conf. limit	6.676	7.333
Upper 90% conf. limit	7.324	7.889
Minimum	3.00	4.00
Median (50th percentile)	7.00	7.00
Maximum	10.00	10.00
Normality test KS	0.1613	0.1613
Normality test P value	P>0.0001	P>0.0001
Passed normality test?	No	No
Coefficient of variation	26.40%	20.82%
Mann-Whitney test one-tailed p-value: 0.0059		
p-Value summary: Means ARE significantly different		

## 8.8 Benchmark 4 Fault ID Timings

Tables 46 and 47 present a comparison of the average times for locating and identifying the faults projected in the Benchmark 4 scenarios. Table 46 shows the results for the 86” non-stereoscopic (keyboard/mouse) interface and Table 47 shows the results using the CAVE environment. Note that the tables present the elapsed times per identified fault for the novice user group, experienced user group, and for all users.

**Table 46 – Benchmark 4 Elapsed Fault ID Elapsed Times – 86” Non-Stereo Interface**

Novice Users	86"Non							
	Pass 1		Pass 2		Pass 3		3 Pass Avg	
	Count	Time	Count	Time	Count	Time	Count	Time
Avg.	5.93	240.00	6.67	239.00	8.00	237.67	6.87	238.89
Sec./Fault	40.45		35.85		29.71		34.79	
Experienced Users	86"Non							
	Pass 1		Pass 2		Pass 3		3 Pass Avg	
	Count	Time	Count	Time	Count	Time	Count	Time
Avg.	6.27	239.67	7.07	238.33	8.07	236.33	7.13	238.11
Sec./Fault	38.24		33.73		29.30		33.38	
All Users	86"Non							
	Pass 1		Pass 2		Pass 3		3 Pass Avg	
	Count	Time	Count	Time	Count	Time	Count	Time
Avg.	6.10	239.83	6.87	238.67	8.03	237.00	7.00	238.50
Sec./Fault	39.32		34.76		29.50		34.07	

**Table 47 – Benchmark 4 Fault ID Elapsed Times – CAVE Environment**

Novice Users	CAVE							
	Pass 1		Pass 2		Pass 3		3 Pass Avg	
	Count	Time	Count	Time	Count	Time	Count	Time
Avg.	6.73	240.00	7.33	240.00	9.40	230.00	7.82	236.67
Sec./Fault	35.64		32.73		24.47		30.26	
Experienced Users	CAVE							
	Pass 1		Pass 2		Pass 3		3 Pass Avg	
	Count	Time	Count	Time	Count	Time	Count	Time
Avg.	5.58	240.00	7.33	240.00	9.33	231.83	7.64	237.28
Sec./Fault	43.03		32.73		24.84		31.04	
All Users	CAVE							
	Pass 1		Pass 2		Pass 3		3 Pass Avg	
	Count	Time	Count	Time	Count	Time	Count	Time
Avg.	6.16	240.00	7.33	240.00	9.37	240.00	7.73	240.00
Sec./Fault	38.99		32.73		25.62		31.03	

## 8.9 Economic Factors

### 8.9.1 VR System Capital and Operating Costs

In a university research environment, a typical capital investment cost of \$577,300 can be expected to implement a CAVE environment of front, left, right, and floor projection including the suspension grid structure, mirrors, RGB projectors, tracking system, cabling, computer system, software, shipping and installation [104]. It is assumed that the building preparation is \$22,700 yielding a capital cost of \$600K which represents a reasonable estimate for a university installation where buildings and other infrastructure already exist.

Operating costs include approximately \$60K/year for a research engineer. With fringe benefits of 25% and overhead costs at 40%, annual operating costs are estimated at \$105K.

### **8.9.2 VR System Cost Savings**

The design review process is a user and time intensive task. It is argued that by providing a more efficient interface, users become more efficient and are able to locate more faults. This reduces disruption and rework costs by the shipbuilder. It should be noted that Storch et al. [105] state that between 35% and 60% of product cost for US shipyards are labor costs, and 40% is taken as a conservative industry estimate. The Benchmark 4 test is designed to help quantify this cost difference.

As discussed in the introduction, this study is supported under funding provided by the Office of Naval Research (ONR) and was conducted at the University of New Orleans Gulf Coast Region Maritime Technology Center at the Northrop Grumman Ship Systems (NGSS), Avondale Shipyards, Avondale, LA. While specific NGSS production and fault-related cost data is proprietary, for the purposes of this study the following ship production costing assumptions are used as reported in published sources:

- Overall LPD prototype cost, \$1.76B [106]
- Overall DD(X) prototype cost, \$3.3B [107]
- Overall Virginia class prototype SSN cost, \$2.6B [108]
- Rework and disruption due to faults (flaws) is virtually all labor
- Labor cost is approximately 40% of ship cost [105].

As noted in Section 8.5 and Table 43, at the completion on the final pass of Benchmark 4 testing for all users, the CAVE environment allows reviewers to locate and identify 1.3 more faults within the 4 minute allocated search time than with the 86" traditional CAD interface (9.333 vs 8.033 faults). This result indicates a 16.2% overall increase in fault location and identification using the CAVE environment. Using the conservative fault cost data above, the following formula

can be used to estimate the savings possible resulting from the use of the CAVE environment during the basic design review process for fault identification:

cost avoidance = overall ship program cost

\* % ship program cost attributed to faults

\* % ship program cost as labor

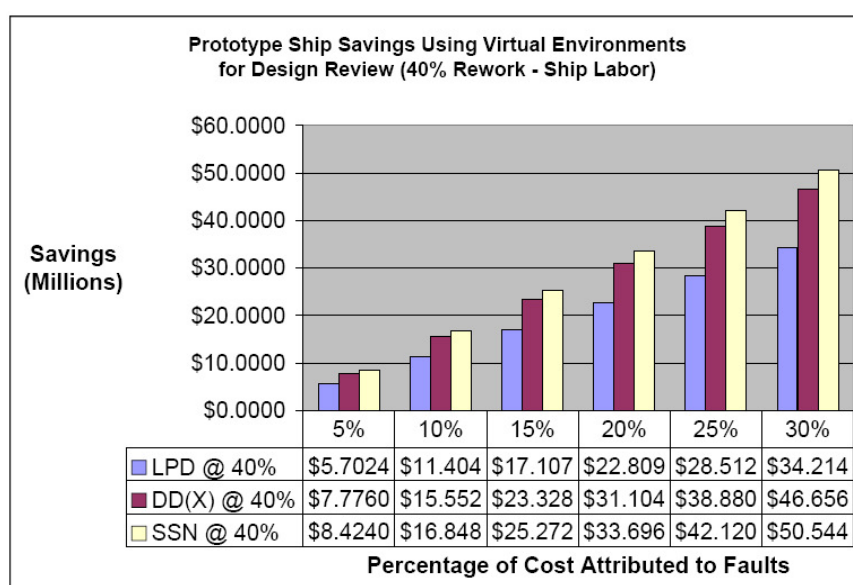
\* % increase in fault identification. (8-1)

Table 48 provides a range of possible savings by ship type based on the variable range of the percentage of cost attributed for fault identification as described by Storch, et al. [105].

**Table 48 – Prototype Ship Cost Savings Using Virtual Environments**

Prototype Ship Cost Savings Using Virtual Environments						
Ship Class	LPD		Virgina Class SSN		DD(X)	
Total Cost	\$1.76 Billion		\$2.60 Billion		\$3.30 Billion	
Costs Attributed to Faults	Low: 5%	High: 30%	Low: 5%	High: 30%	Low: 5%	High: 30%
Rework Labor Costs (40%)	\$35,200,000	\$211,200,000	\$52,000,000	\$312,000,000	\$66,000,000	\$396,000,000
Fault ID Savings (16.2%)	\$5,702,400	\$34,214,400	\$8,424,000	\$50,544,000	\$10,692,000	\$64,152,000

Figure 37 presents the savings possible for each class of ship fro the range of percentages of costs attributed to fault identification.



**Figure 37 – Potential LPD Prototype Rework Savings**



## 9 Conclusions and Recommendations

### 9.1 Competitive Analysis of Environments

Tables 49 presents a tabulation of the results of the ANOVA pair-wise comparisons of the interfaces for Benchmark 1, 2, and 3 measurements for both timings/offsets and user environment ratings for pass 3. The results presented show the comparison of the population means for each environment/interface as described by the software packages (Prisim [94] and SAS [99]) reported for a 90% confidence level.

**Table 49 – ANOVA Pair-wise Environment Comparisons Summary**

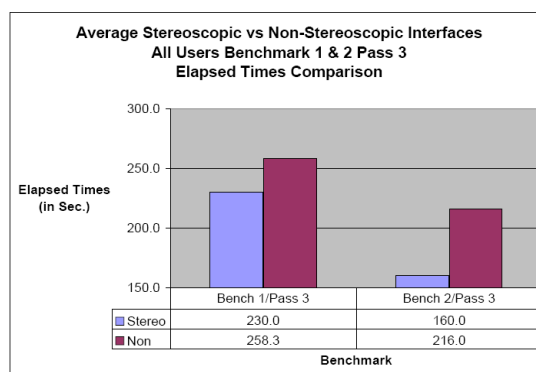
Pass 3 Statistically Significant Environment Comparisons Summary "Better" Environment						
Benchmark 1	Novice users		Experienced Users		All Users	
	Elapsed Times	Overall Ratings	Elapsed Times	Overall Ratings	Elapsed Times	Overall Ratings
Glove vs Wand	Wand	Glove	Wand	Neither	Wand	Glove
Glove vs 86"Non	Neither	Glove	Neither	Glove	Neither	Glove
Glove vs W/S	Neither	Glove	Neither	Glove	Neither	Glove
Wand vs 86"Non	Wand	Wand	Wand	Wand	Wand	Wand
Wand vs W/S	Wand	Wand	Wand	Wand	Wand	Wand
86"Non vs W/S	86" Non	86" Non	Neither	Neither	Neither	Neither
Benchmark 2	Novice users		Experienced Users		All Users	
	Elapsed Times	Overall Ratings	Elapsed Times	Overall Ratings	Elapsed Times	Overall Ratings
Glove vs Wand	Wand	Neither	Neither	Neither	Neither	Neither
Glove vs 86"Non	Glove	Glove	Glove	Glove	Glove	Glove
Glove vs W/S	Glove	Glove	Glove	Glove	Glove	Glove
Wand vs 86"Non	Wand	Wand	Wand	Wand	Wand	Wand
Wand vs W/S	Wand	Wand	Wand	Wand	Wand	Wand
86"Non vs W/S	Neither	Neither	Neither	Neither	W/S	Neither
Benchmark 3	Novice users		Experienced Users		All Users	
	Distance Offsets	Overall Ratings	Distance Offsets	Overall Ratings	Distance Offsets	Overall Ratings
Glove vs Wand	Wand	Glove	Wand	Wand	Glove	Neither
Glove vs 86"Non	Glove	Glove	Glove	Neither	Glove	Glove
Glove vs W/S	Neither	Glove	Glove	Neither	Glove	Glove
Wand vs 86"Non	Wand	Neither	Neither	Wand	Neither	Wand
Wand vs W/S	Wand	Wand	Neither	Wand	Neither	Wand
86"Non vs W/S	W/S	86"Non	W/S	86" Non	W/S	86"Non

Generalizing the environment comparisons presented in Tables 48, one can infer that for all users the stereoscopic wand interface can be demonstrated to be only statistically “better” for navigation (Benchmark 1) for some cases and rated higher in user preference. Furthermore, the results for the stereoscopic voice/glove interface are mixed. However, for find and repair operations (Benchmark 2), a statistical basis for claims that the voice/glove and wand stereoscopic environments are better exist, and the stereoscopic interfaces are preferred by the users.

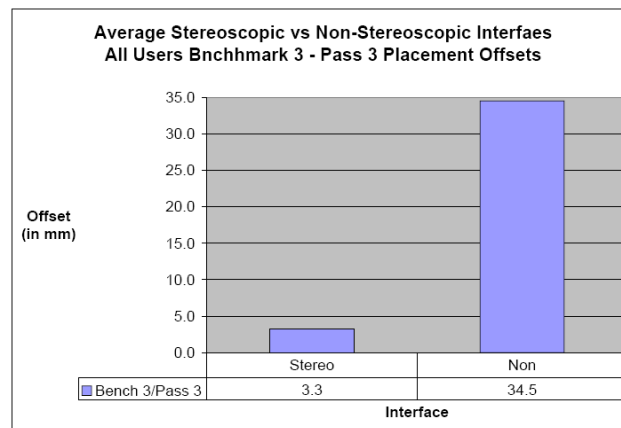
For Benchmark 3 (spatial awareness), the statistics show somewhat better results for the voice/glove interface in both objective and subjective measures. Additionally, under some circumstances the wand interface proves to be statistically better and preferred to non-stereoscopic methods.

## 9.2 Stereoscopic vs Non-Stereoscopic Environment Comparisons

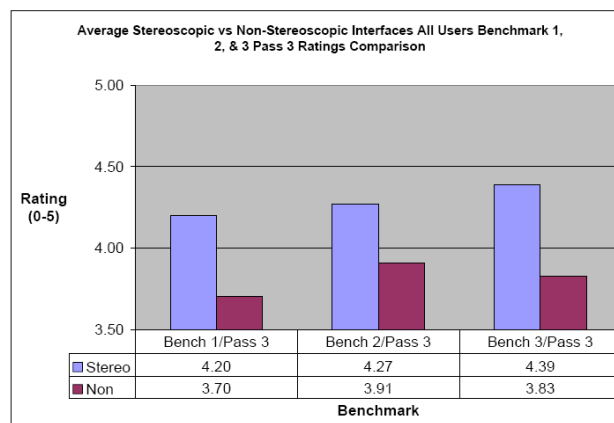
Figures 38, 39, and 40 present a graphical comparison of stereoscopic vs non-stereoscopic environment timing, offset, and overall ratings comparisons. Examination of Figure 38 shows improvement in elapsed times for navigation (Benchmark 1) and distinct improvement for manipulation (Benchmark 2) functions using the stereoscopic environments. Similarly, Figure 39 shows that the stereoscopic environments provide substantially improved spatial awareness, and Figure 40 presents a visual representation of user overall preference for the stereoscopic interfaces. Although the statistical analysis is limited with regard to conclusions reached through test of hypothesis procedures, Figures 38 through 40 strongly imply the superiority of stereoscopic interfaces over non-stereoscopic methods.



**Figure 38 – Stereoscopic vs Non-Stereoscopic Environment Times Comparisons**



**Figure 39 – Stereoscopic vs Non-Stereoscopic Environment Distance Offsets Comparisons**



**Figure 40 – Stereoscopic vs Non-Stereoscopic Environment Overall Ratings Comparisons**

### 9.3 Spatial Awareness

The user's knowledge and understanding of the space is an important component of the design process. The ability to measure user's spatial awareness is a difficult task which is addressed in this study in Benchmark 3 as the ability of the user to locate an icon's position in a unfamiliar space. In production environments improved spatial awareness translates not only to the ability to relate parts/faults within the space but also to enhanced perception of the role of the space and the interactions of components within

the space. This study show remarkable improvements (as much as 695%) in spatial awareness using stereoscopic interfaces over traditional, non-stereoscopic CAD environments. Although this improvement cannot be directly linked to design quality by statistics or logic, it is argued that a designer with significantly improved awareness should provide a higher quality design.

#### ***9.4 Collaborative Stereoscopic Voice/Glove Interface***

The collaborative ANOVA tests show extremely significant results indicating that there is a strong inference that the population means are unequal (within a 99.9% confidence level), thus indicating that  $H_0$  should be rejected and the alternate hypothesis that collaboration improves results accepted. Therefore, the very large improvement in spatial awareness for the voice/glove environment in the collaborative mode vs the individual (better than 207%) is statistically proven.

#### ***9.5 Fault Identification***

Benchmark 4 testing shows that stereoscopic environments appreciably improve the user's ability to identify faults within a design space. Table 45 shows that users in the CAVE environment identify 19.5 more faults per hour than with any other interface. In a complex design/manufacturing organization this capability, coupled with the increases in spatial awareness using stereoscopic methods in collaborative environments, can lead to significant cost savings (as noted in Equation 8-1 of Section 8.9).

#### ***9.6 Competitive Usability***

As discussed in the introduction, usability in software analysis normally indicates the user's preferences for the layout/interaction of a software package, where navigation traditionally relates to information flow from upper-left to lower-right of an interface screen or the data input verification functions for a specific application. Competing multiple interfaces with multiple control functions, especially in a virtual environment, provides an opportunity to investigate properties of the interfaces common to user/application success.

The author is unaware of published applications where competitive or comparative usability is the central tool employed. Although a claim that this method is significantly novel as a technique is not asserted, it is thought that the use of competitive or comparative usability may be a new tool with important benefits for software engineers.

## ***9.7 Additional Testing***

As can be seen in the previous sections on statistically significant pass 3 elapsed times/distance offsets and overall impressions ratings comparisons (Table 49), 75% of the Benchmark 1, 2, and 3 tests provide results that can be said to be statistically significant. While no great improvement in the number of statistically significant results can be noted in the larger populations provided by combining the novice (15 testers) and experienced (15 testers) users into the total group of 30, statistics theory does imply that larger sample sizes generally provide more significant results. With this in mind, future executions of the Benchmark 1, 2, and 3 scenarios with additional novice and experienced users is suggested, if funding in support of those tests can be obtained.

## ***9.8 Enhancements for Further Study***

During the course of the testing as documented in this report, users/testers, test administrators, and test developers often suggested possible enhancements to the interfaces that warrant further testing and evaluation. These enhancements ran the gambit from simply expanding the interface to include a “you-are-here” tracking map in one corner of the workstation or ImmersaDesk display to the multi-screen, immersive, VR CAVE environment to simpler changes such as blinking, color, position, etc. Although the research team would prefer to provide test data and evidence for each potential improvement, rigorous and professional usability testing required that the interfaces remain static during the actual summative tests. To minimize the impact of these constraints, a less formal initial test of some interfaces might be performed.

### ***9.8.1 Tracking Map***

The “you-are-here” tracking map enhancement was suggested by the user groups as an aid to navigation through a new space. As an initial implementation the developers

attempted to insert such a map, in 3D, into the existing ImmersaDesk environment. This initial implementation placed too large a compute burden on the system and slowed user response times to an unusable level. The enhancement was removed for the remainder of the testing due to the significant latency.

However, after considerable discussions with the users, administrators, and developers, the group came to the conclusion that the map enhancement might be implemented in a less burdensome manner, allowing more rapid computation. It is suggested that this enhancement be changed to a callable map activated/deactivated by the user via voice command. Thus, the compute load is not constant, and the user requests “you-are-here” help only on-demand.

### **9.8.2 Notes/Annotations Log**

Initial discussions between users and developers did not suggest a need/use for a notes or annotations page. However, during testing, the users expressed a desire to make notes about the design/space. As the testing progressed into collaborative groups this ability was more and more frequently requested. With the voice/glove interface the notes/annotations can reference the voice file for detailed descriptions/annotations.

### **9.8.3 CAVE-to-ImmersaDesk and CAVE-to-CAVE Collaboration**

With the availability of the CAVE and the 86” ImmersaDesk environments, comparison testing should be expanded to include design reviewers in each environment simultaneously. Control mechanisms for “who’s driving” (ie., which environment/user has the movement/manipulation functions) and means for indicating and transferring these controls would be developed and the scenarios of Benchmarks 1, 2, 3, and 4 might be used to evaluate these mechanisms and functionality.

CAVE-to-CAVE mechanisms evaluations would follow the methods proven in previous testing. A proposed Navy-Industry VE network would provide for testing in other existing, distributed, CAVE and/or ImmersaDesk VE environments at such locations as:

- NRL, Washington, DC
- H/HTC at NSWC Carderock, MD
- NUWC Newport, RI
- Naval Post-graduate School, Monterey, CA.

## **9.9 Closure**

With new technologies available using virtual environments, engineers, designers, and managers have capable new tools based on new applications of hardware and software. However, the use of this technology and technical standards should be integrated into methods that ensure the organization's technical progress. To accomplish this integration, the limitations and benefits of the technologies need to be explored and tested. This thesis has undertaken this as a principle focus.

### **9.9.1 Benchmarks, Measurements, and Ratings**

The study documented in this thesis includes the development of a survey rating questionnaire of 22 items in 4 categories (navigation, locating, manipulation, and general impressions) executed in each of 4 design environments (two non-stereoscopic and two stereoscopic). The study also included development of four distinct benchmark test scenarios designed to provide substantial objective measures (elapsed times, offsets, and fault counts). Each benchmark scenario was designed so that each test could be performed repeatedly. The study executed each benchmark three times with 30 total testers (15 novice and 15 experienced). These tests provided 2570 distinct objective data points and 5,940 distinct subjective rating points for analysis.

Benchmark environment test sequences were randomly assigned to user/testers to help reduce bias that might be introduced into the study based on non-random exposure to the test spaces. Additionally, during the course of collaboration testing, users were randomly assigned to test groups of 3 (within novice and experienced classes). Further, each user was afforded the opportunity to “drive” (operate the pointer or gestural interface) during one of the 3 executions of the collaborative benchmark.

### **9.9.2 Statistical Analysis for Proof of Significance**

Human users, equipment, and factors of randomness influence performance and measurements of performance. With these influences, statistics as a mathematical science dictates that this randomness must be investigated and resolved appropriately in order to make reliable inference from the test data and analysis. In this thesis, the management of randomness has been undertaken using well accepted and standard statistical methods. These methods proscribe that assessment of statistical significance at a given confidence level be achieved before hypotheses can be accepted or rejected. This study has followed these methods and procedures in statistical analysis, and the conclusions drawn from this work are achieved with statistical significance. The methodologies employed in achieving and delivering this statistical significance permits conclusions which rest on a solid mathematical basis. This reliance on rigorous, repeatable, testing with standard statistical analysis should be understood as valid proof of the conclusions presented for this study.

### **9.9.3 Summary**

The study was suggested to provide statistically significant proof of the hypothesis that new VR technologies could provide users with “better” tools for navigation, manipulation, and spatial awareness than exiting traditional, non-stereoscopic interfaces. The results of the study help verify this hypothesis.

- 1) Navigation is improved (lower elapsed times to navigate through a design space) using the stereoscopic environments prove to be 17.9% faster than using the non-stereoscopic interfaces. User survey overall subjective ratings of the environments show the stereoscopic voice/glove interface as preferred over the non-stereoscopic environments by as much as 11%.
- 2) Manipulation (find/repair) times are 26% faster using the stereoscopic environments over the non-stereoscopic interfaces and the voice/glove stereoscopic environment is rated by all the users 6.5% higher than either non-stereoscopic interface.
- 3) Remarkable (695%) improvements in spatial awareness are shown with the use of stereoscopic environments over either of the non-stereoscopic



interfaces with user surveys of overall environment ratings showing a 9.4% higher preference for the stereoscopic voice/glove interface over either non-stereoscopic interface.

- 4) Collaborative group find/repair times show that the stereoscopic voice/glove environment produces results over 3 times faster than did individual users. However, users overall impressions ratings of the collaborative methods were 3.2% lower than working as individuals.
- 5) Collaborative group spatial awareness results using the stereoscopic voice/glove interface show a 46.1% improvement over individual results with overall impressions ratings 14.8% higher than as individuals.
- 6) The CAVE environment allowed users to identify more faults in test space over a 4 minute timeframe than using the 86" non-stereoscopic interface, producing an overall 16.2% overall increase in fault location and identification.

#### **9.9.4 Conclusion**

This study has improved the understanding of the benefits of immersive environments by utilizing usability analysis methods to compare multiple stereoscopic, immersive environments with traditional, non-stereoscopic CAD interfaces. The rigorous usability analysis methods required the development of repeatable benchmark scenarios and user survey questionnaire that can be used to extend understanding benefits by using new environments/enhancements for comparison against traditional systems.

The study shows that the use of stereoscopic interfaces in the design review process can significantly improve the time that users spend in navigating and manipulating objects within the design space. Moreover, the study shows that use of the stereoscopic environments dramatically (695%) improves user spatial awareness.

Finally, the study presents an economic analysis that shows a significant cost savings garnered from a demonstrated 16.2% overall increase in fault location and identification using an immersive CAVE environment for design reviews. These cost savings per ship program are significantly greater than the acquisition and operating costs

of the virtual environments. This economic argument represents conclusive evidence that investment in stereoscopic systems is well justified.

## Bibliography

- [1] Johnson, Michael W., "Boeing Phantom Works - personal communication on VR design," Dr.George Hazelrigg of the National Science Foundation 1-14-2002.
- [2] Dave Phillips and Mike Tull, "Boeing Beats Schedule Again; Delivers X-32B Forebody Early," Boeing Co. New Release, St. Louis, MO: 6-29-1999.
- [3] National Institute of Standards and Technology , "Visualization and Virtual Reality for Manufacturing", <http://ovrt.nist.gov/home.html>, <http://ovrt.nist.gov/home.html>, 7-12-2004.
- [4] D. Maxwell, Rob King, and Alley Butler, "Design Synthesis in a Virtual Environment," in *Proceedings, ASME Computers and Information in Engineering* Pittsburgh, PA: 2001.
- [5] D. Maxwell, "Thesis: Design Synthesis in a Virtual Environment." MS ME, Louisiana Tech University, 2001.
- [6] J. Nielsen, *Usability Engineering*, 1st ed. San Diego, CA: Academic Press, 1993.
- [7] J. D. Gould and C.Lewis, "Designing for Usability: Key Principles and What designers Think," *Communications of the ACM*, vol. 28, no. 3, pp. 300-311, 1985.
- [8] ISO and (International Standards Organization), "Guidance on Usability," 2003.
- [9] D. Hix, J.Edward Swan, Joseph Gabbard, Mike McGee, Jim Durban, and Tony King, "User-Centered Design and Evaluation of a Real-Time Battlefield Visualization Virtual Environment," in *Proceedings, IEEE Virtual Reality* 1999, pp. 96-103.
- [10] J. Gabbard, Deborah Hix, and J.Edward Swan, "User Centered Design and Evaluation of Virtual Environments," *IEEE Computer Graphics and Applications*, vol. 19, no. 6, pp. 2-10, Dec.1999.
- [11] J. Gabbard, Kent Swartz, Kevin Ritchey, and Deobrah Hix, "Usability Evaluation Techniques: A Novel Method for Assessing the Usability of an Immersive Medical Visualization VE," in *VWSIM'99 Proceedings* San Francisco, CA: Visual Worlds and Simulation, 1999.
- [12] J. Nielsen, "Iterative User Interfaces Design," *IEEE Computer*, vol. 26, no. 11, pp. 32-41, Nov.1993.

- [13] I. Poupyrev, Suzanne Weghorst, Mark Billinghurst, and Tadao Ichikawa, "A Framework and Testbed for Studying Manipulation Techniques for Immersive VR," in *Proceedings of the ACM Conference on Virtual Reality Systems and Technology* Lausanne, Switzerland: 1997, pp. 21-28.
- [14] P. Thomas and Robert D. Macredie, "Introduction to the New Usability," in *ACM Transactions on Computer-Human Interaction*, 9 ed 2002, pp. 69-73.
- [15] Sunderland, Ivan E., "The Ultimate Display," Invited lecture, 1965 IFIP Congress, 1965.
- [16] F. P. Brooks, "Is There Any Real Virtue in Virtual Reality?," in *Public lecture cosponsored by the Royal Academy of Engineering and the British Computer Society*, London 1994.
- [17] F. P. Brooks, "What's Real About Virtual Reality?," in *Keynote Address 1999 IEEE Virtual Reality Conference*, 1999.
- [18] Bellis, Mary, "Inventors of the Modern Computer," About, Inc., <http://inventors.about.com/library/weekly/aa081898.htm>, 2005.
- [19] L. Rosenberg, "The Effect of Interocular Distance upon Operator Performance using Stereoscopic Displays to Perform Virtual Depth Tasks," in *IEEE Virtual Reality Annual International Symposium, 1993*. IEEE, Ed. Seattle, WA: IEEE, 1993, pp. 27-32.
- [20] C.-C. Chu, Tushar Dani, and Rajit Gadh, "Evaluation of Virtual Reality Interface for Product Shape Designs," *IIE Transactions*, vol. 30, no. 7, pp. 629-643, July 1998.
- [21] J. LaViola, "Whole-Hand and Speech Input in Virtual Environments," MS Thesis, Computer Science, Brown University, 1999.
- [22] K. Herndon, Andeies Van Dam, and Michael Gleicher, "The Challenges of 3D Interaction," in *Conference on Human Factors in Computing Systems*, 26 ed Boston: ACM, 1994, pp. 36-43.
- [23] K. Kaur, N. Maiden, and A. Sutcliffe, "Design Practice and Usability Problems with Virtual Environments," in *Virtual Reality World 96 Conference*, 11 ed Stuttgart, Germany: IDG Conferenes and Seminars, 1996, pp. 403-426.
- [24] University of Nottingham, CS Dept. Communications Research Group, "COVEN - Collaborative Virtual Environments - Usage Evaluation," in *Public Deliverable Report of ACTS Project N.AC042* 1997.
- [25] Kaur, K., Alistair Sutcliffe, and Neil Maiden, "A Design Advise Tool Presenting Usability Guidance for Virtual Environments," King's Manor, University of York, UK: British HCI Group, 9-30-1999.

- [26] Baum, Stephen, James Boudreaux, Neil Bourassa, and Jay Jenkins, "An Information technology Blueprint for the Twenty-first Century Warship," AN Information Technology Blueprint, [www.pms317.navy.mil/news/aninformationtechnologyblueprint](http://www.pms317.navy.mil/news/aninformationtechnologyblueprint), 1998.
- [27] J. Nomura and Kazuya Sawada, "Virtual Reality Technology and Its Applications to Industrial Use," *Control Engineering Practice*, vol. 7, no. 11, pp. 1381-1394, 1999.
- [28] C. Ware, Kevin Arthur, and Kellogg Booth, "Fish Tank Virtual Reality," in *Proceedings of the Conference on Human Factors in Computing Systems* Amsterdam, Netherlands: 1993, pp. 37-42.
- [29] R. Bolt and Edward Herranz, "Two-Handed Gesture in Multi-Modal Natural Dialog," in *Proceedings of the ACM Symposium on User Interface, Software, and Technology*, Monterey, CA: ACM, 1992, pp. 7-18.
- [30] A. Wexelblat, "An Approach to Natural Gesture in Virtual Environments," *ACM Transactions on Computer-Human Interaction*, vol. 2, no. 3, pp. 179-200, 1995.
- [31] C. Cruz-Neira, Daniel J.Sandin, Thomas A.DeFanti, Robert V.Kenyon, and John C.Hart, "The Cave: Audio Visual Experience Automatic Virtual Environment," *Communications of the ACM*, vol. 35, no. 6, pp. 65-72, June 1992.
- [32] C. Hand, "A Survey of 3D Interaction Techniques," *Computer Graphics Forum*, vol. 16, no. 5, pp. 269-281, 1997.
- [33] A. Stork and Martin Maidhof, "Efficient and Precise Solid Modeling using a 3D Input Device," in *Proceedings of the ACM Conference on Solid Modeling* Atlanta, GA: 1997, pp. 181-194.
- [34] S. N. Trika, P.Bannerjee, and R.L.Kashyap, "Virtual Reality Interfaces for Feature-Based Computer-Aided Design Systems," *Computer Aided Design*, vol. 29, no. 8, pp. 565-574, 1997.
- [35] D. Tushar and Rajit Gadh, "Creation of Concept Shape Designs via a Virtual Reality Interface," *Computer Aided Design*, vol. 29, no. 8, pp. 555-563, Aug. 1997.
- [36] S. Jayaram, Yong Wang, Uma Jayaram, Kevin Lyons, and Peter Hart, "Virtual Assembly Design Environment," in *Proceedings of the IEEE Virtual Reality Annual International Symposium* Los Alamitos, CA: 1999, pp. 172-179.
- [37] S. Jayaram, Hugh Connacher, and Kevin Lyons, "Virtual Assembly Using Virtual Reality Techniques," *Computer Aided Design*, vol. 29, no. 8, pp. 575-584, Aug. 1997.

- [38] S. C.-Y. Lu, M.Shpitalni, and Rajut Gadh, "Virtual and Augmented Reality Technologies for Product Realization," *CIRP Annals - Manufacturing Technology*, vol. 48, no. 2, pp. 471-495, 1999.
- [39] D. Bowman, Donald Johnson, and Larry Hodges, "Testbed Evaluation of Virtual Environment Interaction Techniques," in *Proceedings of the ACM Conference on Virtual Reality Systems and Technology* London, UK: 1999, pp. 26-33.
- [40] P. Evans, Judy Vance, and Veronica Dark, "Assessing the Effectiveness of Traditional and Virtual Reality Interfaces in Spherical Mechanism Design," *ASME Journal of Mechanical Design*, vol. 121, no. 4, pp. 507-514, Dec.1999.
- [41] R. Rosenfield, Xiaojin Zhu, Arthur Toth, Stefanie Shriver, Kevin Lenzo, and Alan Black, "Towards a Universal Speech Interface," in *Proceedings of the International Conference on Spoken Language Processing* Beijing, China: 2000.
- [42] D. V. Dorojkine and Judy M.Vance, "Implementing Speech Recognition in Virtual Reality," in *ASME Design Engineering Technical Conferences on Computers and Information Engineering*. ASME, Ed. Montreal, Canada: 2002.
- [43] K. Kiyokawa, Haruo Takemura, and Naokazu Yokoya, "SeamlessDesign for 3D Object Creation," *IEEE MultiMedia*, vol. 7, no. 1, pp. 22-33, 2000.
- [44] F. Kuester, Mark Duchaineau, Bernd Hamann, Ken Joy, and Kwan-Liu Ma, "DesignersWorkbench: Towards Real-Time Immersive Modeling," in *Proceedings of the SPIE, Stereoscopic Displays and Virtual Reality Systems* San Jose, CA: 2000, pp. 464-472.
- [45] M. Omata, Kentaro Go, and Atsumi Imamiya, "A Gesture-Based Interface for Seamless Communication between Real and Virtual Worlds," in *ERCIM 6th ERCIM WORKSHOP "USER INTERFACES FOR ALL,"* European Consortium for Informatics and Mathematics, Ed. Florence, Italy: ERCIM, 2000.
- [46] W. Broll and Thomas Schardt, "The Virtual Round Table - A Collaborative Augmented Multi-User Environment," in *ACM Proceedings of the 3rd International Conference on Collaborative Virtual Environments* ACM, 2000, pp. 39-45.
- [47] J. Vance and Sergei Volkov, "Virtual Reality Techniques to Improve Product Design," in *Design and Manufacturing Grantees Conference Proceedings, January 2001*, CD-ROM ed NSF, 2001.
- [48] A. Ansar, Denilson Rodrigues, Jaydev P.Desai, Kostas Daniilidis, Vijay Kumar, and Mario F.M.Campos, "Visual and Haptic Collaborative Tele-Presence," *Computers and Graphics*, vol. 25, pp. 789-798, 2001.

- [49] Alan Hedge, "Cornell University Ergonomics Web," Cornell Human Factors and Ergonomics Research Group, <http://ergo.human.cornell.edu>, 11-28-2005.
- [50] B. Connor, Matt Cutts, Henry Fuchs, Marco Jacobs, Russ Fish, Brian Loss, Rich Risenfeld, Loring Holden, Lee Markosian, and Greg Turk, "An Immersive Tool for Wide-Area Collaborative Design," in *TeamCAD, the First Graphics Visualization, and Usability (GVU) Workshop on Collaborative Design* Atlanta, GA: TeamCAD, the First Graphics Visualization, and Usability (GVU) Workshop on Collaborative Design, 1997, pp. 139-143.
- [51] R. Sky and Ralph Buchal, "Modeling and Implementing Concurrent Engineering in a Virtual Collaborative Environment," *Concurrent Engineering: Research and Applications*, vol. 7, no. 4, pp. 279-289, Dec.1999.
- [52] H. Y. Kan, Vincent Duffy, and Chuan-Jun Su, "An Internet Virtual Reality Collaborative Environment for Effective Product Design," *Computers in Industry*, vol. 45, pp. 197-213, 2001.
- [53] S. Woo, Eunjoo Lee, and Tsuyoshi Sasada, "A Multi-User Workspace as a Medium for Communication in Collaborative Design," *Automation in Construction*, vol. 10, pp. 303-308, 2001.
- [54] S. Woo, Eunjoo Lee, and Tsuyoshi Sasada, "The evaluation system for design alternatives in collaborative design," *Automation in Construction*, vol. 10, no. 3, pp. 295-301, Mar.1 A.D.
- [55] R. Lea, Yasuaki Honda, and Kouichi Matsuda, "Issues in the Design of a Scalable Shared Virtual Environment for the Internet," in *Proceedings of the 30th Hawaii International Conference on System Sciences* 1997.
- [56] I. Horvath, Joris Vergeest, and Zoltan Rusak, "Six Ingredients of Collaborative Virtual Design Environments," in *International Design Conference - Design 2002*, Dubrovnik: 2002.
- [57] S. A. Gill and R.A.Ruddle, "Using Virtual Humans to Solve Real Ergonomic Design Problems," in *Proceedings of the International Conference on Simulation* York, UK: 1998, pp. 223-229.
- [58] L. Anderson, James Esser, and Victoria Interrante, "A Virtual Environment for Conceptual Design in Architecture," in *ACM International Conference Proceedings of the Workshop on Virtual Environments* Zurich, Switzerland: 2003, pp. 57-63.
- [59] T. Disz, M.Papka, M.Pellegrino, and R.Stevens, "Sharing Visualization Experiences Among Remote Virtual Environments," in *Proceedings of the International Workshop on High Performance Computing for Computer Graphics and Applications*, 1995.

- [60] C. P. Spagno and Andreas M.Kunz, "Construction of a Three-sided Immersive Telecollaboration System," in *IEEE Conference on Virtual Reality*. IEEE, Ed. Los Angeles, CA: IEEE VR, 2003, pp. 37-44.
- [61] V. D. Lehner and T. A. DeFanti, "Distributed Virtual Reality: Supporting Remote Collaboration in Vehicle Design," *IEEE Computer Graphics and Applications*, vol. 17, no. 2, pp. 13-17, Apr.1997.
- [62] Rowell, Amy, "The Design Benefits of Group VR," *Computer Graphics World*, vol. 20, no. 2, pp. 21-28, Feb. 1997.
- [63] J. Tromp and Dave Snowden, "Virtual Body Language: Providing Appropriate User Interfaces in Collaborative Virtual Environments," in *Proceedings of the ACM Conference on Virtual Reality Software and Technology* Lausanne, Switzerland: 1997, pp. 37-44.
- [64] C. Anumba and J.A.Duke, "Internet and Intranet Usage in a Communications Infrastructure for Virtual Construction Project Teams," in *Proceedings of the 1997 6th IEEE Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises* Cambridge, MA: IEEE, 1997, pp. 56-61.
- [65] G. Bochenek, James Ragusa, and Linda Malone, "Integrating Virtual 3D Display Systems into Product Design Reviews: Some Insights from Empirical Testing," *International Journal of Technology Management*, vol. 21, pp. 340-352, 2001.
- [66] Vance, Judy and Tsung Pin Yeh , "A Virtual Environment for Engineering Design", [www.vrac.iastate.edu](http://www.vrac.iastate.edu/~jmvance/sen_opt.html), [http://www.vrac.iastate.edu/~jmvance/sen\\_opt.html](http://www.vrac.iastate.edu/~jmvance/sen_opt.html), 2001.
- [67] M. Fraser, Tony GLover, Ivan Vaghi, Steve Benford, Chris Greenhalgh, Joh Hindmarsh, and Christian Heath, "Revealing the Realities of Collaborative Virtual Reality," in *ACM Proceedings of the 3rd International Conference on Collaborative Virtual Environments* San Francisco, CA: ACM, 2000, pp. 29-37.
- [68] E. Swing, "Adding Immersion to Collaborative Tools," in *Proceedings of the 5th Symposium on Virtual Reality Modeling Language* Monterey, CA: ACM, 2000, pp. 63-68.
- [69] B. Kolarevic, Gerhard Schmitt, Urs Hirschberg, and David Kurmann, "An Experiment in Design Collaboration," *Automation in Construction*, vol. 9, pp. 73-81, 2000.
- [70] F. Tay and Chen Ming, "A Shared Multi-Media Design Environment for Concurrent Engineering Over the Internet," *Concurrent Engineering: Research and Applications*, vol. 9, no. 1, pp. 55-63, Mar.2001.



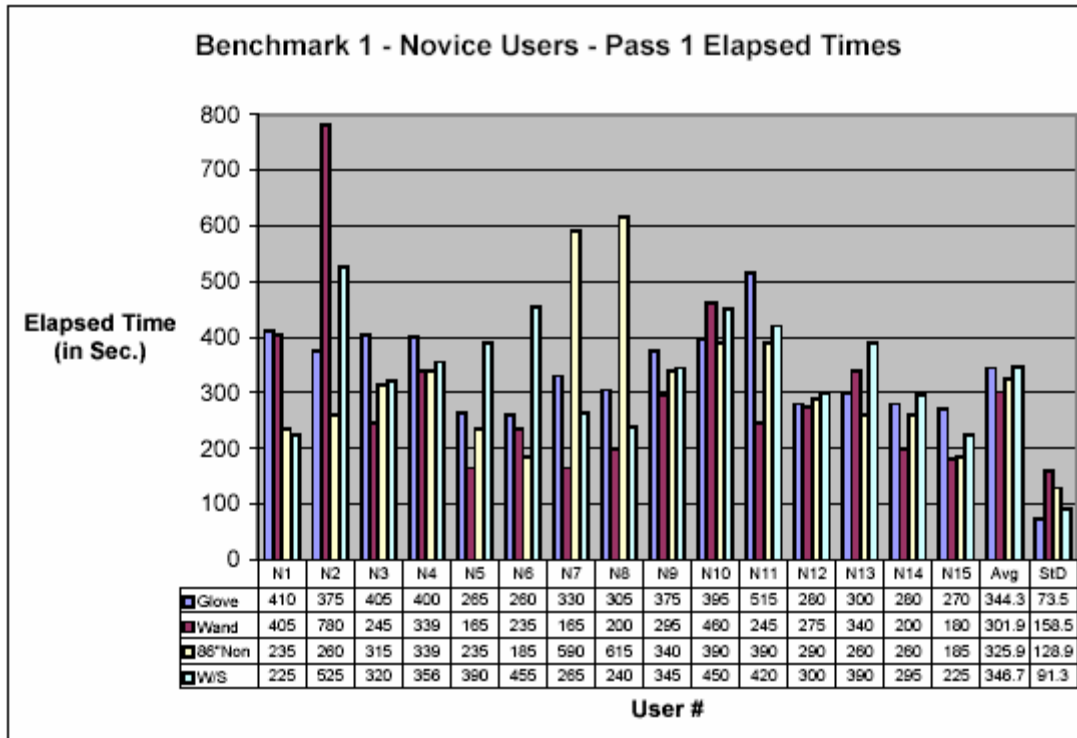
- [71] R. Bergman and J.D.Baker, "Enabling Collaborative Engineering and Science at JPL," *Advances in Software Engineering*, vol. 31, pp. 661-668, 2000.
- [72] Campbell, Dace A. and Maxwell Wells , "A Critique of Virtual Reality in the Architectural Design Process", [www.hitl.washington.edu](http://www.hitl.washington.edu/publications), <http://www.hitl.washington.edu/publications>, 5-23-2003.
- [73] K. A. Chipperfield and Judy Vance, "VRHose - Virtual Hydraulic Hose Routing Design," in *Proceedings of the Material Handling Division, American Society of Mechanical Engineering, MHD*. ASME, Ed. 2003, pp. 57-63.
- [74] S. Jayaram, Uma Jayaram, Rajit Gadh, Hari Srinivasan, and Judy Vance, "Assessment of VR Technology and Its Applications to Engineering Problems," *Journal of Computing and Information Science in Engineering*, vol. 1, pp. 72-83, Mar.2001.
- [75] J. Gaisemier, Oskar von Bohuszewicz, Gerrit Gehnen, and Michael Grafe, "Integration of Real IT-Applications in Virtual Manufacturing Systems," in *Proceedings of the 1996 ASME Design Engineering Technical Conferences and Computers in Engineering Conference* Irvine, CA: ASME, 1996, pp. 1-10.
- [76] R. Gupta, Daniel Whitney, and David Zeltzer, "Prototyping and Design for Assembly Analysis using Multimodal Virtual Environments," *Computer Aided Design*, vol. 29, no. 8, pp. 585-598, Aug.1997.
- [77] Gray, W. D. and Boehm-Davis, "Cognitive Analysis of Dynamic Performance: Cognitive Process Analysis and Modeling," <http://www.cogsci.rpi.edu/cogworks>, Human Factors and Ergonomics Society, Alberquerque, NM, 1997.
- [78] D. Hix and H.R.Hartson, "Usability Analysis," in *Developing User Interfaces – Ensuring Usability Through Product & Process*, New York, NY: John Wiley & Sons, Inc., 1993.
- [79] J. Nielsen, "Finding Usability Problems Through Heuristic Evaluation," in *Proceedings of the ACM CHI'92 Conference*, Monterey, CA: 1992, pp. 373-380.
- [80] J. Whiteside, J.Bennett, and K.Holtzblatt, "Usability Engineering: Our Experience and Evolution," in *Handbook of Human-Computer Interaction*. M.Helander, Ed. Amsterdam, North-Holland: Elsevier, 2003, pp. 791-817.
- [81] R. J. R. M. Jeffries, C. C.Wharton, and K.M.Uyeda, "User interface evaluation in the real world: A comparison of four techniques," in *Proceedings ACM CHI'91 Conference*, New Orleans, LA: 1991, pp. 119-124.

- [82] P. Sarvela and R.McDermott, *Health Education Evaluation and Measurement*. Saddlebrook, NJ: McGraw-Hill, 2001.
- [83] Gabbard, Joseph and Deborah Hix, "Taxonomy of Usability Characteristics in Virtual Environments - Final Report to the Office of Naval Research," csgrad.cs.vt.edu, <http://csgrad.cs.vt.edu/~jgabbard/ve/taxonomy/index.html>, 1997.
- [84] V. Paelke, Geiger, C., Rosenbach, and W., "Structured Design of Interactive Virtual and Augmented Reality Content," in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* Seoul, Korea: ACM, 2000, pp. 75-82.
- [85] K. Kaur, N.Maiden, and A.Sutcliffe, "Interacting with Virtual Environments: An Evaluation of a Model of Interaction," *Interacting with Computers*, vol. 11, no. 4, pp. 403-426, 1999.
- [86] J. Nielsen, "Enhancing the Explanatory Power of Usability Hueristics," in *Proceedings, ACM CHI'94 Conference* Boston, MA: 1994, pp. 152-158.
- [87] J. P. Chin, V.A.Diehl, and K.L.Norman, "Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface," in *ACM Proceedings of CHI Conference on Human Factors in Computing Systems* New York, NY: 1988, pp. 213-218.
- [88] Satter, Kurt, "Course Notes for Systems Analysis and Design, Management 6451, University of New Orleans," 2001.
- [89] C. Hall, Randy Stiles, and Varol Horwitz, "Virtual Reality for Training: Evaluating Knowledge Retention," in *Proceedings of the IEEE 1998 Virtual Reality Annual International Symposium* Atlanta, GA: IEEE, 1998, pp. 184-190.
- [90] A. A. Lund, "Measuring Usability with USE Questionnaire," in *STC Usability Special Interest Group Newsletter*, 8 ed Society for Technical Communication, 2001.
- [91] R. Witmer and Michael Singer, "Measuring presence in Virtual Environments: A Presence Questionnaire," *Presence*, vol. 7, no. 3, pp. 225-240, June1998.
- [92] J. Nielsen and Victoria Phillips, "Estimating the relative usability of two interfaces: heuristic, formal, and empirical methods compared," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* Amsterdam, The Netherlands: 1993, pp. 214-221.
- [93] Macnaughton, Donald B. "The Introductory Statistics Course: The Entity-Property-Relationship Approach", [www.matstat.com](http://www.matstat.com), <http://www.matstat.com/teach/eprt0130.htm>, 1-30-2002.
- [94] GraphPad Software, "Prisim for Windows," San Diego, CA: 2-5-2005.

- [95] SAS Institute, "SAS/IML User's Guide," 8.2 ed. SAS Institute SAS Institute, 1999.
- [96] H. C. Jr. Thode, "Lilliefors' Method," in *Testing for Normality* Marcel Dekker, 2002.
- [97] H. Motulsky, *The InStat Guide to Choosing and Interpreting Statistical Tests*. San Diego, CA: GraphPad Software, 2003.
- [98] GraphPad Software, "InStat," San Diego, CA: 2003.
- [99] SAS Institute, "SAS for Windows," 00M0 ed. SAS Institute SAS Institute, Cary, NC, 1999.
- [100] Datadesk Staff, "Cronbach's Alpha", Datadesk.com, <http://www.datadesk.com/support/template/stats.shtml>, 2005.
- [101] M. B. Miller, "Coefficient Alpha: A Basic Introduction From the Perspectives of Classical Test Theory and Structural Equation Modeling," *Structural Equation Modeling*, vol. 2, no. 3, pp. 255-273, 1995.
- [102] H. Motulsky, *Analyzing Data with GraphPad Prism*. San Diego, CA: GraphPad Software, 1999.
- [103] W. Mindenhall, R. Scheaffer, and D. Wackerly, *Mathematical Statistics With Applications*, 2 ed. Boston, MA: Duxbury Press, 1981.
- [104] Don Garwood, "UNO-Satter-112905 Fakesapce CAVE System Quotation," Marshalltown, IA: 11-29-2005.
- [105] R. L. Storch, Colin P. Harmon, Howard L. Bunch, and Richard C. Moore, *Ship Production*, 2nd ed Cornell Maritime Press, 1995.
- [106] Staff Writer New Orleans Times-Picayune, "Cost Overruns, Budget Uncertainties Hurting USN and Contractors," 3-18-2005.
- [107] Andrea Shalal-Esa, "Pentagon may scale back Navy DD(X) destroyer," Reuters, [www.DefenseTalk.com/news/publish/article\\_004040.php](http://www.DefenseTalk.com/news/publish/article_004040.php), 11-8-2005.
- [108] Wikipedia Staff, "Virginia Class Submarine," Wikimedia, [http://en.wikipedia.org/Virginia\\_class\\_submarine](http://en.wikipedia.org/Virginia_class_submarine), 11-11-2005.

## **Appendix A**

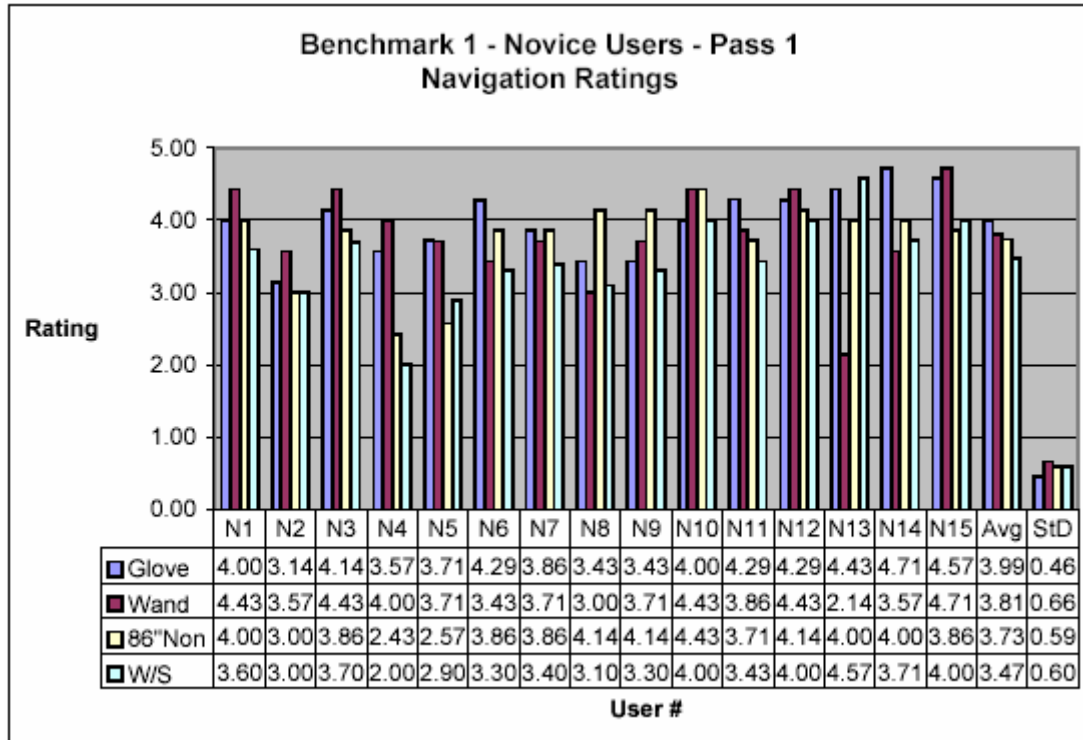
### Novice User Benchmark 1 (Navigation) Detail



**Figure A- 1: B1Np1Tim Novice User Elapsed Times**

**Table A- 1: B1Np1Tim Novice User Elapsed Times Statistics**

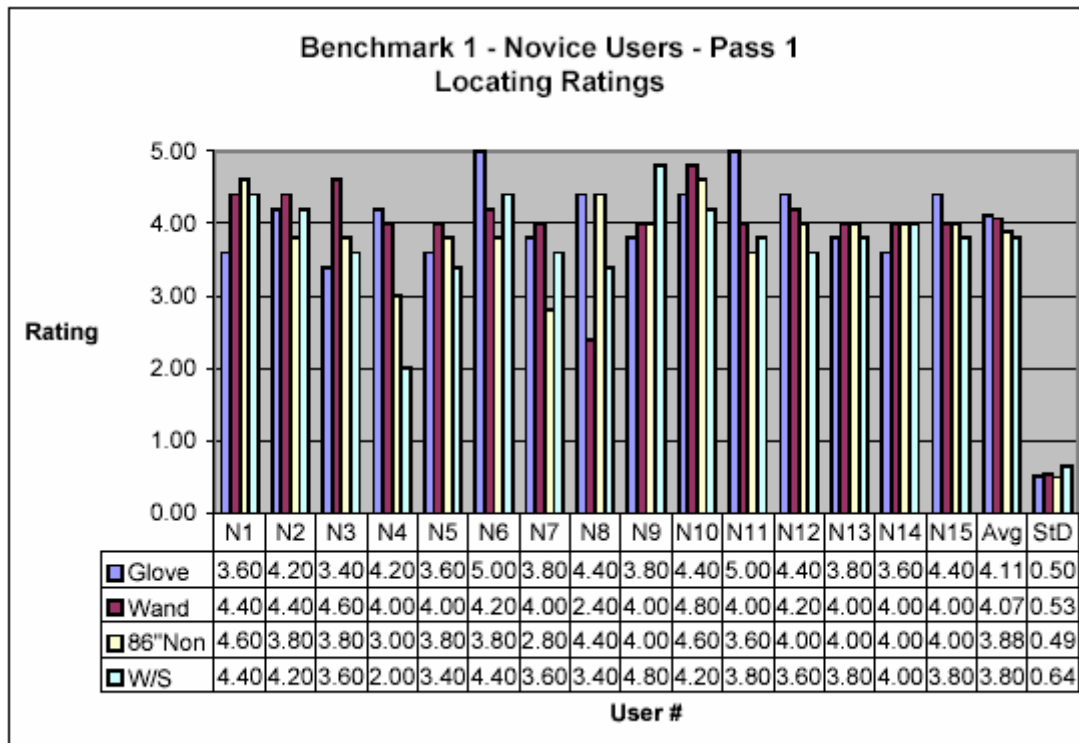
Benchmark 1, Novice Users, Pass 1, Elapsed Time:												
B1Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	15	344.3	73.43	260.0	515.0	> 0.1000	Yes	21.33%				
Wand	15	301.9	158.50	165.0	780.0	0.0895	Yes	52.50%				
86\"Non	15	325.9	128.90	185.0	615.0	> 0.1000	Yes	39.55%				
W/S	15	346.7	91.28	225.0	525.0	> 0.1000	Yes	26.33%				
B1Np1	Homogeneity of Variance					Roy's Greatest Root			F(3,12)		0.74	
	Bartlett's Test		Levene's Test		Equal Var ?	F(1,14)			Pr > F	Equal Means ?	Statistically Significant	Pr > F
	Chi-Sq	Pr > ChiSq	F Value	Pr > F		df	F(1,14)	Pr > F				
Glove vs Wand	7.3102	0.0069			No	14	1.20	0.2920	Yes	Var	Wand	14.04%
Glove vs 86\"Nor	4.0610	0.0439			No	14	0.29	0.6000	Yes	Var	86\"Non	5.65%
Glove vs W/S	0.6311	0.4270			Yes	14	0.01	0.6300	Yes	Neither		Neither
Wand vs 86\"Non	0.5751	0.4482			Yes	14	0.17	0.6827	Yes	Neither		Neither
Wand vs W/S	3.9256	0.0476			No	14	1.82	0.1984	Yes	Var	Wand	14.84%
86\"Non vs W/S	1.5782	0.2090			Yes	14	0.21	0.6534	Yes	Neither		Neither



**Figure A- 2: B1Np1Nav Novice User Navigation Ratings**

**Table A- 2: B1Np1Nav Novice User Navigation Ratings Statistics**

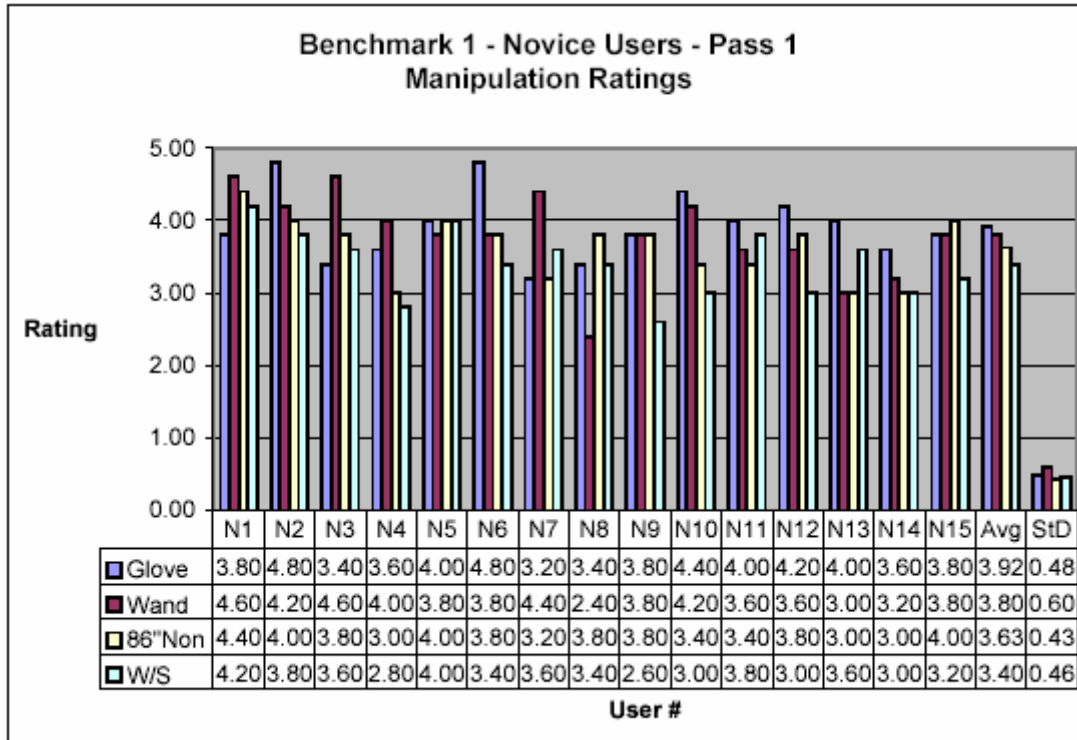
Benchmark 1, Novice Users, Pass 1, Navigation Ratings											
B1Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	3.991	0.460	3.14	4.71	> 0.1000	Yes	11.53%			
Wand	15	3.809	0.661	2.14	4.71	> 0.1000	No	17.36%			
86\"Non	15	3.733	0.589	2.43	4.43	0.0002	No	15.76%			
W/S	15	3.467	0.603	2.00	4.57	> 0.1000	Yes	17.39%			
B1Np1	Homogeneity of Variance				Roy's Greatest Root:				Pr > F		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	1.7417	0.1869	0.0760	0.3916	Yes	14	0.86	0.3704	Yes	Neither	Neither
Glove vs 86\"Non					Yes	14	3.00	0.1052	Yes	Neither	Neither
Glove vs W/S	1.1300	0.2878	0.1400	0.7106	Yes	14	20.03	0.0005	No	Means	Glove 15.11%
Wand vs 86\"Non					Yes	14	0.11	0.7433	Yes	Neither	Neither
Wand vs W/S	0.1153	0.7342	0.0100	0.9378	Yes	14	2.09	0.1707	Yes	Neither	Neither
86\"Non vs W/S					Yes	14	6.08	0.0272	No	Means	86\" Non 7.67%



**Figure A- 3: B1Np1Loc Novice User Locating Ratings**

**Table A- 3: B1Np1Loc Novice User Locating Ratings Statistics**

Benchmark 1, Novice Users, Pass 1, Locating Ratings												
B1Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	15	4.107	0.501	3.40	5.00	> 0.1000	Yes	12.19%				
Wand	15	4.670	0.527	2.40	4.80	< 0.0001	No	11.29%				
86\"Non	15	3.880	0.495	2.80	4.60	0.0245	No	12.74%				
W/S	15	3.800	0.641	2.00	4.80	> 0.1000	Yes	16.88%				
B1Np1		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)		0.87	Pr > F
		Bartlett's Test		Levene's Test		df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				1.2000	0.2822	Yes	14	5.15	0.0396	No	Means	Wand 13.71%
Glove vs 86\"Non		2.4135	0.1203			Yes	14	3.66	0.0764	No	Means	Glove 5.85%
Glove vs W/S		0.3133	0.5757			Yes	14	13.25	0.0027	No	Means	Glove 8.08%
Wand vs 86\"Non				0.2500	0.6205	Yes	14	0.06	0.8103	Yes	Neither	Neither
Wand vs W/S				0.1800	0.6786	Yes	14	3.19	0.0958	No	Means	Wand 22.89%
86\"Non vs W/S		0.7689	0.3806			Yes	14	2.64	0.1265	Yes	Neither	Neither

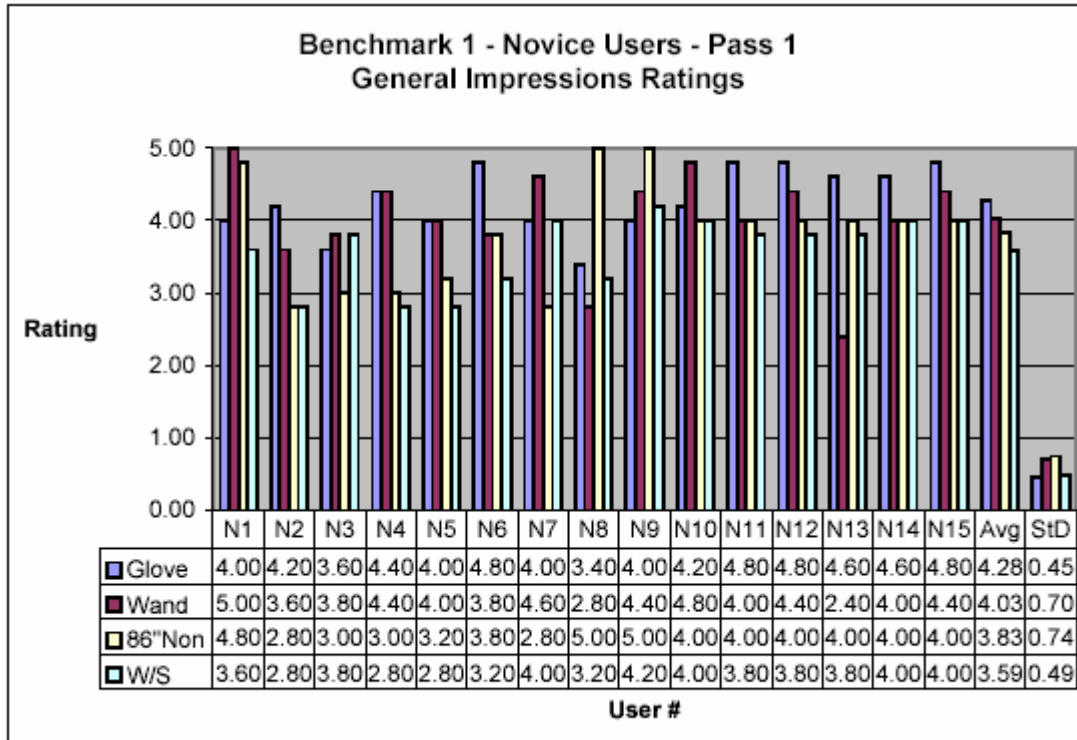


**Figure A- 4: B1Np1Mov Novice User Manipulation Ratings**

**Table A- 4: B1Np1Mov Novice User Manipulation Ratings Statistics**

Benchmark 1, Novice Users, Pass 1, Manipulation Ratings											
B1Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	3.920	0.477	3.20	4.80	> 0.1000	Yes	12.17%			
Wand	15	3.800	0.600	2.40	4.60	> 0.1000	Yes	15.79%			
86\"Non	15	3.627	0.433	3.00	4.40	0.0094	No	11.95%			
W/S	15	3.400	0.460	2.60	4.20	> 0.1000	Yes	13.52%			
B1Np1	Homogeneity of Variance					Roy's Greatest Root: F(3,12)			3.34	Pr > F	0.0561
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.7066	0.4006			Yes	14	0.40	0.5379	Yes	Neither	Neither
Glove vs 86\"Non			0.1900	0.3360	Yes	14	4.19	0.0598	No	Means	Glove 8.08%
Glove vs W/S	0.0180	0.8933			Yes	14	10.03	0.0069	No	Means	Glove 15.29%
Wand vs 86\"Non			1.3100	0.2628	Yes	14	1.15	0.3027	Yes	Neither	Neither
Wand vs W/S	0.9461	0.3307			Yes	14	5.38	0.0359	No	Means	Wand 11.76%
86\"Non vs W/S			0.1132	0.7366	Yes	14	3.36	0.0924	No	Means	86\" Non 6.68%

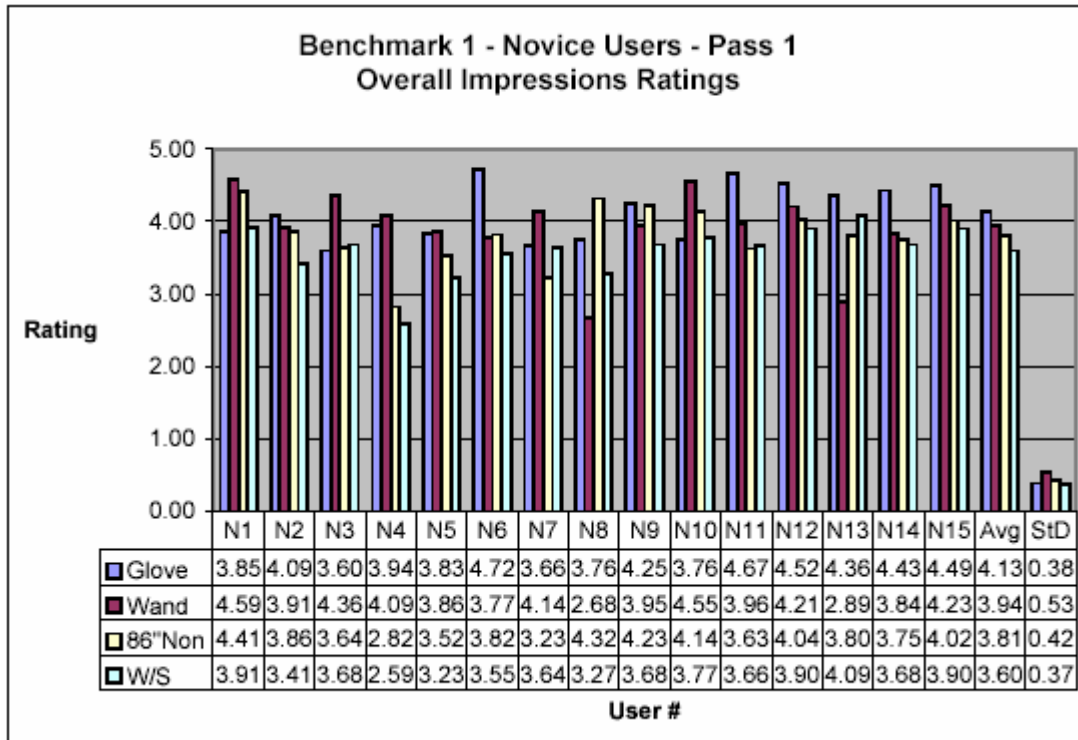




**Figure A- 5: B1Np1Gen Novice User General Impressions Ratings**

**Table A- 5: B1Np1Gen Novice User General Impressions Ratings Statistics**

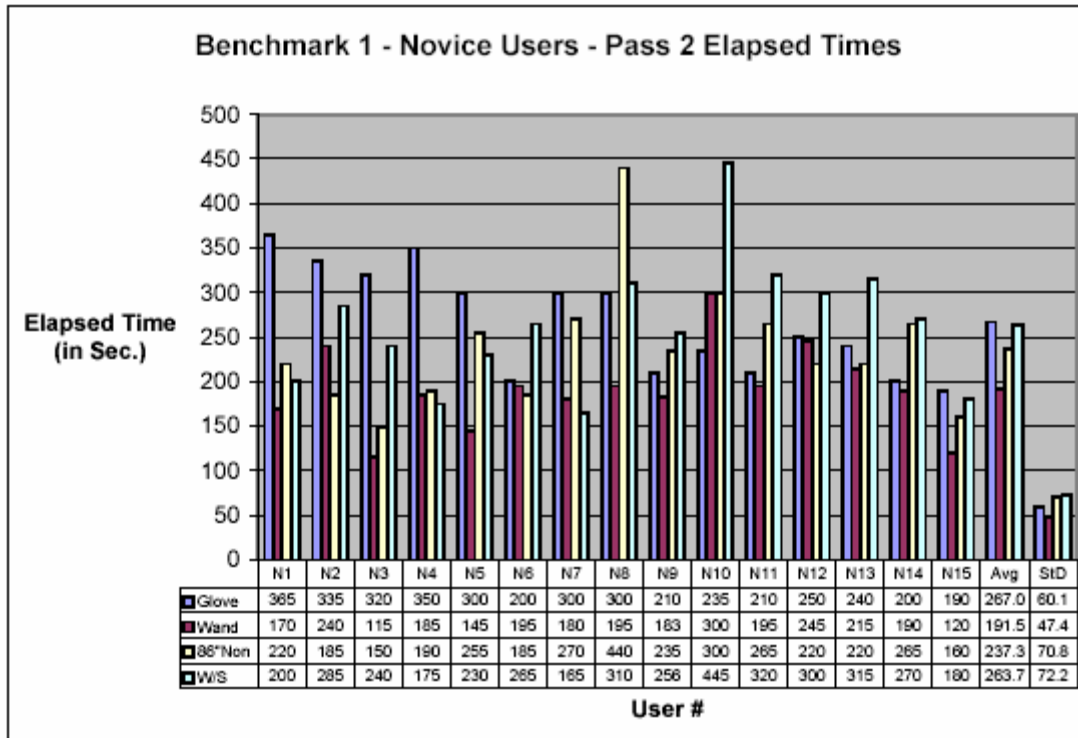
Benchmark 1, Novice Users, Pass 1, General Impressions Ratings										
B1Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	15	4.280	0.452	3.40	4.80	> 0.1000	Yes	10.57%		
Wand	15	4.027	0.701	2.40	5.00	> 0.1000	Yes	17.40%		
86\"Non	15	3.827	0.744	2.80	5.00	> 0.1000	Yes	19.44%		
W/S	15	3.587	0.463	2.80	4.20	0.005	No	12.90%		
Homogeneity of Variance				Roy's Greatest Root: F(3,12)			5.93	Pr > F		
B1Np1	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand	2.5079	0.1133			Yes	14	1.53	0.2372	Yes	Neither
Glove vs 86\"Non	3.2185	0.0728			Yes	14	3.87	0.0694	No	Means
Glove vs W/S			0.2000	0.6543	Yes	14	19.16	0.0006	No	Means
Wand vs 86\"Non	0.0491	0.8247			Yes	14	0.06	0.4707	Yes	Neither
Wand vs W/S			1.3600	0.2539	Yes	14	5.23	0.0383	No	Means
86\"Non vs W/S			3.7700	0.0624	Yes	14	1.68	0.2164	Yes	Neither



**Figure A- 6: B1Np1Ovr Novice User Pass 1 Overall Impressions Ratings**

**Table A- 6: B1Np1Ovr Novice User Overall Pass 1 Impressions Ratings Statistics**

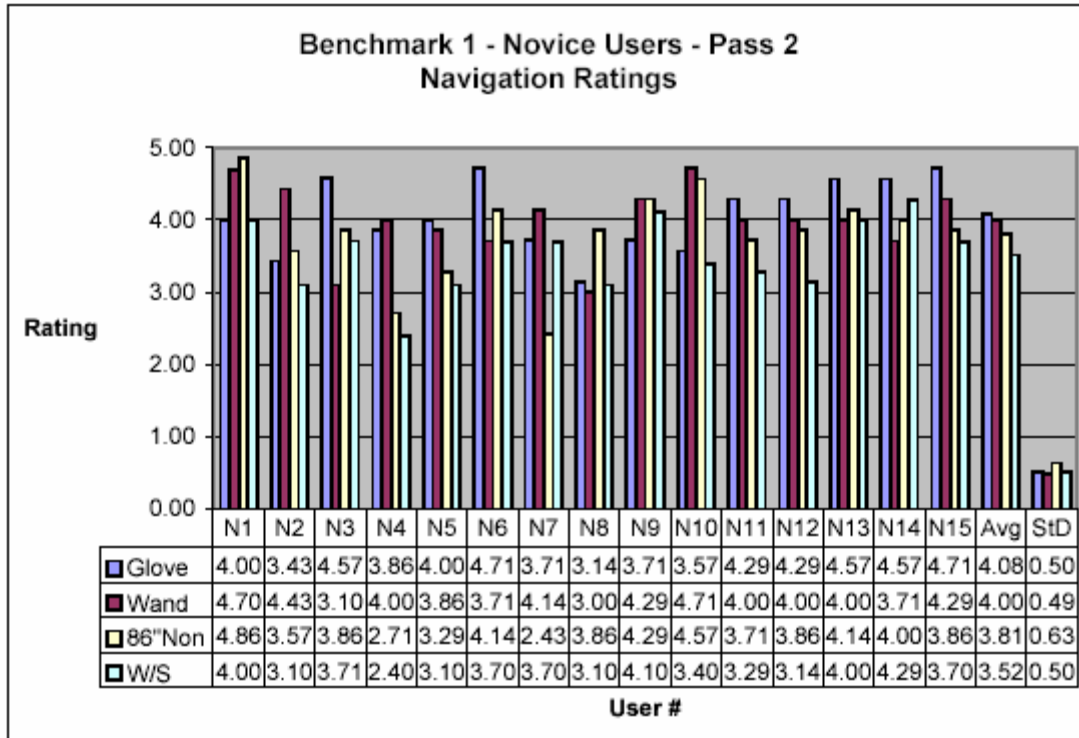
Benchmark 1, Novice Users, Pass 1, Overall Impressions Ratings											
B1Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	4.129	0.384	3.60	4.72	> 0.1000	Yes	9.29%			
Wand	15	3.935	0.530	2.68	4.59	0.0165	No	13.46%			
86\"Non	15	3.815	0.420	2.82	4.41	> 0.1000	Yes	11.01%			
W/S	15	3.597	0.366	2.59	4.09	0.0654	Yes	10.17%			
B1Np1		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	7.75	Pr > F	0.0038
		Bartlett's Test		Levene's Test		df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				1.2000	0.3206	Yes	14	1.56	0.3012	Yes	Neither
Glove vs 86\"Non		0.1110	0.7390			Yes	14	5.14	0.0397	No	Means Glove 8.23%
Glove vs W/S		0.0304	0.8616			Yes	14	21.62	0.0004	No	Means Glove 14.79%
Wand vs 86\"Non				0.5000	0.4853	Yes	14	0.44	0.5172	Yes	Neither
Wand vs W/S				1.0000	0.3261	Yes	14	4.62	0.0496	No	Means Wand 9.40%
86\"Non vs W/S		0.2571	0.6121			Yes	14	5.59	0.0330	No	Means 86\" Non 6.06%



**Figure A- 7: B1Np2Tim Novice User Elapsed Times**

**Table A- 7: B1Np2Tim Novice User Elapsed Times Statistics**

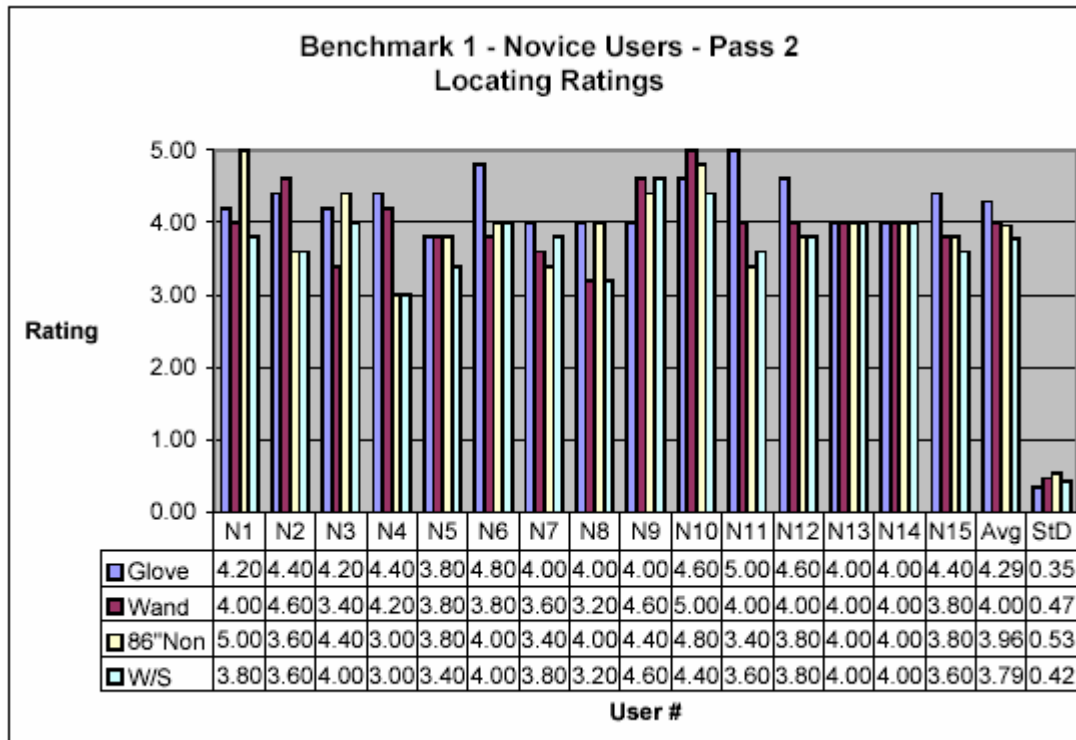
Benchmark 1, Novice Users, Pass 2, Elapsed Time:												
B1Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	15	267.0	60.14	190.0	365.0	> 0.1000	Yes	22.52%				
Wand	15	191.5	47.37	115.0	300.0	0.0929	Yes	24.74%				
86\"Non	15	237.3	70.78	150.0	440.0	> 0.1000	Yes	29.83%				
W/S	15	263.7	72.24	165.0	445.0	> 0.1000	Yes	27.39%				
B1Np2		Homogeneity of Variance				Roy's Greatest Root			F(3,12)		22.96	Pr > F < 0.0001
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		7.6330	0.3823			Yes	14	13.12	0.2800	Yes	Neither	Neither
Glove vs 86\"Nor		0.3571	0.5501			Yes	14	1.52	0.2384	Yes	Neither	Neither
Glove vs W/S		0.4514	0.5017			Yes	14	0.01	0.9090	Yes	Neither	Neither
Wand vs 86\"Non		2.1244	0.1450			Yes	14	6.19	0.0260	No	Means	Wand 23.92%
Wand vs W/S		2.3389	0.1262			Yes	14	35.27	0.0000	No	Means	Wand 37.70%
86\"Non vs W/S		0.0056	0.9404			Yes	14	1.75	0.2070	Yes	Neither	Neither



**Figure A- 8: B1Np2Nav Novice User Navigation Ratings**

**Table A- 8: B1Np2Nav Novice User Navigation Ratings Statistics**

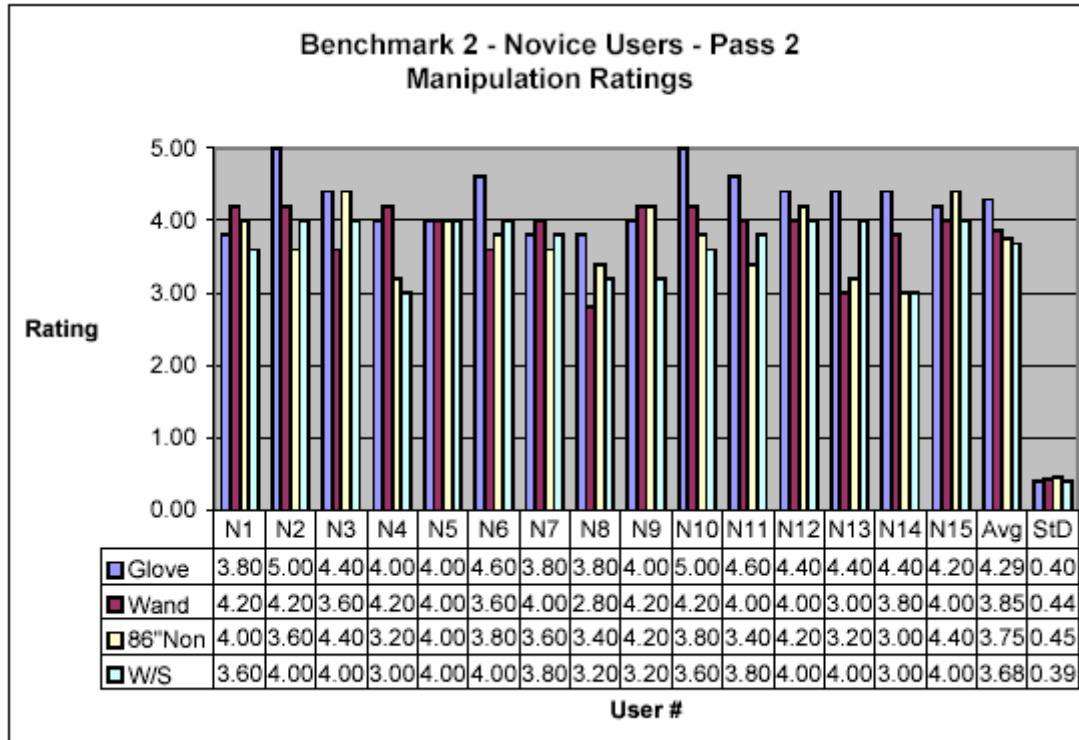
Benchmark 1, Novice Users, Pass 2, Navigation Ratings											
B1Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	4.075	0.500	3.14	4.71	> 0.1000	Yes	12.27%			
Wand	15	3.996	0.490	3.00	4.71	> 0.1000	Yes	12.27%			
86\"Non	15	3.810	0.634	2.43	4.86	> 0.1000	Yes	16.64%			
W/S	15	3.515	0.501	2.41	4.29	> 0.1000	Yes	14.26%			
B1Np2		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)		Pr > F	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand		0.0055	0.9408			Yes	14	0.17	0.6877	Yes	Neither
Glove vs 86\"Non		0.7518	0.3859			Yes	14	1.94	0.1850	Yes	Neither
Glove vs W/S		0.0001	0.9936			Yes	14	16.32	0.0012	No	Means Glove 15.93%
Wand vs 86\"Non		0.8842	0.3471			Yes	14	1.03	0.3280	Yes	Neither
Wand vs W/S		0.0068	0.9344			Yes	14	7.84	0.0142	No	Means Wand 13.68%
86\"Non vs W/S		0.7383	0.3902			Yes	14	4.12	0.0618	No	Means 86\" Non 8.39%



**Figure A- 9: B1Np2Loc Novice User Locating Ratings**

**Table A- 9 B1Np2Loc Novice User Locating Ratings Statistics**

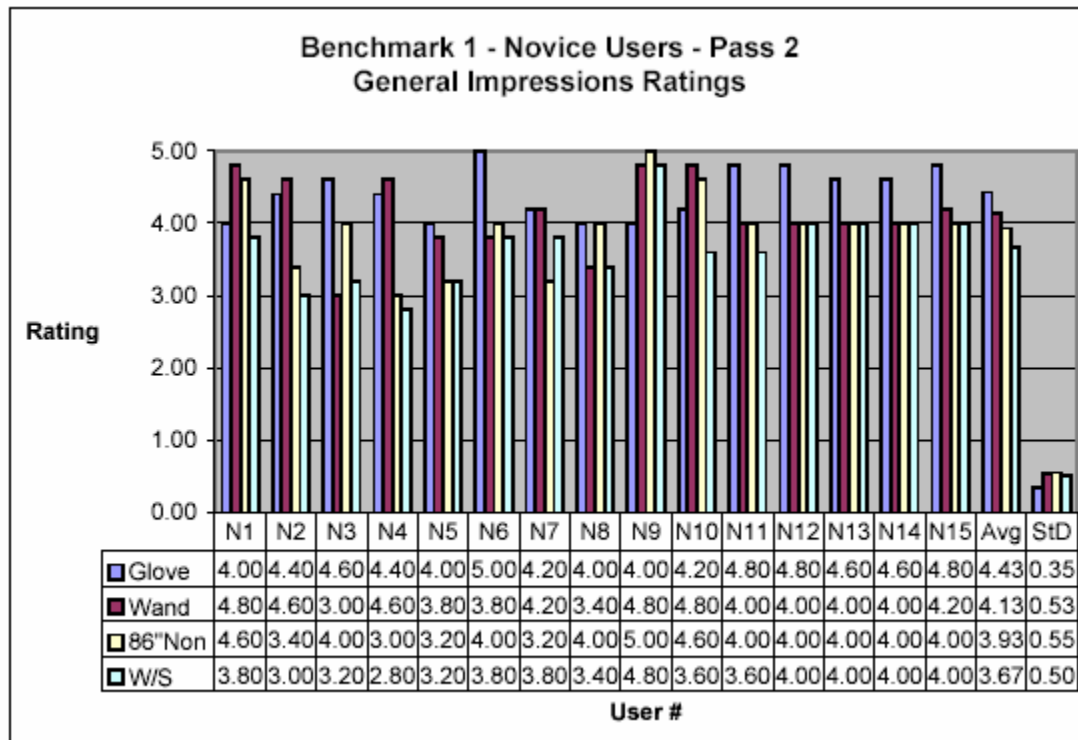
Benchmark 1, Novice Users, Pass 2, Locating Ratings										
B1Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	15	4.293	0.345	3.80	5.00	> 0.1000	Yes	8.04%		
Wand	15	3.996	0.466	3.20	5.00	0.0273	No	11.66%		
86\"Non	15	3.960	0.530	3.00	5.00	0.0961	Yes	13.39%		
W/S	15	3.787	0.417	3.00	4.60	> 0.1000	Yes	11.02%		
Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	4.49	Pr > F	
B1Np2	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand			1.2000	0.2822	Yes	14	5.15	0.0396	No	Means
Glove vs 86\"Non	2.4135	0.1203			Yes	14	3.66	0.0764	No	Means
Glove vs W/S	0.3133	0.5757			Yes	14	13.25	0.0027	No	Means
Wand vs 86\"Non			0.2500	0.6205	Yes	14	0.06	0.8103	Yes	Neither
Wand vs W/S			0.1800	0.6786	Yes	14	3.19	0.0958	No	Means
86\"Non vs W/S	0.7689	0.3806			Yes	14	2.64	0.1265	Yes	Neither



**Figure A- 10: B1Np2Mov Novice User Manipulation Ratings**

**Table A- 10: B1Np2Mov Novice User Manipulation Ratings Statistics**

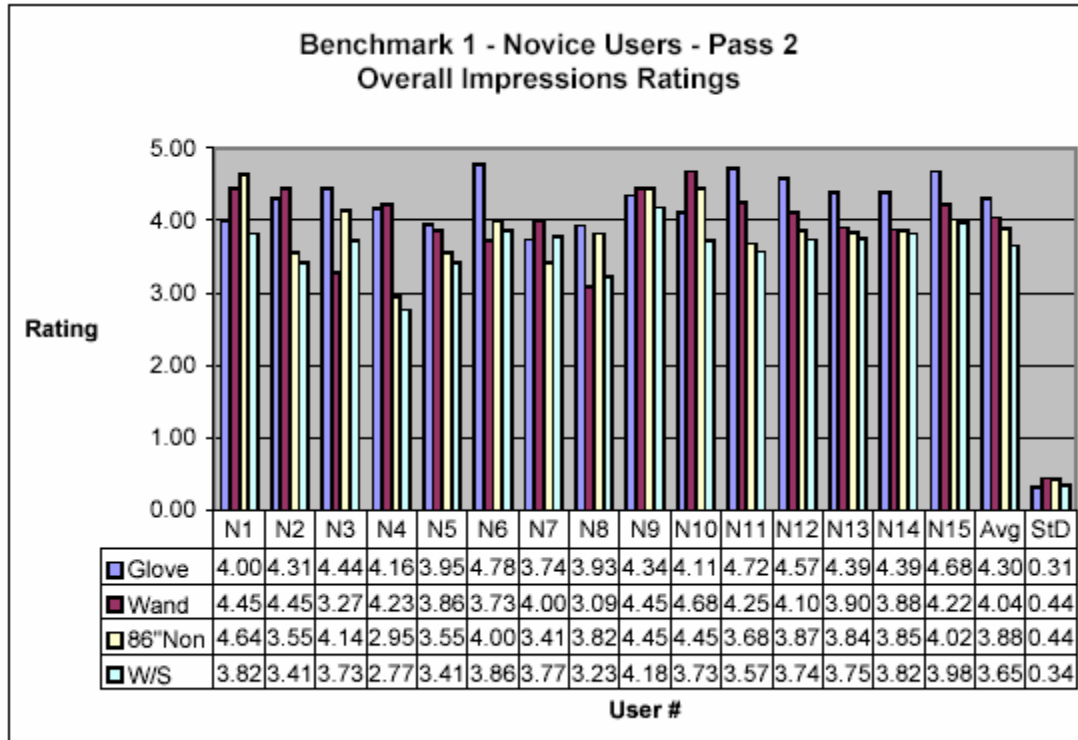
Benchmark 1, Novice Users, Pass 2, Manipulation Ratings												
B1Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	15	4.293	0.399	3.80	5.00	> 0.1000	Yes	9.29%				
Wand	15	3.853	0.437	2.80	4.20	0.0009	No	11.35%				
86\"Non	15	3.747	0.450	3.00	4.40	> 0.1000	Yes	12.01%				
W/S	15	3.680	0.391	3.00	4.00	0.0075	No	10.63%				
B1Np2	Homogeneity of Variance				Roy's Greatest Root: F(3,12)				9.67	Pr > F	0.0016	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.1100	0.7461	Yes	14	9.74	0.0075	No	Means	Glove	11.42%
Glove vs 86\"Non	0.1961	0.6579			Yes	14	11.63	0.0042	No	Means	Glove	14.57%
Glove vs W/S			0.0100	0.9215	Yes	14	27.84	0.0001	No	Means	Glove	16.66%
Wand vs 86\"Non			0.0100	0.9092	Yes	14	0.62	0.4428	Yes	Neither	Neither	
Wand vs W/S			0.1600	0.6929	Yes	14	1.27	0.2778	Yes	Neither	Neither	
86\"Non vs W/S			0.6100	0.4430	Yes	14	0.36	0.5596	Yes	Neither	Neither	



**Figure A- 11: B1Np2Gen Novice User General Impressions Ratings**

**Table A- 11: B1Np2Gen Novice User General Impressions Ratings Statistics**

Benchmark 1, Novice Users, Pass 2, General Impressions Ratings											
B1Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	4.427	0.345	4.00	5.00	> 0.1000	Yes	7.80%			
Wand	15	4.133	0.528	3.00	4.80	> 0.1000	Yes	12.77%			
86\"Non	15	3.933	0.554	3.00	5.00	0.0023	No	14.08%			
W/S	15	3.667	0.500	2.80	4.80	> 0.1000	Yes	13.62%			
B1Np2		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	8.37	Pr > F 0.0028
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant
Glove vs Wand		2.3543	0.1249	3.2100	0.0839	Yes	14	2.51	0.1352	Yes	Neither
Glove vs 86\"Non						Yes	14	7.90	0.0139	No	Means
Glove vs W/S		1.8023	0.1794			Yes	14	24.77	0.0002	No	Means
Wand vs 86\"Non				0.0400	0.8383	Yes	14	1.33	0.2683	Yes	Neither
Wand vs W/S		0.0397	0.8421			Yes	14	7.98	0.0135	No	Means
86\"Non vs W/S				0.1600	0.6880	Yes	14	6.29	0.0251	No	Means

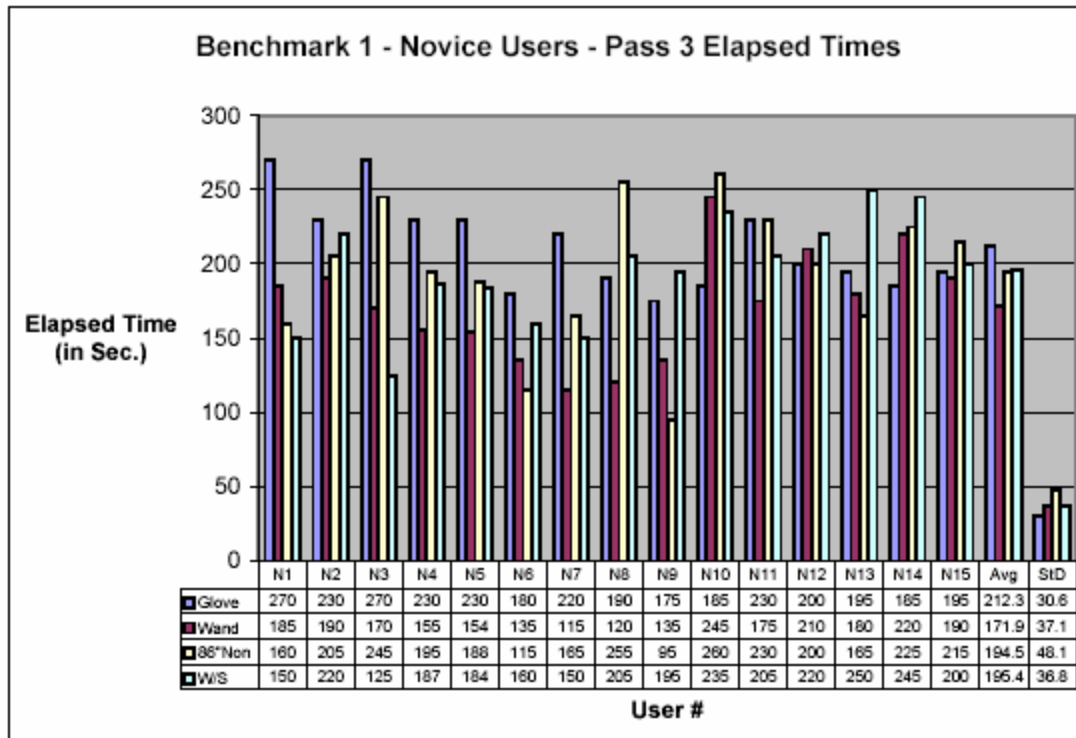


**Figure A- 12: B1Np2Ovr Novice User Overall Impressions Ratings**

**Table A- 12: B1Np2Ovr Novice User Overall Impressions Ratings Statistics**

Benchmark 1, Novice Users, Pass 2, Overall Impressions Ratings												
B1Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	15	4.301	0.314	3.74	4.78	> 0.1000	Yes	7.30%				
Wand	15	4.037	0.439	3.09	4.68	> 0.1000	Yes	10.88%				
86\"Non	15	3.881	0.437	2.95	4.64	> 0.1000	Yes	11.27%				
W/S	15	3.651	0.339	2.77	4.18	0.0081	No	9.29%				
B1Np2		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	16.53	Pr > F	0.0001
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		1.4947	0.2215			Yes	14	3.71	0.0748	No	Means	Glove 6.54%
Glove vs 86\"Non		1.4593	0.2270			Yes	14	10.48	0.0060	No	Means	Glove 10.82%
Glove vs W/S				0.0700	0.7917	Yes	14	44.52	0.0000	No	Means	Glove 17.80%
Wand vs 86\"Non		0.0002	0.9880			Yes	14	1.16	0.2995	Yes	Neither	Neither
Wand vs W/S				0.7400	0.3969	Yes	14	8.86	0.0100	No	Means	Wand 10.57%
86\"Non vs W/S				0.7300	0.3994	Yes	14	8.91	0.0098	No	Means	86\" Non 6.30%

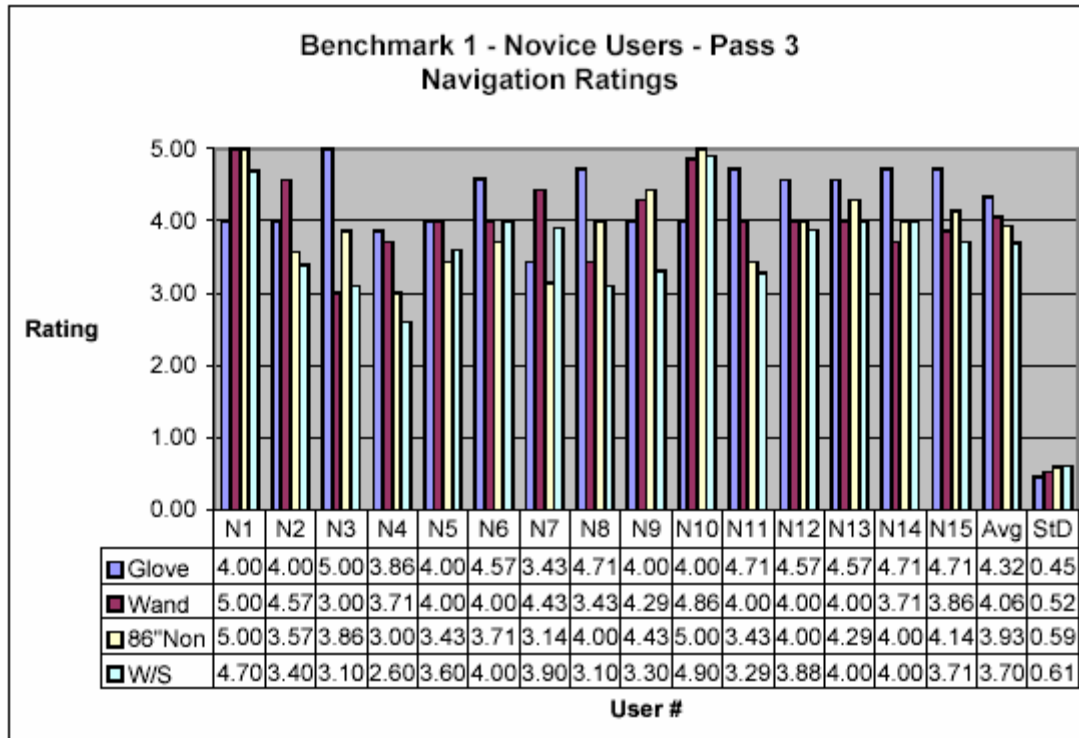




**Figure A- 13: B1Np3Tim Novice User Elapsed Times**

**Table A- 13: B1Np3Tim Novice User Elapsed Times Statistics**

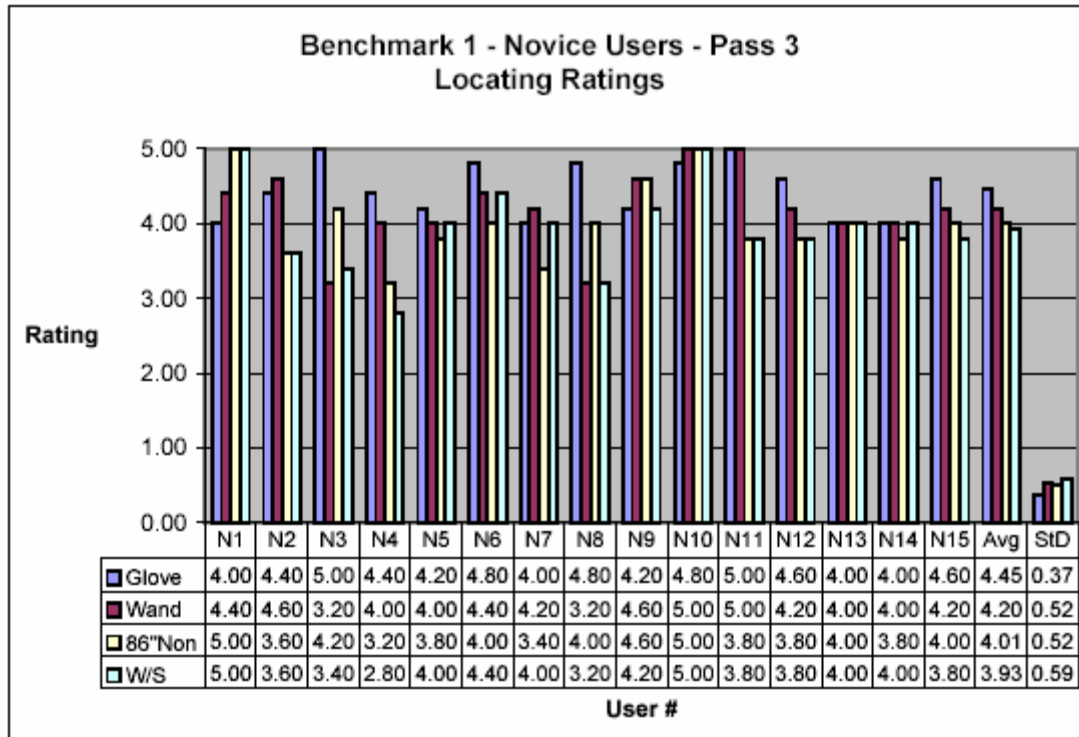
Benchmark 1, Novice Users, Pass 3, Elapsed Time:																	
B1Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.									
Glove	15	212.3	30.64	175.0	270.0	> 0.1000	Yes	14.43%									
Wand	15	171.9	37.11	115.0	245.0	> 0.1000	Yes	21.59%									
86"Non	15	194.5	48.14	95.0	260.0	> 0.1000	Yes	24.75%									
W/S	15	196.4	36.80	125.0	250.0	> 0.1000	Yes	18.74%									
Homogeneity of Variance					Roy's Greatest Root		F(3,12)		4.39		Pr > F 0.0186						
B1Np3		Bartlett's Test		Levene's Test		Equal Var ?		F(1,14)		Pr > F		Equal Means ?		Statistically Significant		Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better						
Glove vs Wand		0.4924	0.4828			Yes	14	10.23	0.0064	No	Means	Wand	23.50%				
Glove vs 86"Nor		2.7107	0.0997			Yes	14	1.75	0.2069	Yes	Neither	Neither					
Glove vs W/S		0.4509	0.5019			Yes	14	1.17	0.2969	Yes	Neither	Neither					
Wand vs 86"Non		0.9305	0.3347			Yes	14	3.76	0.0729	No	Means	Wand	13.15%				
Wand vs W/S		0.0009	0.9757			Yes	14	6.72	0.0213	No	Means	Wand	14.25%				
86"Non vs W/S		0.9893	0.3199			Yes	14	6.72	0.0213	No	Means	86" Non	0.98%				



**Figure A- 14: B1Np3Nav Novice User Navigation Ratings**

**Table A- 14: B1Np3Nav Novice User Navigation Ratings Statistics**

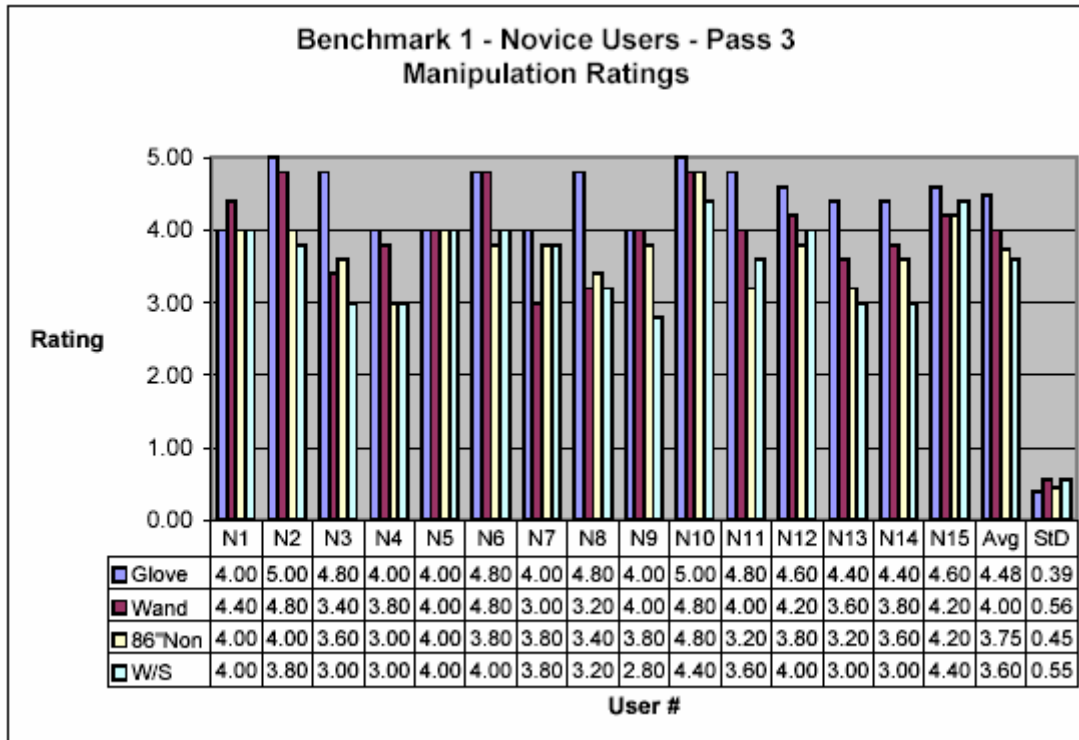
Benchmark 1, Novice Users, Pass 3, Navigation Ratings											
B1Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	4.323	0.445	3.43	5.00	0.0165	No	10.30%			
Wand	15	4.057	0.521	3.00	5.00	0.0725	Yes	12.84%			
86\"Non	15	3.933	0.592	3.00	5.00	> 0.1000	Yes	15.04%			
W/S	15	3.699	0.606	2.60	4.90	> 0.1000	Yes	16.38%			
Homogeneity of Variance				Roy's Greatest Root:				F(3,12)			
Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better	
Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand		0.4200	0.5203	Yes	14	1.35	0.2648	Yes	Neither	Neither	
Glove vs 86\"Non		1.4500	0.2382	Yes	14	4.78	0.0463	No	Means	Glove	9.92%
Glove vs W/S		1.4300	0.2422	Yes	14	8.96	0.0097	No	Means	Glove	16.87%
Wand vs 86\"Non	0.2192	0.6396		Yes	14	0.64	0.4360	Yes	Neither	Neither	
Wand vs W/S	0.3090	0.5783		Yes	14	9.23	0.0089	No	Means	Wand	9.68%
86\"Non vs W/S	0.0078	0.9298		Yes	14	3.71	0.0748	No	Means	Neither	



**Figure A- 15: B1Np3Loc Novice User Locating Ratings**

**Table A- 15: B1Np3Loc Novice User Locating Ratings Statistics**

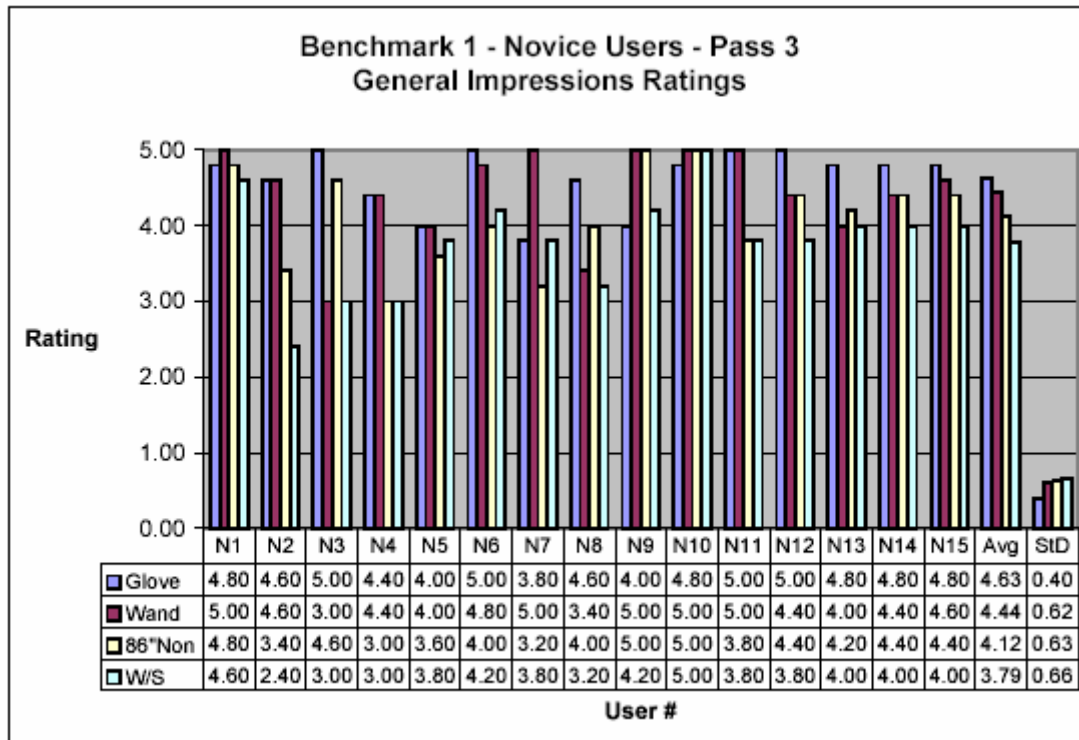
Benchmark 1, Novice Users, Pass 3, Locating Ratings										
B1Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	15	4.453	0.374	4.00	5.00	> 0.1000	Yes	8.40%		
Wand	15	4.200	0.524	3.20	5.00	0.5350	Yes	12.47%		
86\"Non	15	4.013	0.515	3.20	5.00	0.1690	Yes	12.84%		
W/S	15	3.933	0.589	2.80	5.00	> 0.1000	Yes	14.97%		
B1Np3	Homogeneity of Variance				Roy's Greatest Root: F(3,12)			2.58	Pr > F	0.1313
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						Statistically Better
Glove vs Wand	1.5064	0.2197			Yes	14	2.26	0.1548	Yes	Neither
Glove vs 86\"Non			1.5600	0.2222	Yes	14	7.68	0.0150	No	Means
Glove vs W/S	2.6958	0.1006			Yes	14	6.90	0.0199	No	Means
Wand vs 86\"Non			0.0000	0.9507	Yes	14	1.30	0.2737	Yes	Neither
Wand vs W/S	0.1549	0.6672			Yes	14	4.00	0.0653	No	Means
86\"Non vs W/S			0.3931	0.5307	Yes	14	0.63	0.4417	Yes	Neither



**Figure A- 16: B1Np3Mov Novice User Manipulation Ratings**

**Table A- 16: B1Np3Mov Novice User Manipulation Ratings Statistics**

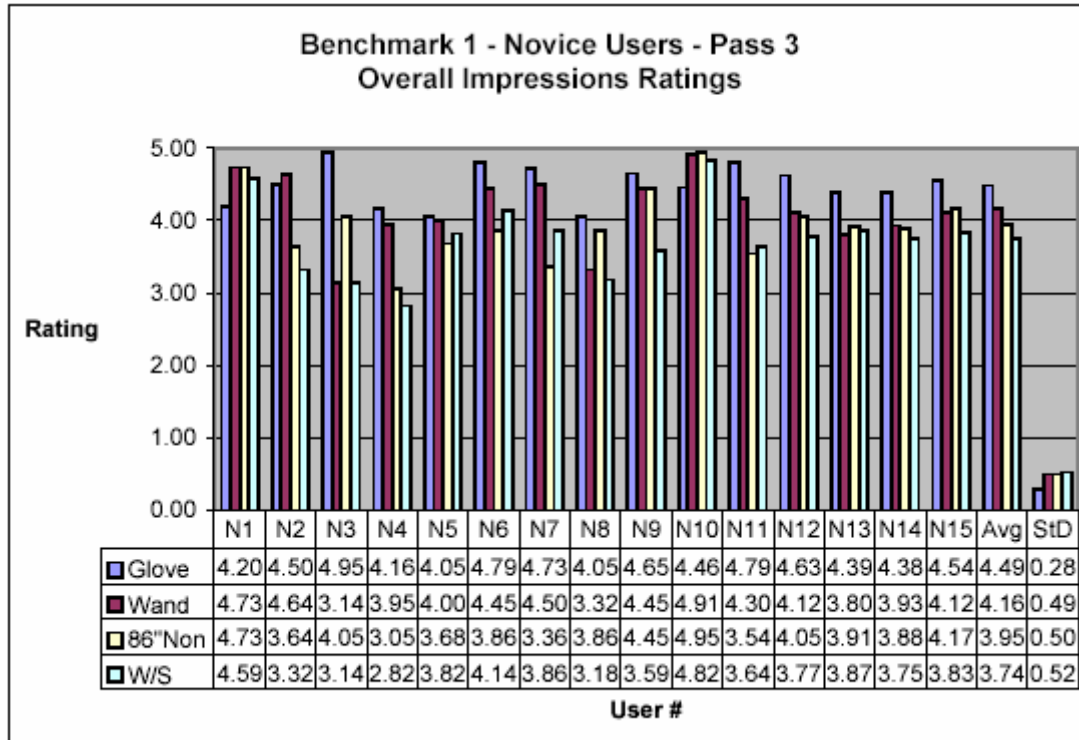
Benchmark 1, Novice Users, Pass 3, Manipulation Ratings														
B1Np3	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	Coeff. of Var.		
Glove	15	4.480		0.391		4.00		5.00		0.0426	No	8.73%		
Wand	15	4.000		0.561		3.00		4.80		> 0.1000	Yes	14.02%		
86"Non	15	3.747		0.450		3.00		4.80		> 0.1000	Yes	12.01%		
W/S	15	3.600		0.550		2.80		4.40		> 0.1000	Yes	15.29%		
B1Np3		Homogeneity of Variance					Roy's Greatest Root:			F(3,12)		10.19	Pr > F	0.0013
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better		
		Chi-Sq	Pr > ChiSq	F Value	Pr > F									
Glove vs Wand				2.8200	0.1040	Yes	14	11.24	0.0047	No	Means	Glove	12.00%	
Glove vs 86"Non				0.3200	0.5760	Yes	14	29.01	0.0000	No	Means	Glove	19.56%	
Glove vs W/S				4.9500	0.0342	Yes	14	33.43	0.0000	No	Means	Glove	24.44%	
Wand vs 86"Non		0.6453	0.4218			Yes	14	4.16	0.0606	No	Means	Wand	6.75%	
Wand vs W/S		0.0046	0.9457			Yes	14	9.13	0.0092	No	Means	Wand	11.11%	
86"Non vs W/S		0.5417	0.4617			Yes	14	2.31	0.1510	Yes	Neither	Neither		



**Figure A- 17: B1Np3Gen Novice User General Impressions Ratings**

**Table A- 17: B1Np3Gen Novice User General Impressions Ratings**

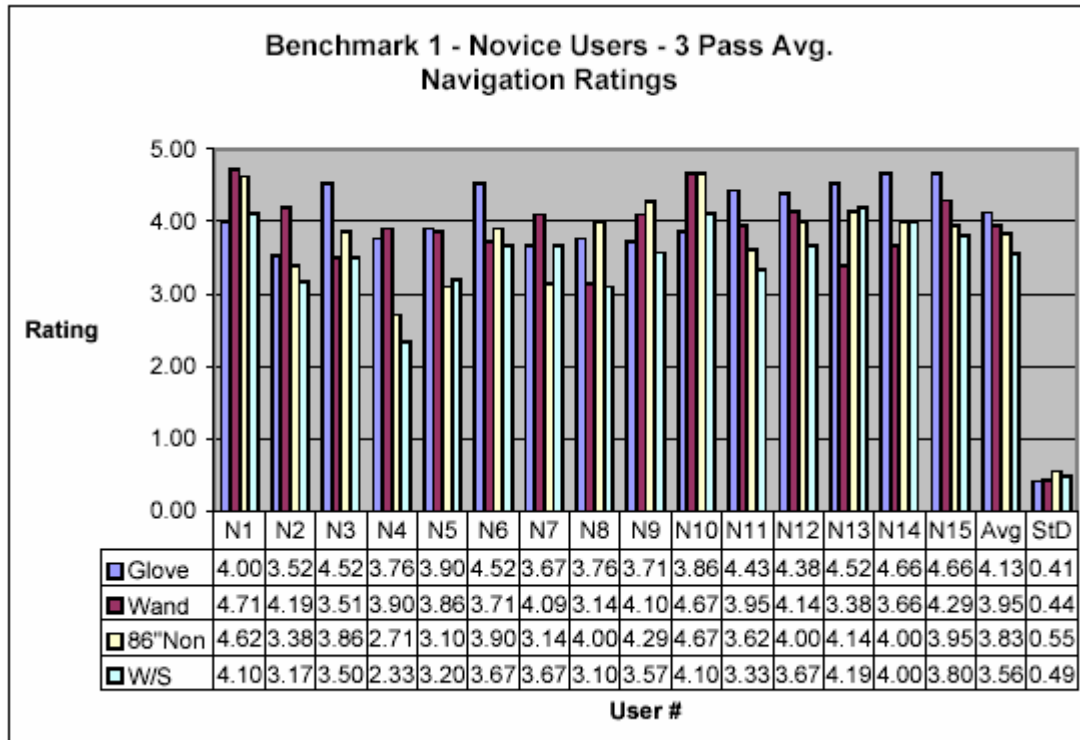
Benchmark 1, Novice Users, Pass 3, General Impressions Ratings												
B1Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	15	4.627	0.399	3.80	5.00	0.0049	No	8.62%				
Wand	15	4.440	0.615	3.00	5.00	0.0819	Yes	13.85%				
86\"Non	15	4.120	0.627	3.00	5.00	> 0.1000	Yes	15.22%				
W/S	15	3.787	0.661	2.40	5.00	0.0188	No	17.45%				
B1Np3	Homogeneity of Variance					Roy's Greatest Root:			F(3,12)		7.69	Pr > F 0.004
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			1.8600	0.1832	Yes	14	0.84	0.3740	Yes	Neither	Neither	
Glove vs 86\"Non	2.6711	0.1022			Yes	14	10.55	0.0580	No	Means	Glove	12.31%
Glove vs W/S			2.6700	0.1135	Yes	14	19.06	0.0006	No	Means	Glove	22.18%
Wand vs 86\"Non			0.0100	0.9367	Yes	14	2.07	0.1721	Yes	Neither	Neither	
Wand vs W/S			0.0700	0.7927	Yes	14	16.61	0.0011	No	Means	Wand	17.24%
86\"Non vs W/S			0.0500	0.8227	Yes	14	5.34	0.0367	No	Means	86\" Non	8.79%



**Figure A- 18: B1Np3Ovr Novice User Overall Impressions Ratings**

**Table A- 18: B1Np3Ovr Novice User Overall Impressions Ratings Statistics**

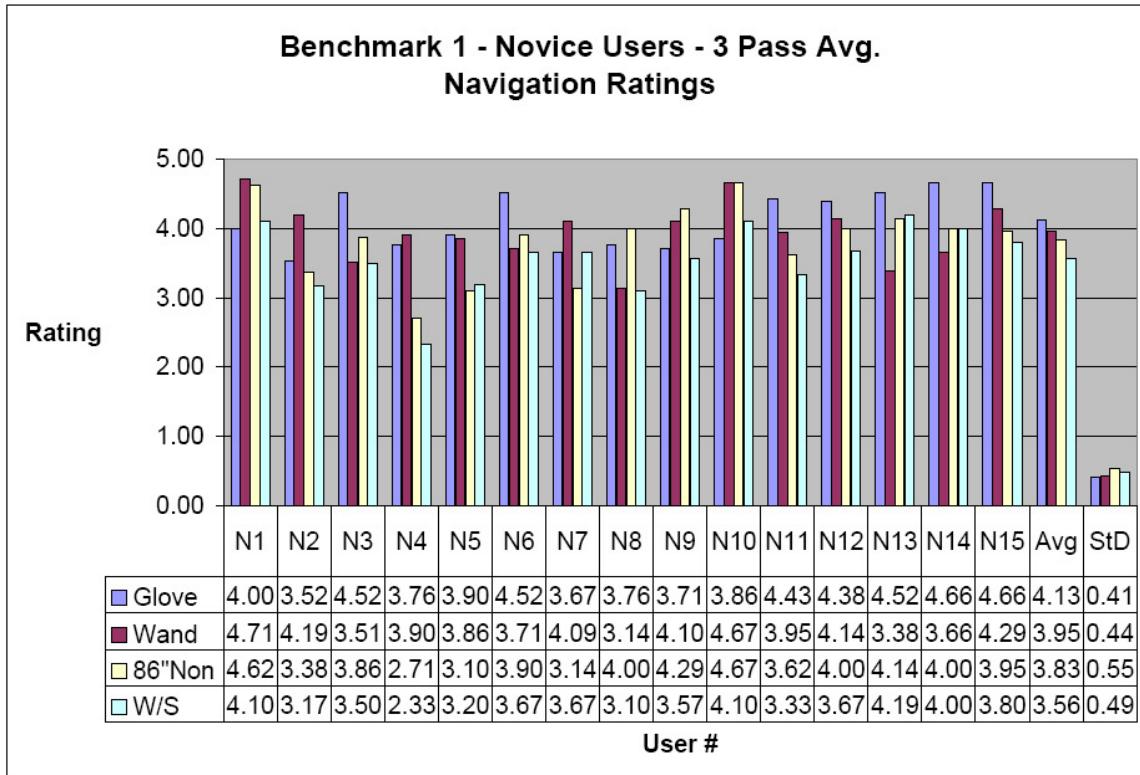
Benchmark 1, Novice Users, Pass 3, Overall Impressions Ratings												
B1Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	15	4.485	0.280	4.05	4.95	> 0.1000	Yes	6.25%				
Wand	15	4.157	0.494	3.14	4.91	> 0.1000	Yes	11.89%				
86\"Non	15	3.945	0.496	3.05	4.95	> 0.1000	Yes	12.57%				
W/S	15	3.683	0.557	2.82	4.82	> 0.1000	Yes	15.11%				
B1Np3	Homogeneity of Variance				Roy's Greatest Root:			F(3,12)		9.09		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Pr > F	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							Statistically Better	
Glove vs Wand	4.1398	0.0419			No	14	5.37	0.0361	No	Var/Mean	Glove	7.89%
Glove vs 86\"Non	6.2003	0.0450			No	14	13.85	0.0023	No	Var/Mean	Glove	13.69%
Glove vs W/S	5.9180	0.0150			No	14	26.04	0.0002	No	Var/Mean	Glove	21.78%
Wand vs 86\"Non	0.0000	0.9907			Yes	14	2.05	0.1743	Yes	Neither	Wand	5.37%
Wand vs W/S	0.1891	0.6636			Yes	14	20.33	0.0213	No	Means	Wand	12.87%
86\"Non vs W/S	0.1791	0.6721			Yes	14	5.49	0.0344	No	Means	86\" Non	7.11%



**Figure A- 19: B1N3pAvgTim Novice Users Elapsed Times**

**Table A- 19: B1N3pAvgTim Novice Users Elapsed Times Statistics**

Benchmark 1, Novice Users, 3 Pass Avg., Elapsed Time:										
B1N3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	15	274.5	44.01	213.3	348.3	> 0.1000	Yes	16.03%		
Wand	15	221.8	69.17	153.3	403.3	> 0.1000	Yes	31.19%		
86\"Non	15	252.6	69.45	167.7	436.7	0.0135	No	27.49%		
W/S	15	268.6	54.36	191.7	376.7	> 0.1000	Yes	20.24%		
Homogeneity of Variance				Roy's Greatest Root		F(3,12)	6.57	Pr > F		
B1N3pA		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					
Glove vs Wand		2.6742	0.1020			Yes	14	8.30	0.0120	No
Glove vs 86\"Nor				1.4300	0.2413	Yes	14	1.23	0.2860	Yes
Glove vs W/S		0.5987	0.4391			Yes	14	0.10	0.7610	Yes
Wand vs 86\"Non				0.0001	0.9909	Yes	14	1.32	0.2690	Yes
Wand vs W/S		0.7771	0.3780			Yes	14	11.81	0.0040	No
86\"Non vs W/S				0.5400	0.4683	Yes	14	0.51	0.4900	Yes
										Statistically Significant
										Statistically Better
										Wand 23.76%
										Neither
										Neither
										Neither
										Wand 21.10%
										Neither

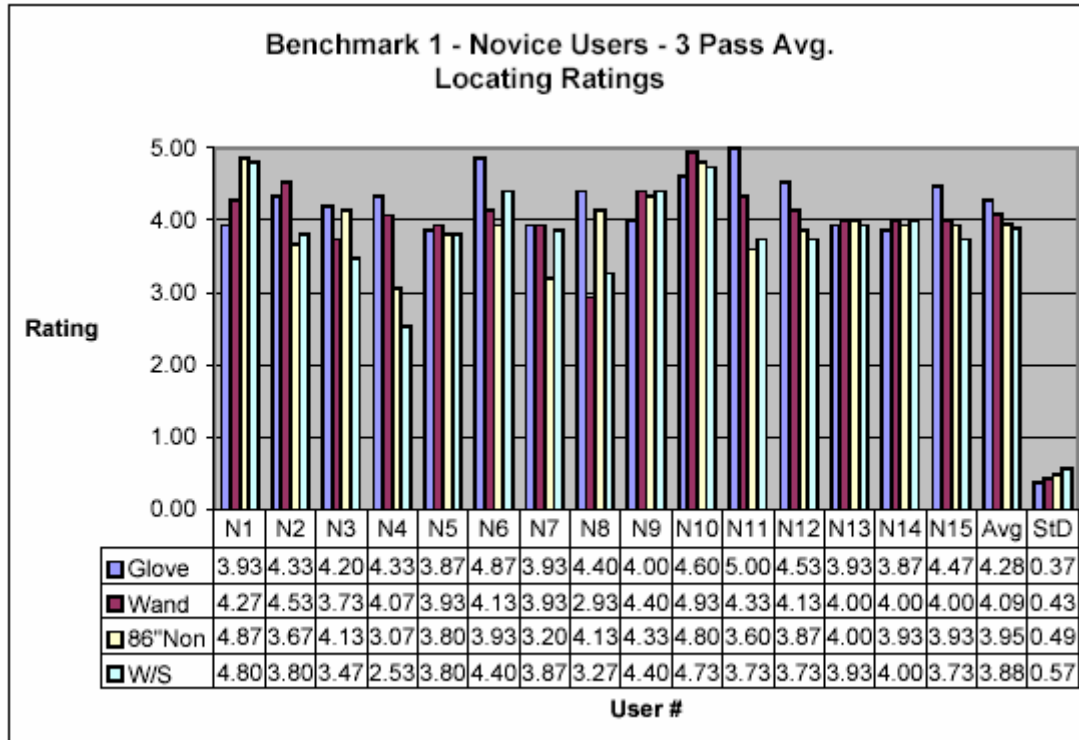


**Figure A- 20: B1N3pAvgNav Novice Users Navigation Ratings**

**Table A- 20: B1N3pAvgNav Novice Users Navigation Ratings Statistics**

Benchmark 1, Novice Users, 3 Pass Avg., Navigation Ratings											
B1N3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	4.125	0.409	3.52	4.71	> 0.1000	Yes	9.92%			
Wand	15	3.953	0.437	3.14	4.71	> 0.1000	Yes	11.05%			
86"Non	15	3.825	0.551	2.71	4.67	> 0.1000	Yes	14.41%			
W/S	15	3.560	0.488	2.33	4.19	> 0.1000	Yes	13.70%			
B1N3pA	Homogeneity of Variance					Roy's Greatest Root:			F(3,12) Pr > F 0.1313		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.0576	0.8104			Yes	14	0.98	0.3395	Yes	Neither	Neither
Glove vs 86"Non	1.1784	0.2777			Yes	14	3.89	0.0687	No	Means	Glove 7.84%
Glove vs W/S	0.4120	0.5210			Yes	14	21.20	0.0004	No	Means	Glove 15.87%
Wand vs 86"Non	0.7218	0.3955			Yes	14	0.66	0.4290	Yes	Neither	Neither
Wand vs W/S	0.1624	0.6869			Yes	14	7.39	0.0166	No	Means	Wand 11.04%
86"Non vs W/S	0.2019	0.6532			Yes	14	8.46	0.0114	No	Means	86" Non 7.44%

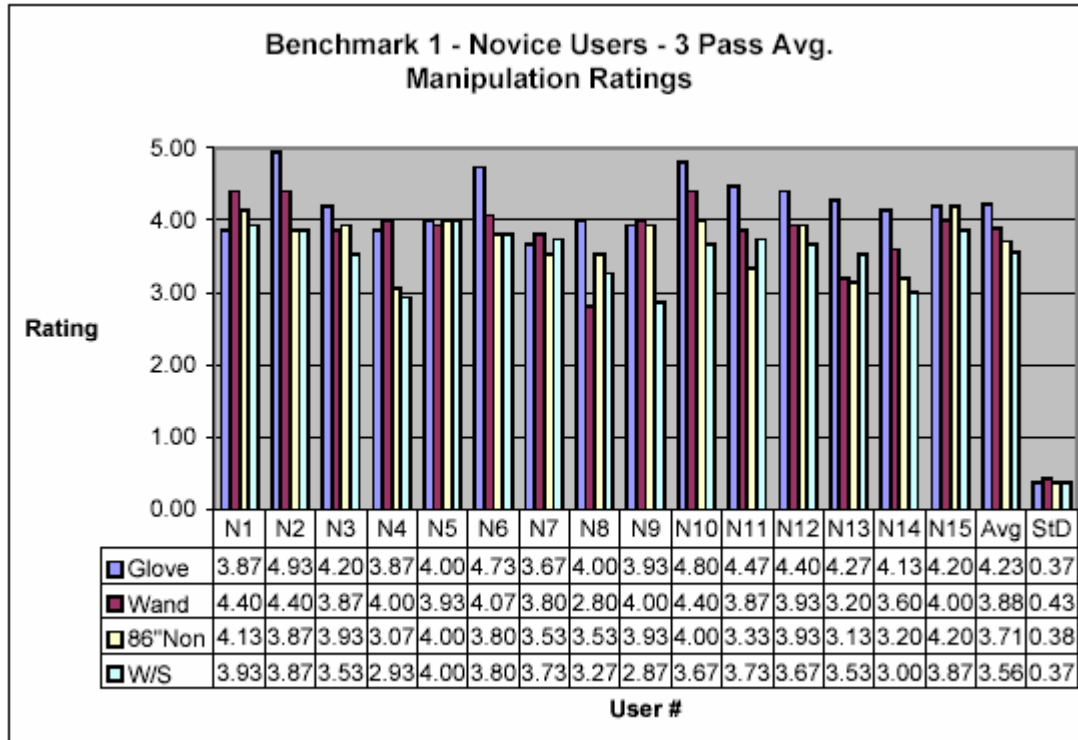




**Figure A- 21: B1N3pAvg Novice User Locating Ratings**

**Table A- 21: B1N3pAvg Novice User Locating Ratings Statistics**

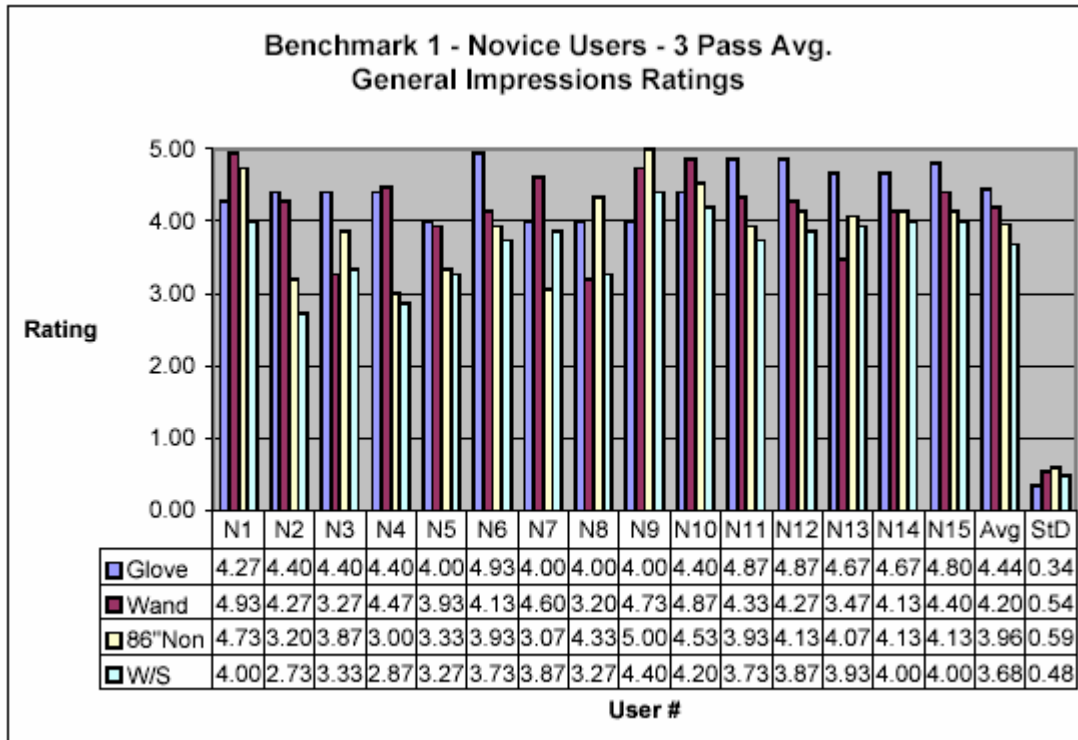
Benchmark 1, Novice Users, 3 Pass Avg., Locating Ratings											
B1N3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	4.284	0.366	3.87	5.00	> 0.1000	Yes	8.54%			
Wand	15	4.087	0.434	2.93	4.93	0.0392	No	10.62%			
86\"Non	15	3.951	0.489	3.07	4.87	> 0.1000	Yes	12.38%			
W/S	15	3.879	0.571	2.53	4.80	> 0.1000	Yes	14.73%			
B1N3pA		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	1.47	Pr > F	0.0081
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant
		Chi-Sq	Pr > ChiSq	F value	Pr > F						
Glove vs Wand				0.2500	0.6207	Yes	14	2.20	0.1604	Yes	Neither
Glove vs 86\"Non		1.1205	0.2898			Yes	14	4.18	0.0602	No	Means
Glove vs W/S		2.5971	0.1071			Yes	14	5.02	0.0419	No	Means
Wand vs 86\"Non				0.1400	0.7113	Yes	14	0.85	0.3735	Yes	Neither
Wand vs W/S				0.6500	0.4284	Yes	14	2.63	0.1270	Yes	Neither
86\"Non vs W/S		0.3254	0.5684			Yes	14	0.49	0.4951	Yes	Neither



**Figure A- 22: B1N3pAvgMov Novice User Manipulation Ratings**

**Table A- 22: B1N3pAvgMov Novice User Manipulation Ratings Statistics**

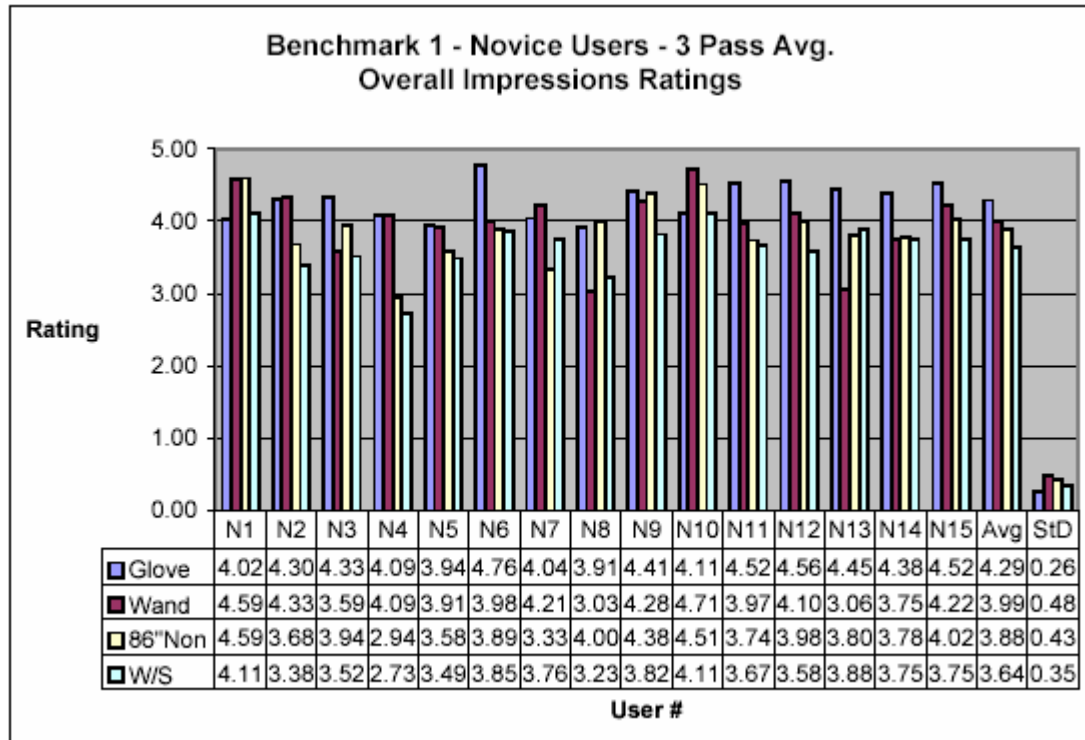
Benchmark 1, Novice Users, 3 Pass Avg., Manipulation Ratings											
B1N3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	4.231	0.372	3.67	4.93	> 0.1000	Yes	8.78%			
Wand	15	3.885	0.432	2.80	4.40	0.0446	No	11.11%			
86"Non	15	3.705	0.376	3.07	4.20	0.0984	Yes	10.15%			
W/S	15	3.560	0.372	2.87	4.00	> 0.1000	Yes	10.44%			
B1N3pA	Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	10.19	Pr > F	0.0013
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.2500	0.6182	Yes	14	8.43	0.0116	No	Means	Glove 8.91%
Glove vs 86"Non	0.0018	0.9661			Yes	14	29.01	0.0000	No	Means	Glove 14.20%
Glove vs W/S	0.0000	0.9989			Yes	14	37.63	0.0000	No	Means	Glove 18.85%
Wand vs 86"Non			0.2300	0.6323	Yes	14	4.16	0.0606	No	Means	Wand 4.86%
Wand vs W/S			0.1500	0.7021	Yes	14	5.38	0.0359	No	Means	Wand 9.13%
86"Non vs W/S	0.0017	0.9672			Yes	14	2.49	0.1367	Yes	Neither	Neither



**Figure A- 23: B1N3pAvgGen Novice User General Impressions Ratings**

**Table A- 23: B1N3pAvgGen Novice User General Impressions Ratings Statistics**

Benchmark 1, Novice Users, 3 Pass Avg., General Impressions Ratings											
B1N3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	4.445	0.344	4.00	4.93	> 0.1000	Yes	7.73%			
Wand	15	4.200	0.537	3.20	4.93	> 0.1000	Yes	12.80%			
86\"Non	15	3.959	0.594	3.00	5.00	> 0.1000	No	14.99%			
W/S	15	3.680	0.482	2.73	4.40	0.0801	Yes	13.10%			
Homogeneity of Variance				Roy's Greatest Root:				F(3,12)	11.17	Pr > F	0.0007
Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better	
Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand	2.6120	0.1061		Yes	14	2.19	0.1611	Yes	Neither	Neither	
Glove vs 86\"Non	3.8470	0.0498		No	14	7.88	0.0140	No	Var/Means	Glove	12.28%
Glove vs W/S	1.5159	0.2182		Yes	14	29.15	0.0000	No	Means	Glove	20.79%
Wand vs 86\"Non	0.1337	0.7147		Yes	14	1.60	0.2267	Yes	Neither	Wand	6.09%
Wand vs W/S	0.1593	0.6898		Yes	14	13.07	0.0028	No	Means	Wand	14.13%
86\"Non vs W/S	0.5818	0.4456		No	14	6.99	0.0193	No	Var/Means	86\" Non	7.58%



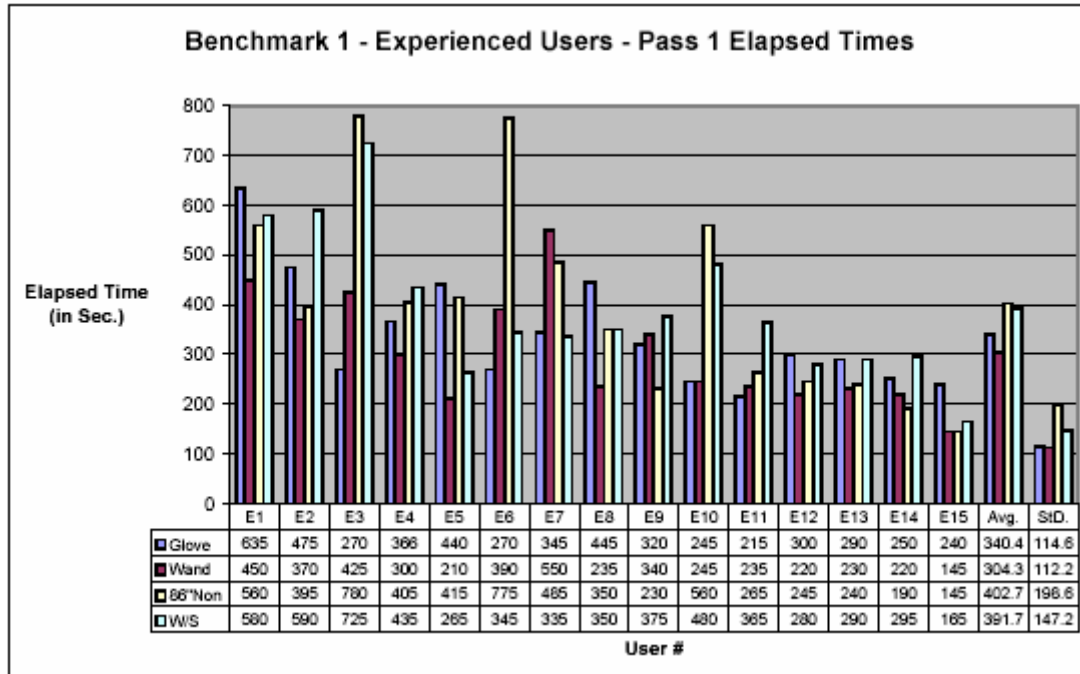
**Figure A- 24: B1N3pAvgOvr Novice User Overall Impressions Ratings**

**Table A- 24: B1N3pAvgOvr Novice User Overall Impressions Ratings Statistics**

Benchmark 1, Novice Users, 3 Pass Avg., Overall Impressions Ratings											
B1N3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	4.29	0.26	3.91	4.76	> 0.1000	Yes	5.99%			
Wand	15	3.99	0.48	3.03	4.71	> 0.1000	Yes	12.01%			
86\"Non	15	3.88	0.43	2.94	4.59	> 0.1000	Yes	11.01%			
W/S	15	3.64	0.35	2.73	4.11	> 0.1000	Yes	9.62%			
B1N3pA	Homogeneity of Variance				Roy's Greatest Root			F(3,12)		15.45	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Pr > F
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	4.9383	0.0263			No	14	4.50	0.0521	No	Var/Mean	Glove 7.55%
Glove vs 86\"Non	3.3435	0.0675			Yes	14	11.23	0.0048	No	Means	Glove 10.63%
Glove vs W/S	1.2830	0.2573			Yes	14	44.18	0.0000	No	Means	Glove 17.76%
Wand vs 86\"Non	0.1794	0.6719			Yes	14	0.61	0.4481	Yes	Neither	Wand 2.86%
Wand vs W/S	1.2988	0.2544			Yes	14	7.28	0.0173	No	Means	Wand 9.50%
86\"Non vs W/S	0.5219	0.4700			Yes	14	9.56	0.0080	No	Means	86\" Non 6.45%

## **Appendix B**

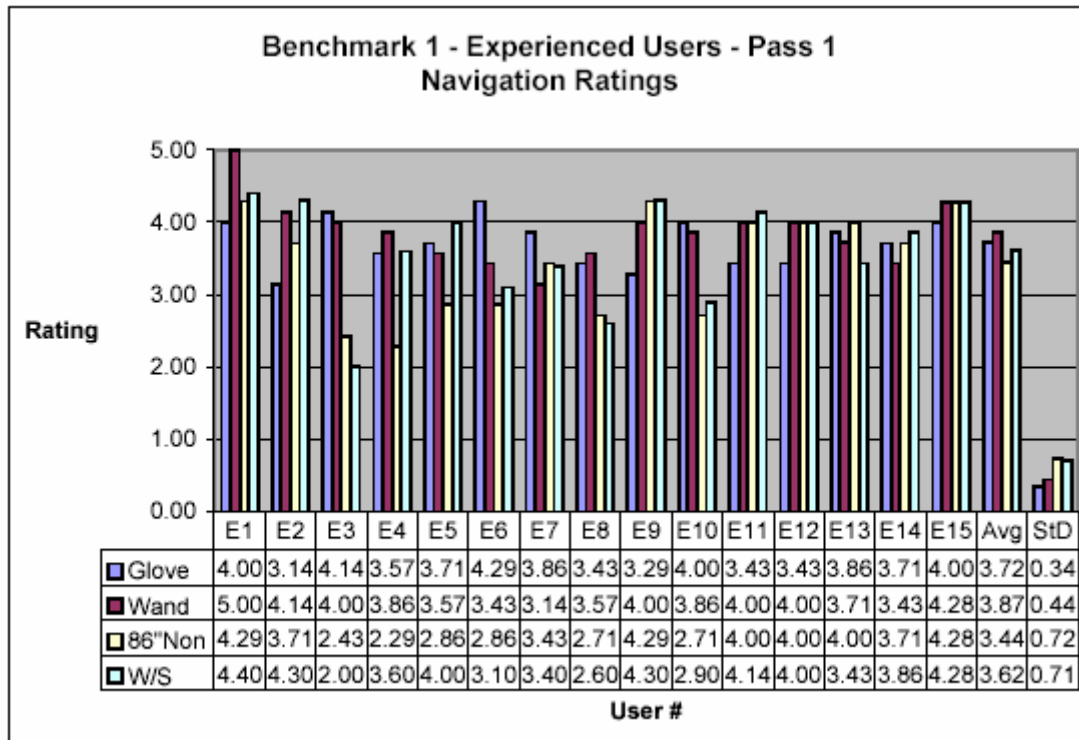
### Experienced User Benchmark 1 (Navigation) Detail



**Figure B- 1: B1Ep1Tim Experienced User Elapsed Times**

**Table B- 1: B1Ep1Tim Experienced User Elapsed Times Statistics**

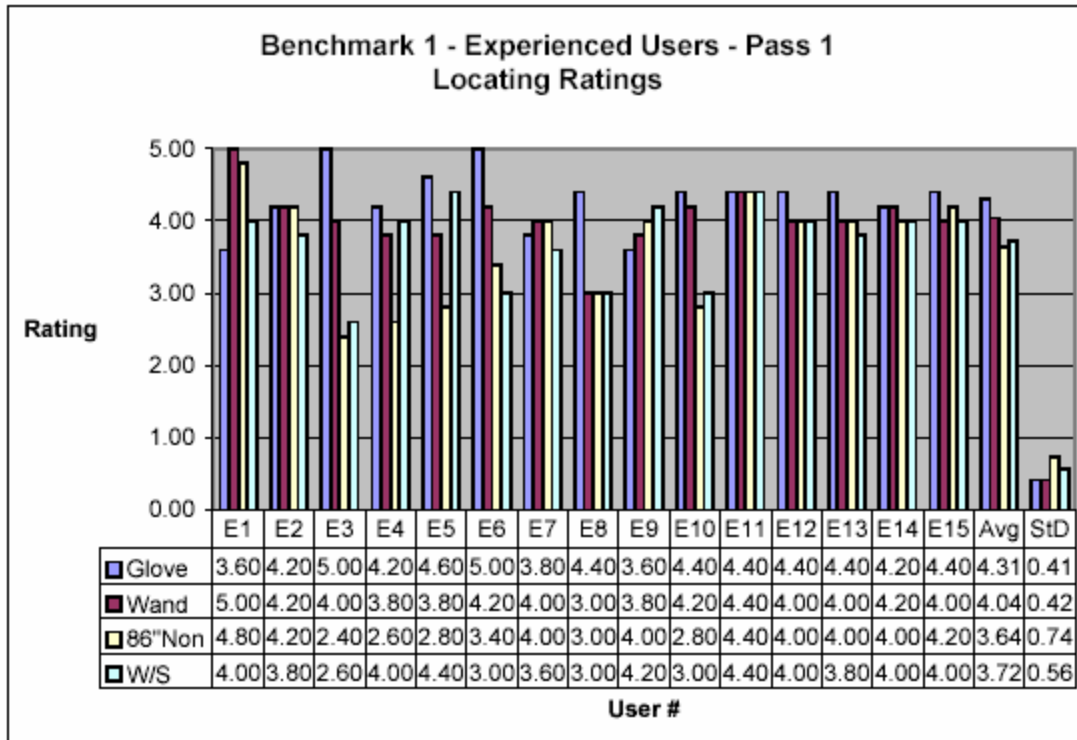
Benchmark 1, Experienced Users, Pass 1, Elapsed Times											
B1Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	340.40	114.60	215.00	635.00	> 0.1000	Yes	33.67%			
Wand	15	304.30	112.20	145.00	550.00	0.0255	No	36.87%			
86*Non	15	402.30	196.60	145.00	780.00	> 0.1000	Yes	49.37%			
W/S	15	391.70	147.20	165.00	725.00	0.069	No	37.58%			
B1Ep1	Homogeneity of Variance				Roy's Greatest Root:				F(3,12)		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Pr > F
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.0100	0.9394	Yes	14	1.20	0.2924	Yes	Neither	Neither
Glove vs 86*Non	3.8877	0.0486			No	14	1.30	0.2739	Yes	Var	Glove 18.18%
Glove vs W/S			0.7100	0.4070	Yes	14	1.71	0.2121	Yes	Neither	Neither
Wand vs 86*Non			4.0300	0.0544	Yes	14	6.22	0.0257	No	Means	Wand 32.21%
Wand vs W/S			0.9700	0.3338	Yes	14	7.62	0.1984	Yes	Neither	Neither
86*Non vs W/S			1.3800	0.2499	Yes	14	0.08	0.7832	Yes	Neither	Neither



**Figure B- 2: B1Ep1Nav Experienced User Navigation Ratings**

**Table B- 2: B1Ep1Nav Experienced User Navigation Ratings Statistics**

Benchmark 1, Experienced Users, Pass 1, Navigation Ratings												
B1Ep1	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	Coeff. of Var.
Glove	15	3.724		0.335		3.14		4.29		> 0.1000	Yes	9.00%
Wand	15	3.866		0.220		3.14		5.00		> 0.1000	Yes	5.69%
86"Non	15	3.438		0.723		2.29		4.29		> 0.1000	Yes	21.03%
W/S	15	3.621		0.713		2.00		4.40		> 0.1000	Yes	19.70%
B1Ep1		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	2.13	Pr > F	0.1502
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
		Chi-Sq	Pr > ChiSq	F value	Pr > F							
Glove vs Wand		0.9922	0.3192			Yes	14	0.95	0.3459	Yes	Neither	
Glove vs 86"Non		7.3130	0.0068			No	14	1.62	0.2243	Yes	Var	Glove 8.32%
Glove vs W/S		7.0792	0.0078			No	14	0.19	0.6678	Yes	Var	Glove 2.84%
Wand vs 86"Non		3.2038	0.0735			Yes	14	6.31	0.0248	No	Means	Wand 12.45%
Wand vs W/S		3.0386	0.0813			Yes	14	2.06	0.1735	Yes	Neither	Neither
86"Non vs W/S		0.0025	0.9605			Yes	14	1.98	0.1809	Yes	Neither	W/S 5.32%

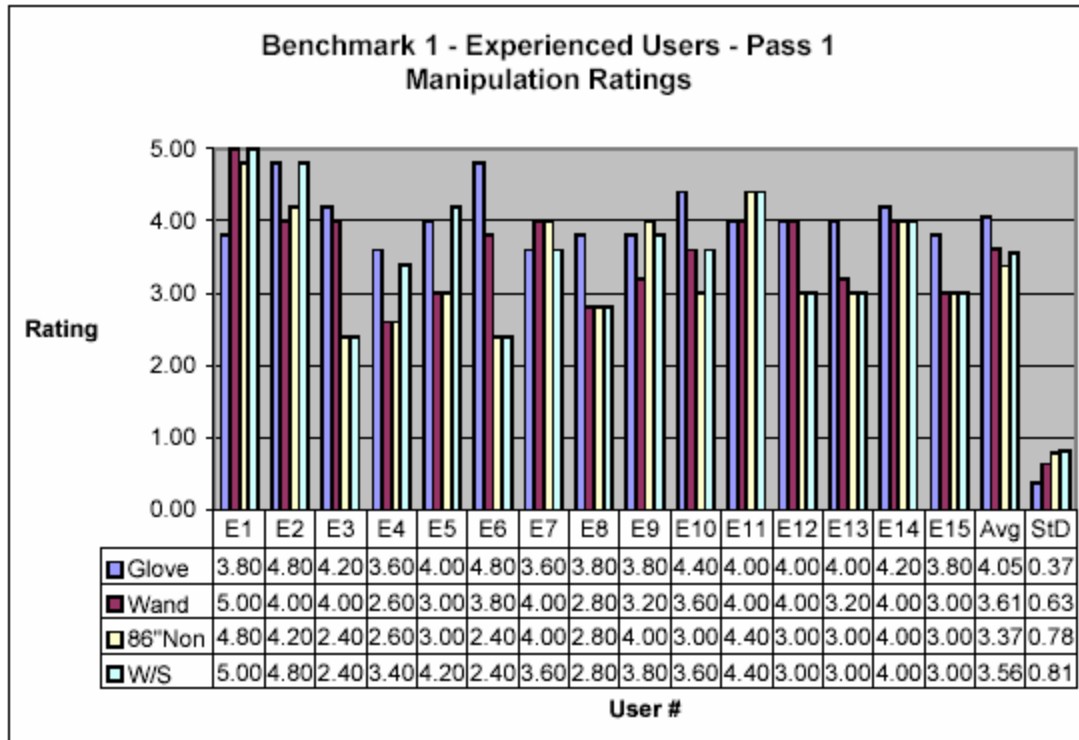


**Figure B- 3: B1Ep1Loc Experienced User Locating Ratings**

**Table B- 3: B1Ep1Loc Experienced User Locating Ratings Statistics**

Benchmark 1, Experienced Users, Pass 1, Locating Ratings												
B1Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	15	4.307	0.413	3.00	5.00	0.0721	Yes	9.59%				
Wand	15	4.040	0.415	2.40	5.00	0.0563	Yes	10.28%				
86"Non	15	3.640	0.741	2.60	4.80	0.0017	No	20.37%				
W/S	15	3.720	0.560	2.00	4.40	0.0398	No	15.04%				
B1Ep1		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	4.45	Pr > F	0.1143
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		0.0005	0.9837			Yes	14	2.53	0.1337	Yes	Neither	Neither
Glove vs 86"Non				6.8600	0.0141	No	14	6.22	0.0258	No	Var/Mean	Glove 18.32%
Glove vs W/S				1.6400	0.2109	Yes	14	7.20	0.0178	No	Means	Glove 15.78%
Wand vs 86"Non				5.3800	0.0279	No	14	6.27	0.0253	No	Var/Mean	Wand 10.99%
Wand vs W/S				1.1100	0.3020	Yes	14	4.09	0.0628	No	Means	Wand 8.60%
86"Non vs W/S				2.1400	0.1548	Yes	14	0.23	0.6361	Yes	Neither	Neither





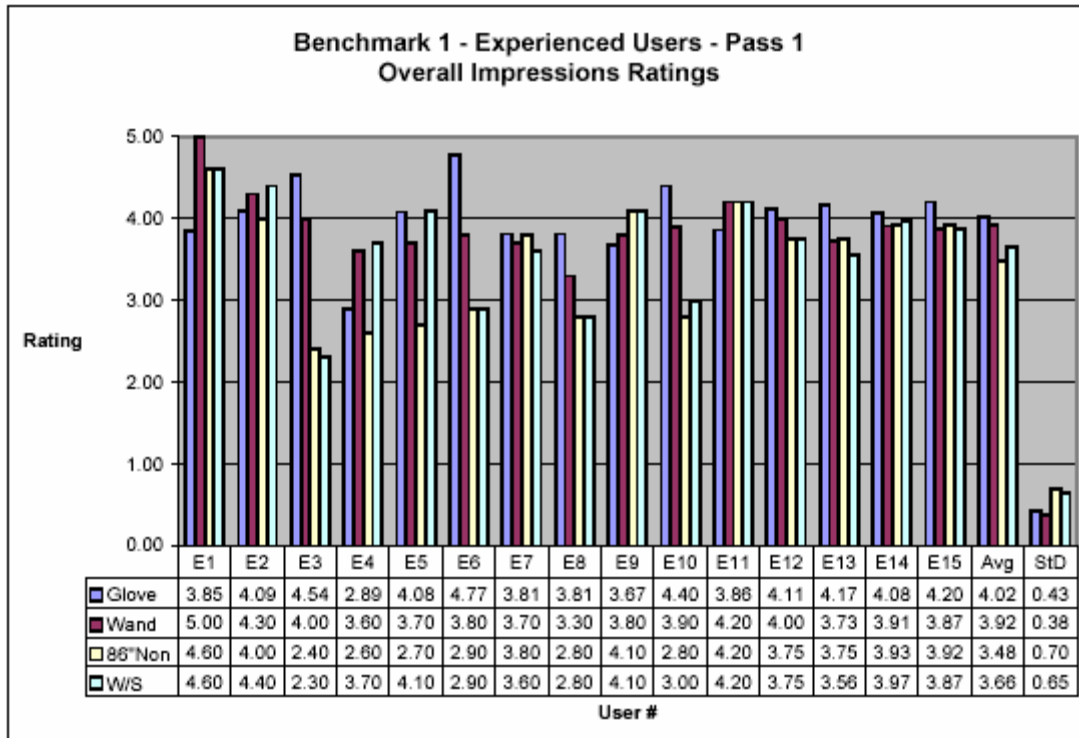
**Figure B- 4: B1Ep1Mov Experienced User Manipulation Ratings**

**Table B- 4: B1Ep1Mov Experienced User Manipulation Ratings Statistics**

Benchmark 1, Experienced Users, Pass 1, Manipulation Ratings												
B1Ep1	# Users	Mean	St. Dev.		Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	4.053	0.374		3.60	4.80	0.0425	No	9.23%			
Wand	15	3.613	0.635		2.60	5.00	0.0917	Yes	17.56%			
86"Non	15	3.343	0.778		2.40	4.80	0.0019	No	23.27%			
W/S	15	3.560	0.811		2.40	5.00	> 0.1000	Yes	22.79%			
B1Ep1		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	2.83	Pr > F	0.0832
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				3.1400	0.0875	Yes	14	7.14	0.0182	No	Means	Glove 12.18%
Glove vs 86"Non				9.3300	0.0490	No	14	8.50	0.0113	No	Var/Mean	Glove 21.24%
Glove vs W/S				7.6300	0.0100	No	14	4.50	0.0522	No	Var/Mean	Glove 13.85%
Wand vs 86"Non				1.0200	0.3201	Yes	14	2.04	0.1754	Yes	Neither	Wand 8.08%
Wand vs W/S		0.8084	0.3686			Yes	14	0.07	0.7995	Yes	Neither	Neither
86"Non vs W/S				0.0500	0.8187	Yes	14	2.91	0.1103	Yes	Neither	Neither



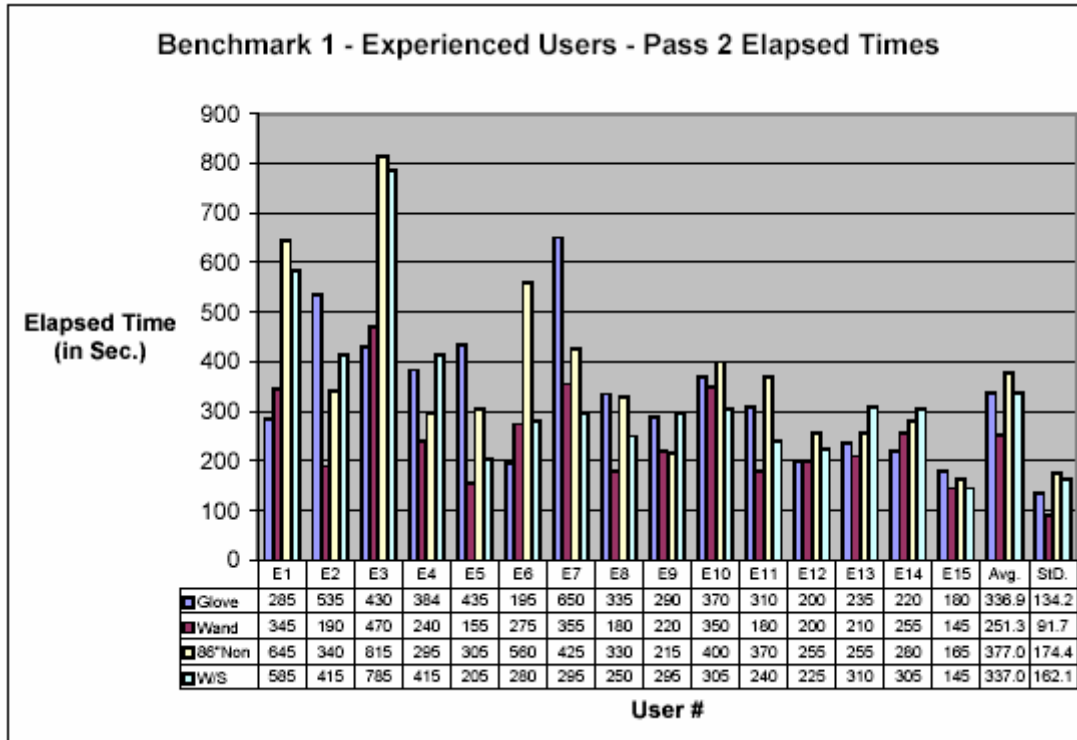
Benchmark 1, Experienced Users, Pass 1, General Impressions Ratings												
B1Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	15	4.267	0.419	3.60	5.00	> 0.1000	Yes	9.81%				
Wand	15	4.210	0.353	3.80	5.00	< 0.0001	No	8.38%				
86"Non	15	3.547	0.791	2.20	4.80	0.0003	No	22.30%				
W/S	15	3.747	0.795	2.20	5.00	0.0012	No	21.20%				
B1Ep1	Homogeneity of Variance			Roy's Greatest Root:			F(3,12)	2.84	Pr > F 0.0829			
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.4200	0.5241	Yes	14	0.86	0.3688	Yes	Neither	Neither	
Glove vs 86"Non			8.2600	0.0077	No	14	7.70	0.0149	No	Var/Mean	Glove 20.30%	
Glove vs W/S			4.7000	0.0389	No	14	3.66	0.0763	No	Var/Mean	Glove 13.88%	
Wand vs 86"Non			9.8500	0.0040	No	14	9.60	0.0079	No	Var/Mean	Wand 18.69%	
Wand vs W/S			5.6800	0.0242	No	14	6.11	0.0269	No	Var/Mean	Wand 12.36%	
86"Non vs W/S			0.0000	0.9821	Yes	14	2.06	0.1733	Yes	Neither	Neither	



**Figure B- 6: B1Ep1Ovr Experienced User Pass 1 Overall Impressions Ratings**

**Table B- 6: B1Ep1Ovr Experienced User Overall Pass 1 Impressions Ratings Statistics**

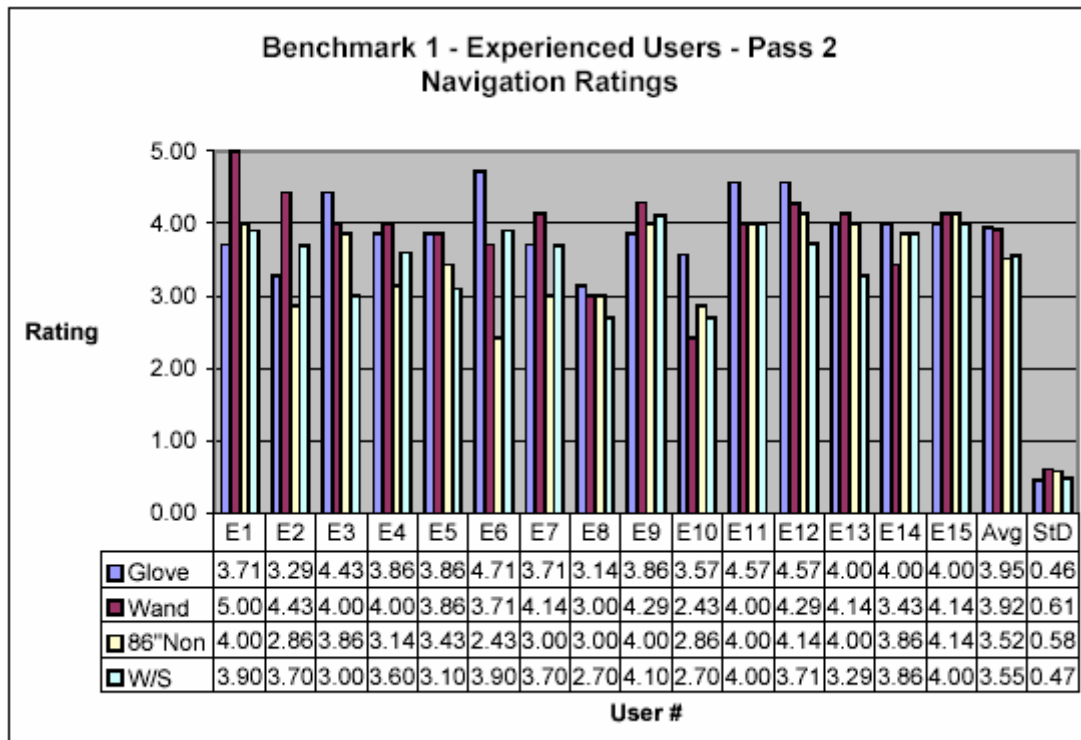
Benchmark 1, Experienced Users, Pass 1, Overall Impressions Ratings									
B1Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.	
Glove	15	4.022	0.431	3.30	4.77	> 0.1000	Yes	10.70%	
Wand	15	3.921	0.383	2.40	5.00	0.0535	Yes	9.77%	
86*Non	15	3.483	0.701	2.30	4.60	0.0136	No	20.13%	
W/S	15	3.657	0.648	2.59	4.60	> 0.1000	Yes	17.71%	
B1Ep1		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	2.63
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F
		Chi-Sq	Pr > ChiSq	F Value	Pr > F				
Glove vs Wand		0.1843	0.6677			Yes	14	0.53	0.4798
Glove vs 86*Non				4.8900	0.0354	No	14	5.76	0.0308
Glove vs W/S		2.1980	0.1382			Yes	14	2.38	0.1452
Wand vs 86*Non				6.8900	0.0139	No	14	9.18	0.0090
Wand vs W/S		3.5745	0.0587			Yes	14	3.44	0.0846
86*Non vs W/S				0.1800	0.6760	Yes	14	2.10	0.1697
								Equal Means ?	Statistically Significant
									Statistically Better
									Neither
									Var/Mean
									Neither
									Var/Mean
									Wand
									12.58%
									Wand
									7.22%
									Neither



**Figure B- 7: B1Ep2Tim Experienced User Elapsed Times**

**Table B- 7: B1Ep2Tim Experienced User Elapsed Times Statistics**

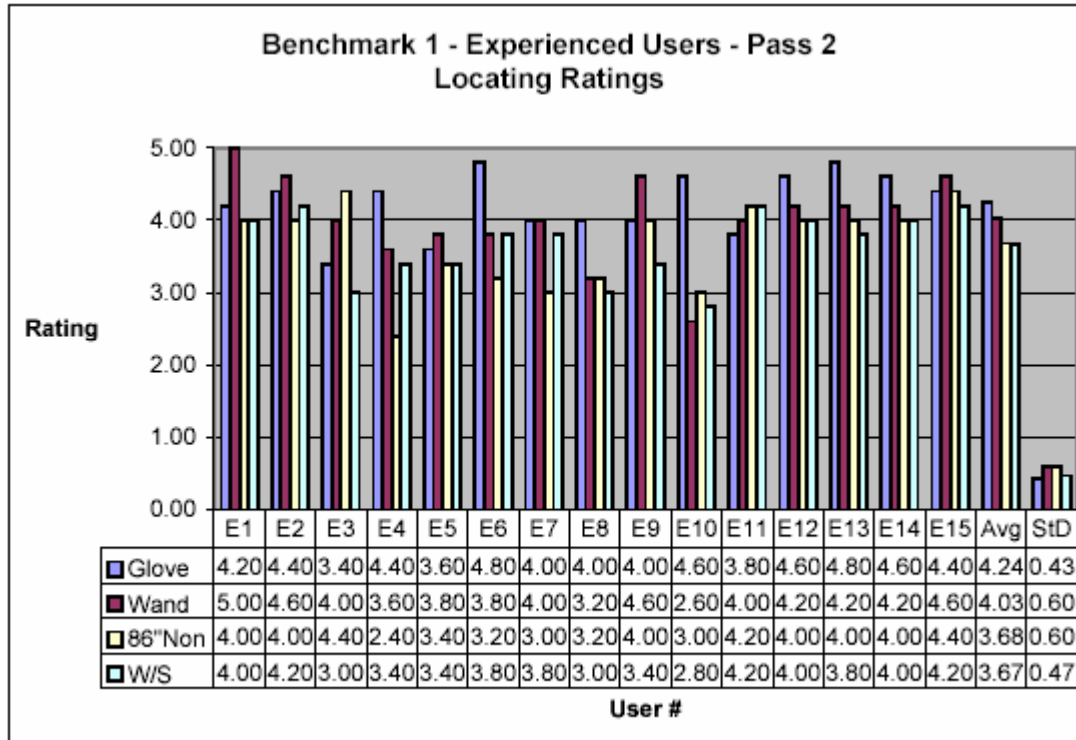
Benchmark 1, Experienced Users, Pass 2, Elapsed Times											
B1Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	336.90	134.20	180.00	650.00	> 0.1000	Yes	39.83%			
Wand	15	251.30	91.66	145.00	470.00	> 0.1000	Yes	36.47%			
86*Non	15	377.00	174.40	165.00	815.00	> 0.1000	Yes	46.26%			
W/S	15	337.00	162.10	145.00	785.00	0.0008	No	48.10%			
B1Ep2		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	8.74	Pr > F	0.0022
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand		1.9171	0.1662			Yes	14	6.01	0.0280	No	Means
Glove vs 86*Non				0.6500	0.4252	Yes	14	0.65	0.4332	Yes	Neither
Glove vs W/S		0.4794	0.4887			Yes	14	0.00	0.9989	Yes	Neither
Wand vs 86*Non				2.4500	0.1291	Yes	14	19.12	0.0006	No	Means
Wand vs W/S		4.1719	0.0411			No	14	9.43	0.0083	No	Var/Mean
86*Non vs W/S				0.0400	0.8350	Yes	14	0.16	0.2070	Yes	Neither



**Figure B- 8: B1Ep2Nav Experienced User Navigation Ratings**

**Table B- 8: B1Ep2Nav Experienced User Navigation Ratings Statistics**

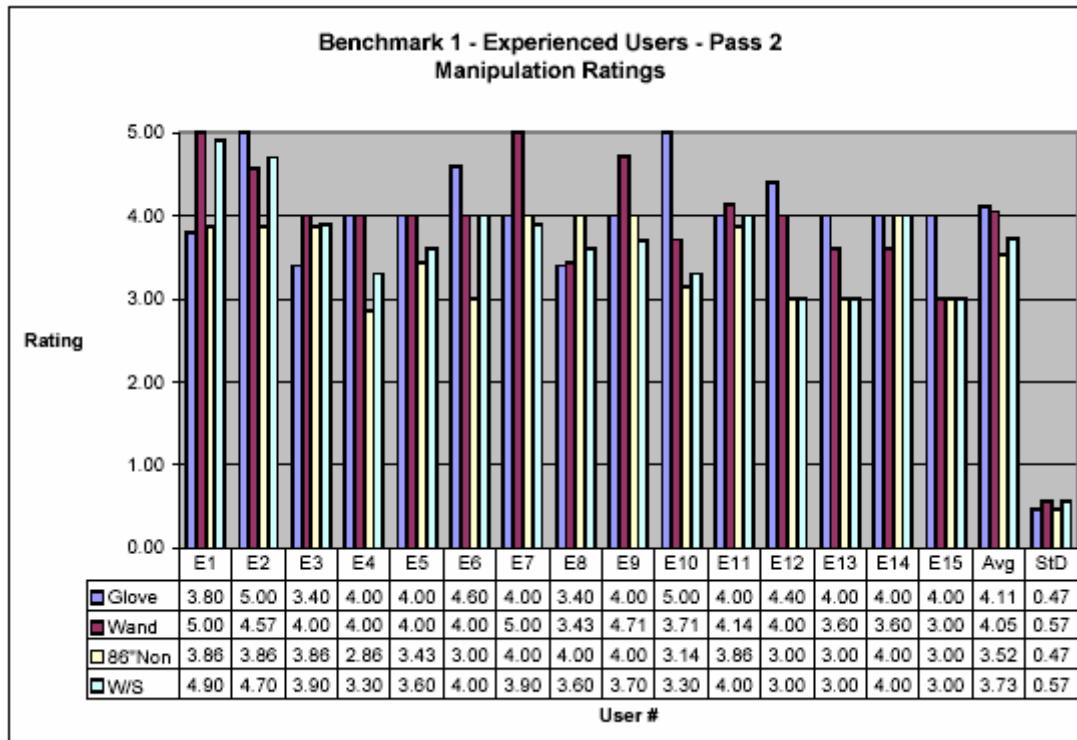
Benchmark 1, Experienced Users, Pass 2, Navigation Ratings											
B1Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	3.952	0.460	3.14	4.71	> 0.1000	Yes	11.63%			
Wand	15	3.924	0.611	2.43	5.00	0.0575	Yes	15.56%			
86\"Non	15	3.515	0.578	2.43	4.14	0.0081	No	16.44%			
W/S	15	3.551	0.474	2.70	4.10	0.0418	No	13.33%			
B1Ep2		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	7.91	Pr > F	0.0035
		Bartlett's Test	Levene's Test	Equal Var ?	df	F(1,14)	Pr > F	Means	Statistically Significant	Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand		1.0726	0.3004			Yes	14	0.02	0.8769	Yes	Neither
Glove vs 86\"Non				1.5600	0.2219	Yes	14	7.69	0.0150	No	Means
Glove vs W/S				0.0200	0.8879	Yes	14	9.07	0.0093	No	Means
Wand vs 86\"Non				0.0400	0.8369	Yes	14	6.60	0.0223	No	Means
Wand vs W/S				0.6600	0.4246	Yes	14	9.50	0.0081	No	Means
86\"Non vs W/S				1.2500	0.2723	Yes	14	0.05	0.8230	Yes	Neither



**Figure B- 9: B1Ep2Loc Experienced User Locating Ratings**

**Table B- 9: B1Ep2Loc Experienced User Locating Ratings Statistics**

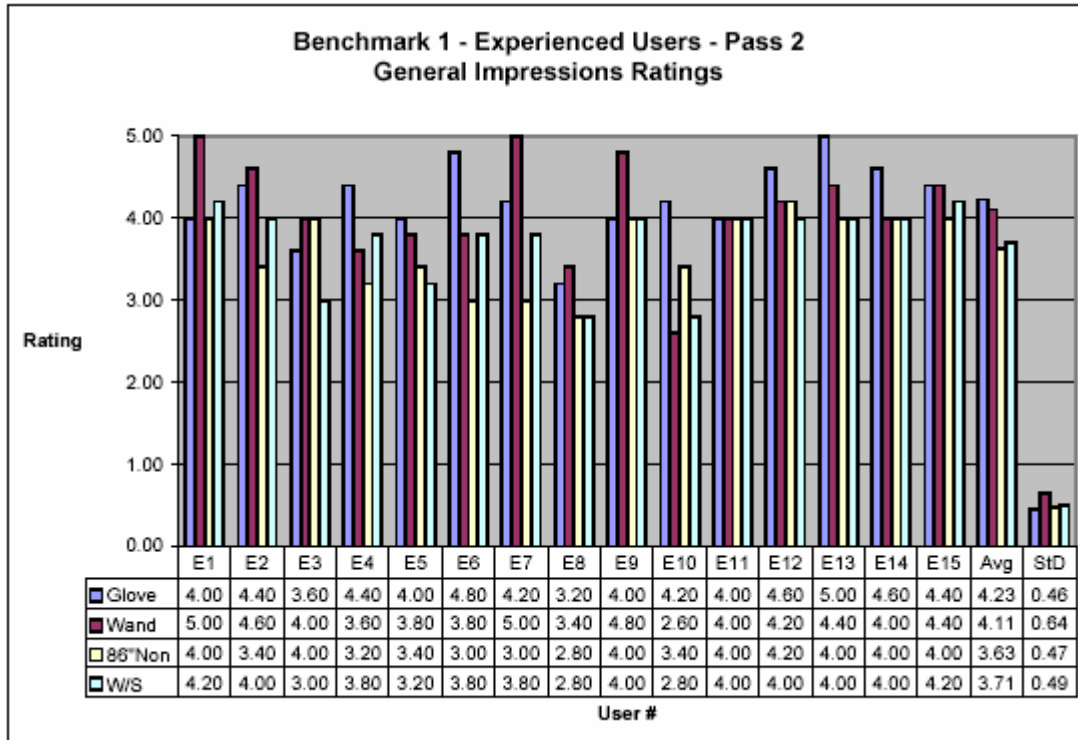
Benchmark 1, Experienced Users, Pass 2, Locating Ratings											
B1Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	4.240	0.429	3.40	4.80	> 0.1000	Yes	10.12%			
Wand	15	4.027	0.599	2.60	5.00	> 0.1000	Yes	14.88%			
86\"Non	15	3.680	0.599	2.40	4.40	0.0006	No	16.28%			
W/S	15	3.667	0.470	2.80	4.20	0.0692	Yes	12.82%			
B1Ep2		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	12.67	Pr > F	0.0005
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand		1.4853	0.2230			Yes	14	1.23	0.2869	Yes	Neither
Glove vs 86\"Non				2.0500	0.1638	Yes	14	7.47	0.0162	No	Means
Glove vs W/S		0.1130	0.7367			Yes	14	17.66	0.0009	No	Means
Wand vs 86\"Non				0.0000	0.9984	Yes	14	7.43	0.0164	No	Means
Wand vs W/S		0.7906	0.3739			Yes	14	11.26	0.0047	No	Means
86\"Non vs W/S				1.2300	0.2772	Yes	14	0.01	0.9289	Yes	Neither



**Figure B- 10: B1Ep2Mov Experienced User Manipulation Ratings**

**Table B- 10: B1Ep2Mov Experienced User Manipulation Ratings Statistics**

Benchmark 1, Experienced Users, Pass 2, Manipulation Ratings									
B1Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.	
Glove	15	4.107	0.471	3.40	5.00	0.0002	No	11.48%	
Wand	15	4.051	0.569	3.00	5.00	> 0.1000	Yes	14.06%	
86°Non	15	3.525	0.467	2.86	4.00	0.0009	No	13.25%	
W/S	15	3.727	0.573	3.00	4.90	> 0.1000	Yes	15.36%	
B1Ep2		Homogeneity of Variance			Roy's Greatest Root:			F(3,12)	
		Bartlett's Test	Levene's Test	Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Pr > F
		Chi-Sq	Pr > ChiSq	F Value	Pr > F			Statistically Significant	Statistically Better
Glove vs Wand		0.4808	0.4881	0.0000	0.9633	Yes	14	0.09	0.7630
Glove vs 86°Non				0.0000	0.9633	Yes	14	8.45	0.0115
Glove vs W/S		0.5090	0.4756	0.9700	0.3337	Yes	14	4.11	0.0622
Wand vs 86°Non				0.9700	0.3337	Yes	14	14.35	0.0020
Wand vs W/S		0.0004	0.9839	0.8900	0.3526	Yes	14	7.16	0.0181
86°Non vs W/S				0.8900	0.3526	Yes	14	3.17	0.0968



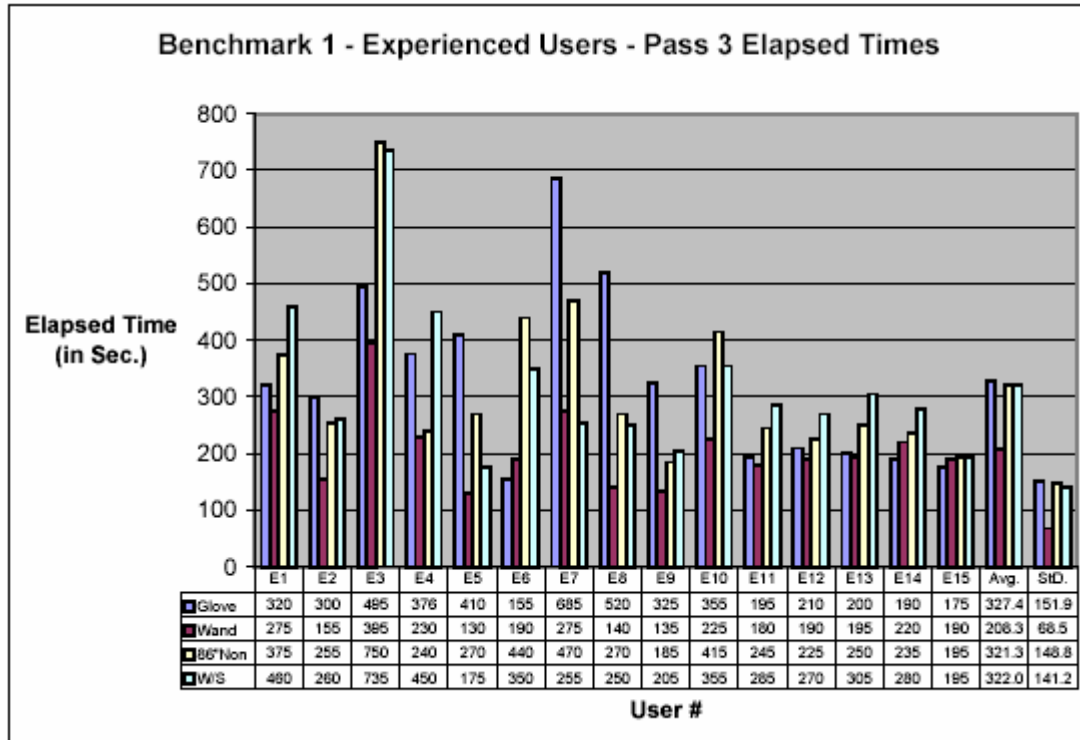
**Figure B- 11: B1Ep2Gen Experienced User General Impressions Ratings**

**Table B- 11: B1Ep2Gen Experienced User General Impressions Ratings Statistics**

Benchmark 1, Experienced Users, Pass 2, General Impressions Ratings											
B1Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	4.227	0.459	3.20	5.00	> 0.1000	Yes	10.86%			
Wand	15	4.107	0.641	2.60	5.00	> 0.1000	Yes	15.60%			
86*Non	15	3.627	0.471	2.80	4.20	0.0002	No	12.99%			
W/S	15	3.707	0.495	2.80	4.20	0.0005	No	13.35%			
B1Ep2	Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	18.44	Pr > F < 0.0001	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	1.4785	0.2240			Yes	14	0.41	0.5336	Yes	Neither	Neither
Glove vs 86*Non			0.0100	0.9042	Yes	14	16.02	0.0013	No	Means	Glove 16.54%
Glove vs W/S			0.1000	0.7594	Yes	14	22.32	0.0003	No	Means	Glove 14.03%
Wand vs 86*Non			1.2300	0.2764	Yes	14	8.20	0.0125	No	Means	Wand 13.23%
Wand vs W/S			0.8500	0.3643	Yes	14	12.35	0.0034	No	Means	Wand 10.79%
86*Non vs W/S			0.0700	0.8001	Yes	14	0.39	0.5411	Yes	Neither	Neither



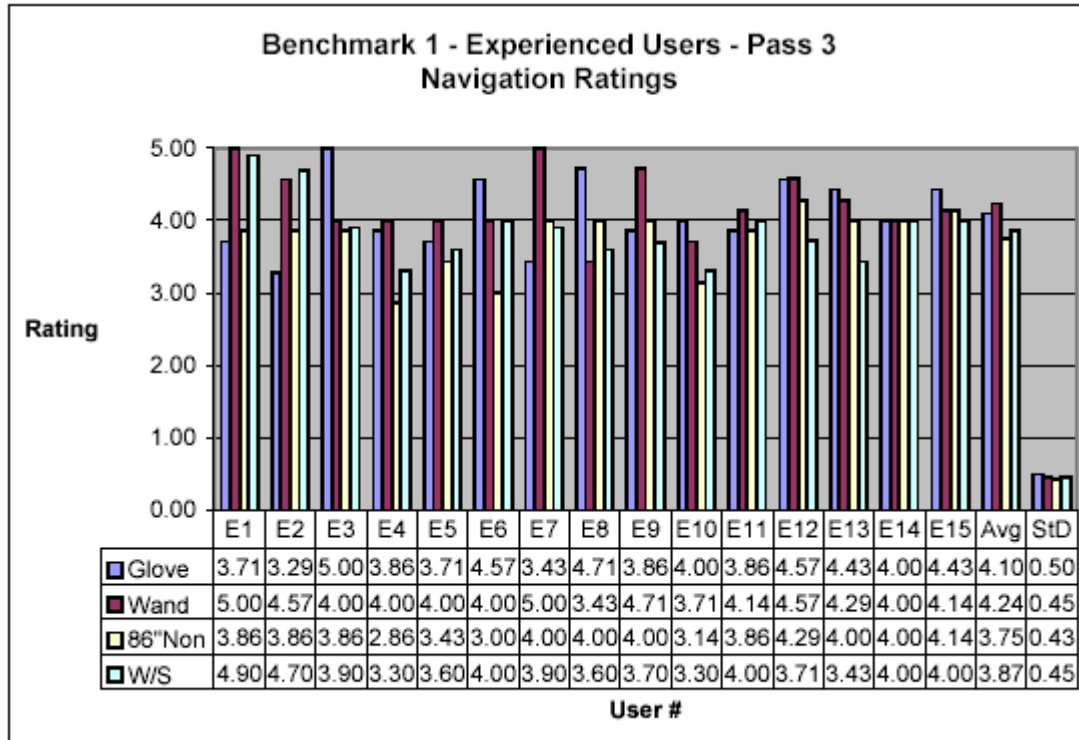




**Figure B- 13: B1Ep3Tim Experienced User Elapsed Times**

**Table B- 13 B1Ep3Tim Experienced User Elapsed Times**

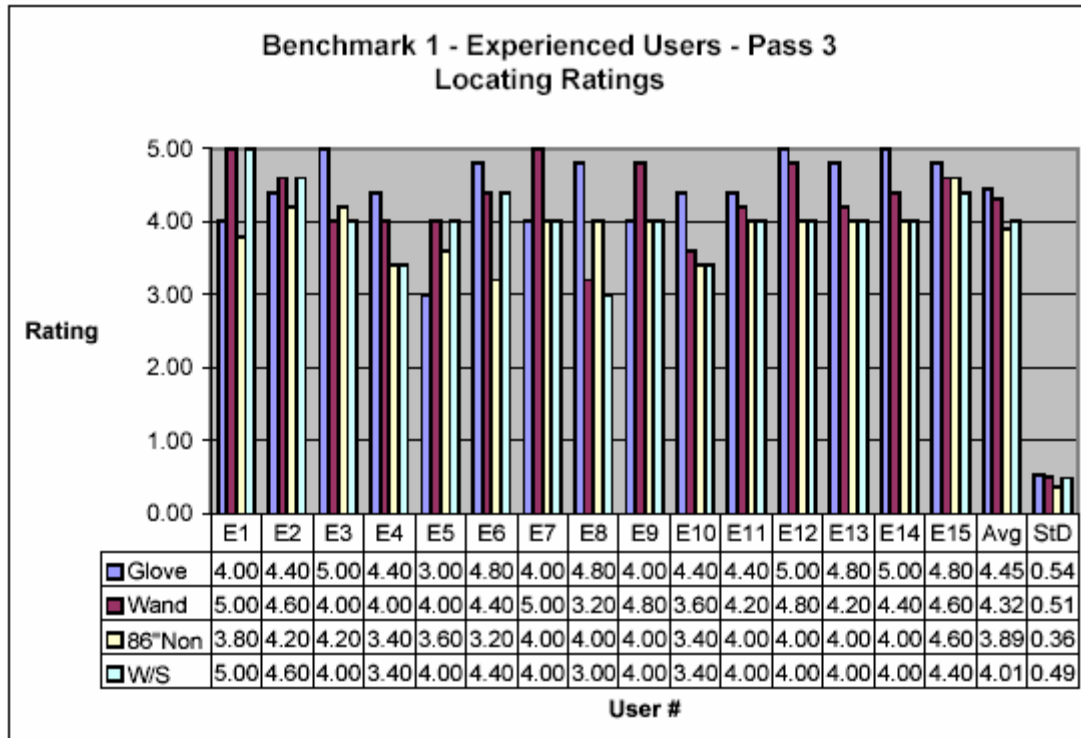
Benchmark 1, Experienced Users, Pass 3, Elapsed Times													
B1Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.					
Glove	15	327.40	151.90	155.00	685.00	> 0.1000	Yes	46.40%					
Wand	15	208.30	68.47	130.00	395.00	> 0.1000	Yes	32.87%					
86"Non	15	321.30	148.60	185.00	750.00	0.0007	No	46.25%					
W/S	15	322.00	141.20	175.00	735.00	0.0615	No	43.85%					
B1Ep3		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)		12.74	Pr > F	0.0005
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand		7.8080	0.0052			No	14	10.32	0.0063	No	Var/Mean	Wand	57.18%
Glove vs 86"Non				0.0000	0.9518	Yes	14	0.02	0.8803	Yes	Neither	Neither	
Glove vs W/S				1.2800	0.2620	Yes	14	0.00	0.9117	Yes	Neither	Neither	
Wand vs 86"Non				1.8500	0.0064	No	14	19.30	0.0006	No	Var/Mean	Wand	54.25%
Wand vs W/S				4.7000	0.0389	No	14	24.55	0.0002	No	Var/Mean	Wand	54.58%
86"Non vs W/S				0.0200	0.8984	Yes	14	0.00	0.9788	Yes	Neither	Neither	



**Figure B- 14: B1Ep3Nav Experienced User Navigation Ratings**

**Table B- 14 B1Ep3Nav Experienced User Navigation Ratings**

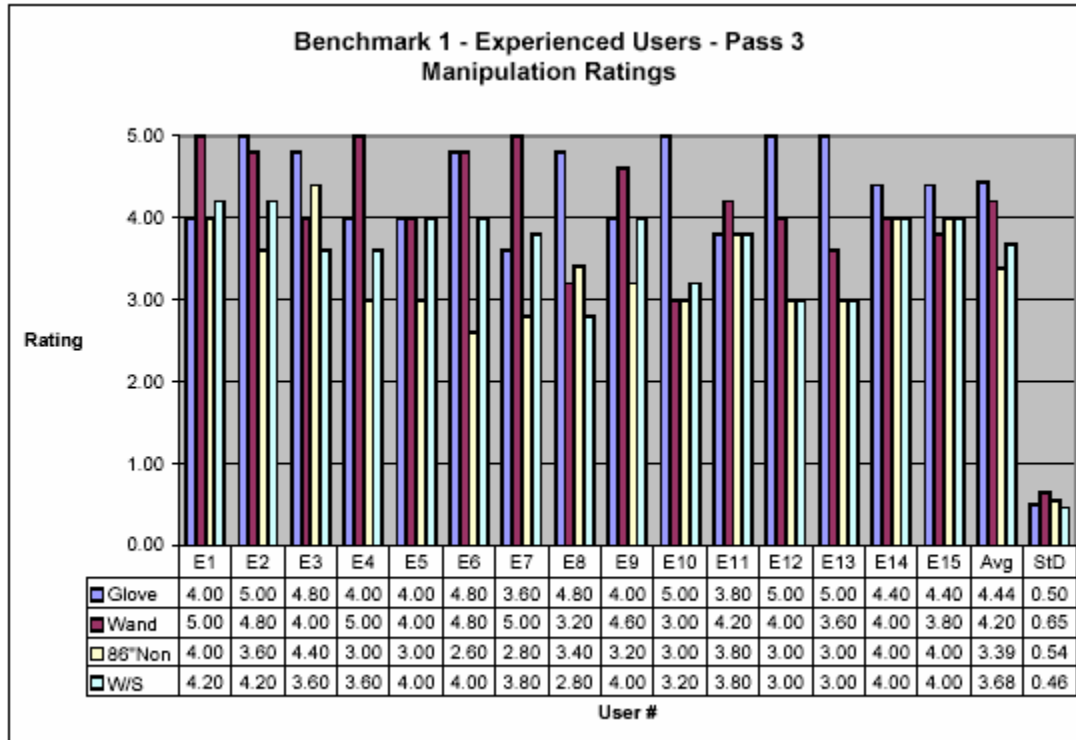
Benchmark 1, Experienced Users, Pass 1, Navigation Ratings																			
B1Ep1	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	Coeff. of Var.							
Glove	15	3.724		0.335		3.14		4.29		> 0.1000	Yes	9.00%							
Wand	15	3.866		0.220		3.14		5.00		> 0.1000	Yes	5.69%							
86"Non	15	3.438		0.723		2.29		4.29		> 0.1000	Yes	21.03%							
W/S	15	3.621		0.713		2.00		4.40		> 0.1000	Yes	19.70%							
B1Ep1		Homogeneity of Variance				Roy's Greatest Root:				F(3,12)		2.13		Pr > F		0.1502			
		Bartlett's Test		Levene's Test		Equal Var ?		df		F(1,14)		Pr > F		Equal Means ?		Statistically Significant		Statistically Better	
		Chi-Sq	Pr > ChiSq	F value	Pr > F														
Glove vs Wand		0.9922	0.3192			Yes	14	0.95	0.3459	Yes	2	Neither					Neither		
Glove vs 86"Non		7.3130	0.0068			No	14	1.62	0.2243	Yes	Var			Glove		8.32%			
Glove vs W/S		7.0792	0.0078			No	14	0.19	0.6678	Yes	Var			Glove		2.84%			
Wand vs 86"Non		3.2038	0.0735			Yes	14	6.31	0.0248	No	Means			Wand		12.45%			
Wand vs W/S		3.0386	0.0813			Yes	14	2.06	0.1735	Yes	Neither			Neither					
86"Non vs W/S		0.0025	0.9605			Yes	14	1.98	0.1809	Yes	Neither			W/S		5.32%			



**Figure B- 15: B1Ep3Loc Experienced User Locating Ratings**

**Table B- 15: B1Ep3Loc Experienced User Locating Ratings Statistics**

Benchmark 1, Experienced Users, Pass 3, Locating Ratings										
B1Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	15	4.453	0.542	3.00	5.00	0.8880	Yes	12.18%		
Wand	15	4.320	0.512	3.20	5.00	> 0.1000	Yes	11.84%		
86\"Non	15	3.893	0.362	3.20	4.60	0.0021	No	9.29%		
W/S	15	4.013	0.493	3.00	5.00	0.1500	No	12.28%		
B1Ep3		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					Statistically Significant
Glove vs Wand		0.0460	0.8303			Yes	14	0.43	0.5248	Yes
Glove vs 86\"Non				1.1700	0.2892	Yes	14	14.91	0.0017	No
Glove vs W/S				0.0900	0.7682	Yes	14	4.79	0.0462	No
Wand vs 86\"Non				1.6300	0.2122	Yes	14	9.39	0.0084	No
Wand vs W/S				0.0200	0.8866	Yes	14	12.43	0.0034	No
86\"Non vs W/S				1.1000	0.3026	Yes	14	0.74	0.4029	Yes
										Statistically Better
										Neither
										Means
										Glove
										14.38%
										Means
										Glove
										10.96%
										Means
										Wand
										10.97%
										Means
										Wand
										7.65%
										Neither

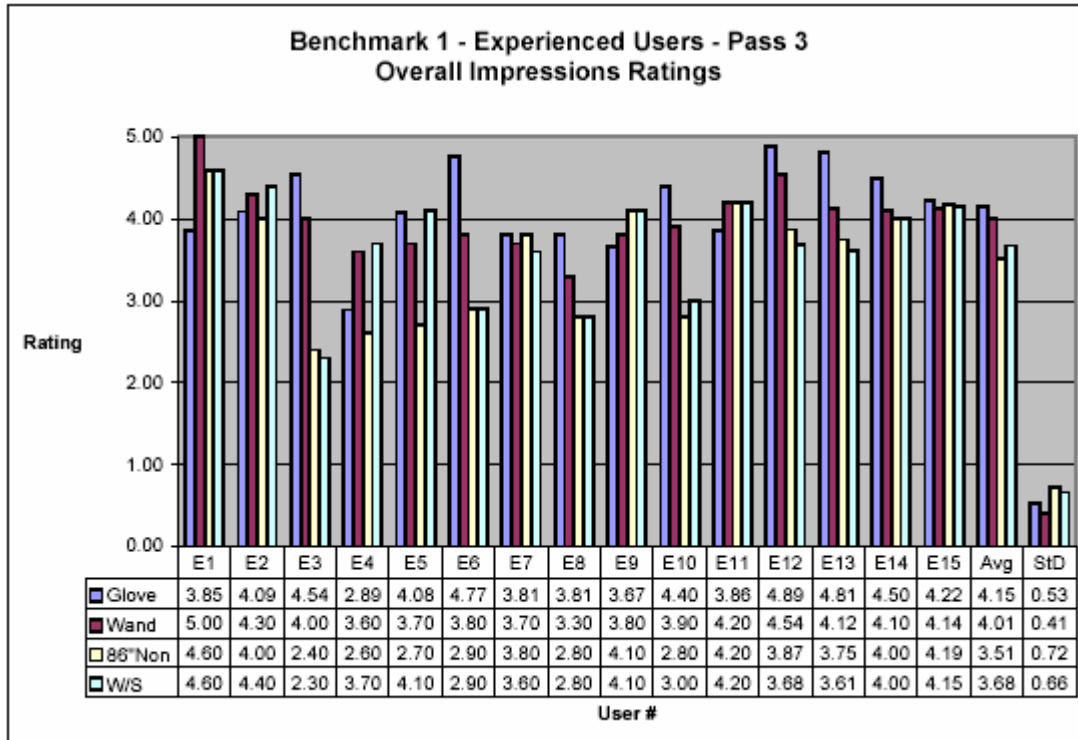


**Figure B- 16: B1Ep3Mov Experienced User Manipulation Ratings**

**Table B- 16: B1Ep3Mov Experienced User Manipulation Ratings Statistics**

Benchmark 1, Experienced Users, Pass 3, Manipulation Ratings										
B1Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	15	4.440	0.503	3.60	5.00	0.0321	No	11.32%		
Wand	15	4.200	0.650	3.00	5.00	> 0.1000	Yes	15.48%		
86\"Non	15	3.387	0.542	2.60	4.40	0.0336	No	16.01%		
W/S	15	3.680	0.465		4.20	0.0469	No	12.63%		
B1Ep3		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	11.99	Pr > F
		Bartlett's Test		Levene's Test		df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					
Glove vs Wand				1.8700	0.1825	Yes	14	0.84	0.3739	Yes
Glove vs 86\"Non				0.2200	0.6411	Yes	14	29.58	0.0000	No
Glove vs W/S				0.2200	0.6402	Yes	14	12.31	0.0035	No
Wand vs 86\"Non				0.8900	0.3535	Yes	14	12.72	0.0031	No
Wand vs W/S				2.5400	0.1223	Yes	14	17.40	0.0009	No
86\"Non vs W/S				0.6700	0.4198	Yes	14	3.54	0.0810	No

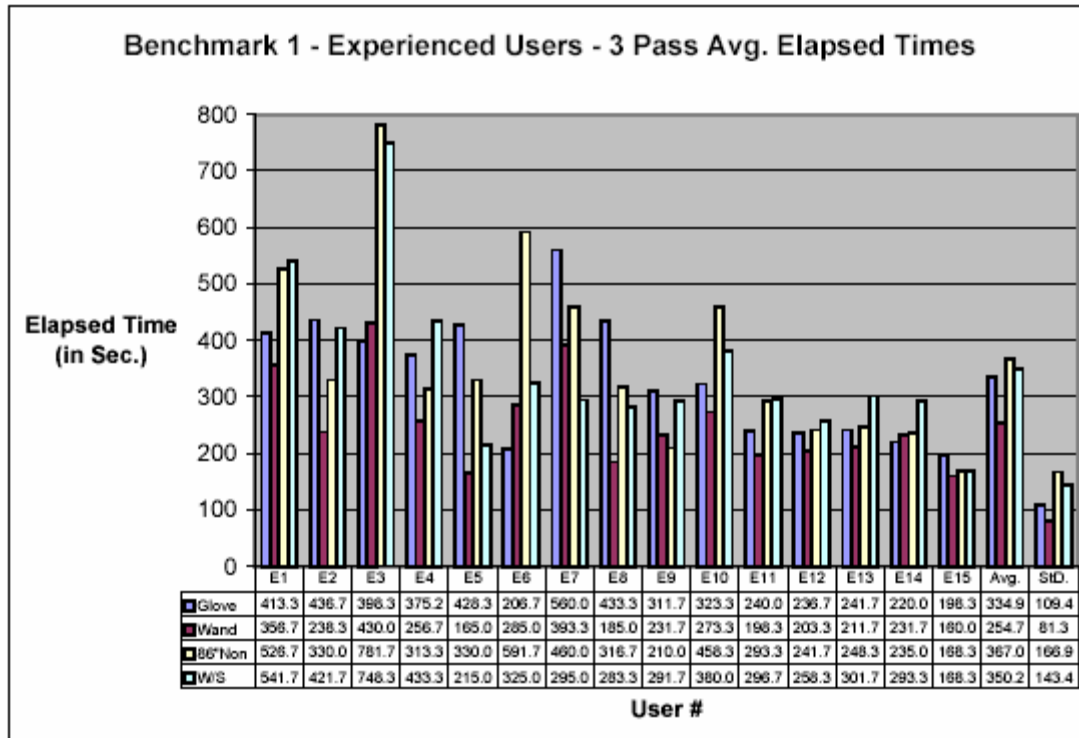




**Figure B- 18: B1Ep3Ovr Experienced User Overall Impressions Ratings**

**Table B- 18: B1Ep3Ovr Experienced User Overall Impressions Ratings**

Benchmark 1, Experienced Users, Pass 3, Overall Impressions Ratings												
B1Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	15	4.146	0.531	2.89	4.89	> 0.1000	Yes	12.80%				
Wand	15	4.013	0.413	3.30	5.00	> 0.1000	Yes	10.28%				
86"Non	15	3.514	0.723	2.40	4.60	0.0349	No	20.58%				
W/S	15	3.676	0.659	2.30	4.60	> 0.1000	Yes	17.92%				
B1Ep3		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	3.71	Pr > F 0.0425		
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		0.8466	0.3575			Yes	14	0.84	0.3751	Yes	Neither	Neither
Glove vs 86"Non				2.6100	0.1171	Yes	14	7.67	0.0151	No	Means	Glove 17.99%
Glove vs W/S		0.6266	0.4286			Yes	14	3.57	0.0798	No	Means	Glove 12.79%
Wand vs 86"Non				8.1100	0.0081	No	14	12.97	0.0029	No	Var/Means	Wand 14.20%
Wand vs W/S		2.8550	0.0911			Yes	14	5.31	0.0371	No	Means	Wand 9.17%
86"Non vs W/S				0.2700	0.6051	Yes	14	1.79	0.2026	Yes	Neither	Neither

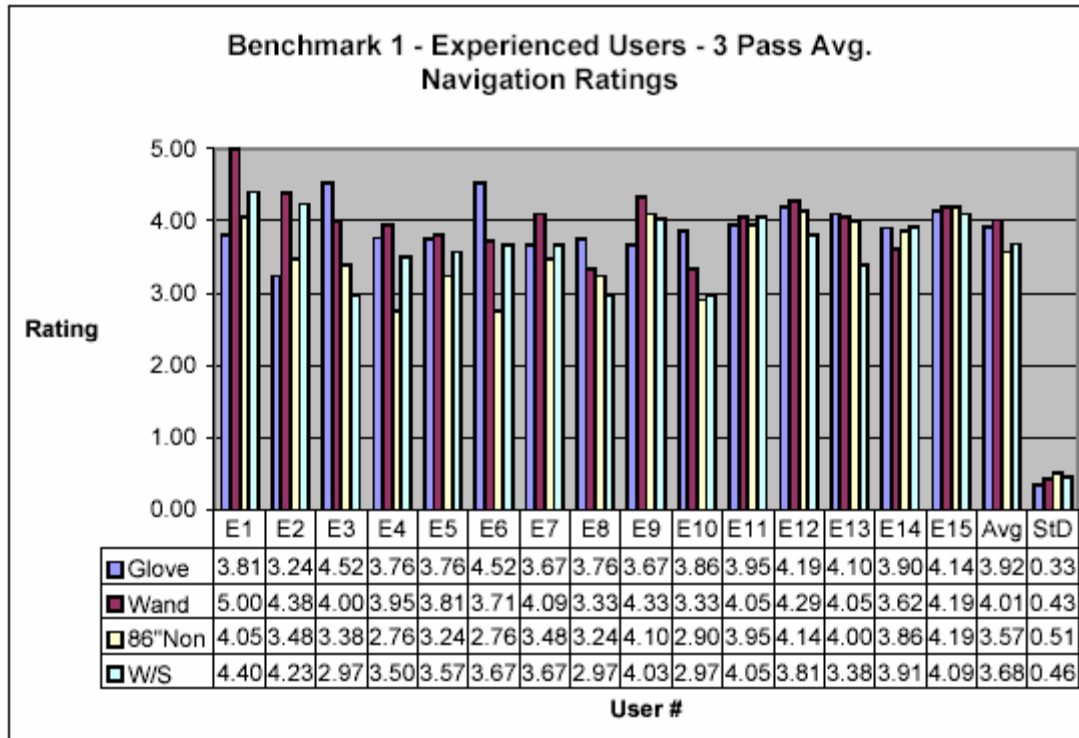


**Figure B- 19: B1E3pAvgTim Experienced Users Elapsed Times**

**Table B- 19: B1E3pAvgTim Experienced Users Elapsed Times Statistics**

Benchmark 1, Experienced Users, 3 Pass Avg., Elapsed Times												
B1E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	15	334.90	109.40	198.30	560.00	0.0976	Yes	32.67%				
Wand	15	254.70	81.28	160.00	430.00	> 0.1000	Yes	31.91%				
86"Non	15	367.00	166.90	168.30	781.70	0.0099	No	45.48%				
W/S	15	350.20	143.40	168.30	748.30	0.0237	No	40.95%				
B1E3pA		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	11.20	Pr > F	0.0009
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		1.1759	0.2782			Yes	14	9.64	0.0078	No	Means	Wand 31.49%
Glove vs 86"Non				1.6600	0.2083	Yes	14	0.58	0.4593	Yes	Neither	Neither
Glove vs W/S				0.5500	0.4650	Yes	14	0.16	0.6954	Yes	Neither	Neither
Wand vs 86"Non				3.0800	0.0901	Yes	14	16.10	0.0013	No	Means	Wand 44.09%
Wand vs W/S				3.0800	0.0901	Yes	14	15.16	0.0016	No	Means	Wand 37.50%
86"Non vs W/S				0.2000	0.6575	Yes	14	0.39	0.5428	Yes	Neither	Neither



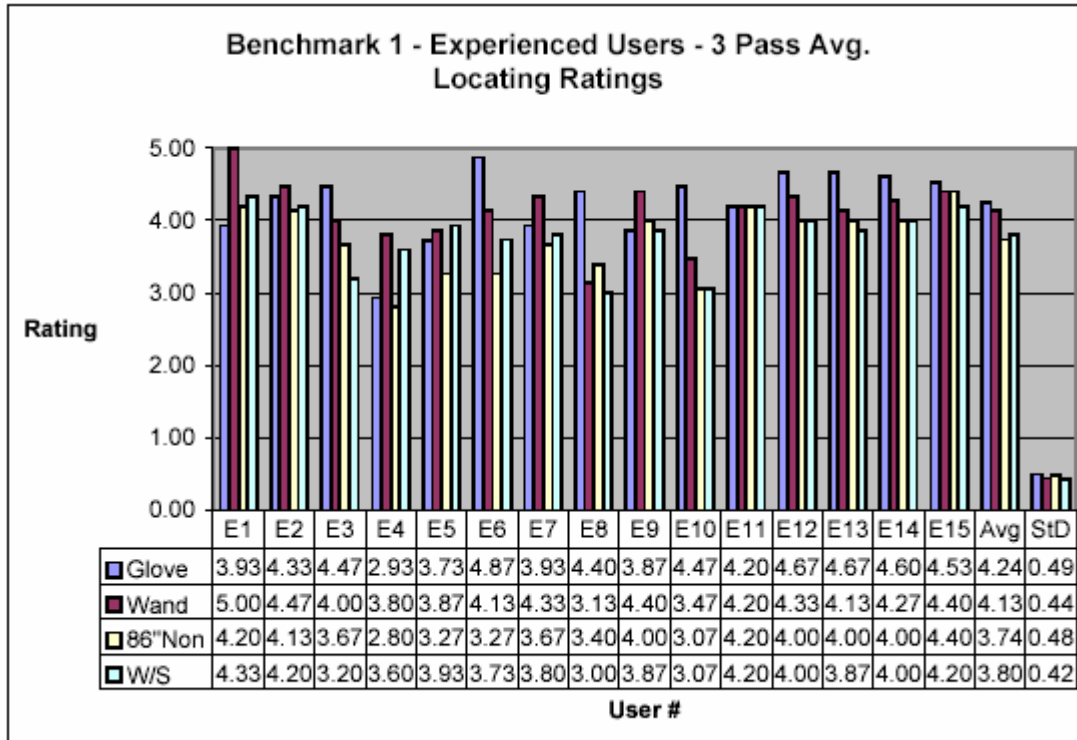


**Figure B- 20: B1E3pAvgNav Experienced Users Navigation Ratings**

**Table B- 20: B1E3pAvgNav Experienced Users Navigation Ratings**

Benchmark 1, Experienced Users, Pass 2, Navigation Ratings											
B1Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	3.952	0.460	3.14	4.71	> 0.1000	Yes	11.63%			
Wand	15	3.924	0.611	2.43	5.00	0.0575	Yes	15.56%			
86"Non	15	3.515	0.578	2.43	4.14	0.0081	No	16.44%			
W/S	15	3.551	0.474	2.70	4.10	0.0418	No	13.33%			
B1Ep2	Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	7.91	Pr > F 0.0035	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Means	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F					2		
Glove vs Wand	1.0726	0.3004			Yes	14	0.02	0.8769	Yes	Neither	Neither
Glove vs 86"Non			1.5600	0.2219	Yes	14	7.69	0.0150	No	Means	Glove 12.43%
Glove vs W/S			0.0200	0.8879	Yes	14	9.07	0.0093	No	Means	Glove 11.29%
Wand vs 86"Non			0.0400	0.8369	Yes	14	6.60	0.0223	No	Means	Wand 11.64%
Wand vs W/S			0.6600	0.4246	Yes	14	9.50	0.0081	No	Means	Wand 10.50%
86"Non vs W/S			1.2500	0.2723	Yes	14	0.05	0.8230	Yes	Neither	Neither

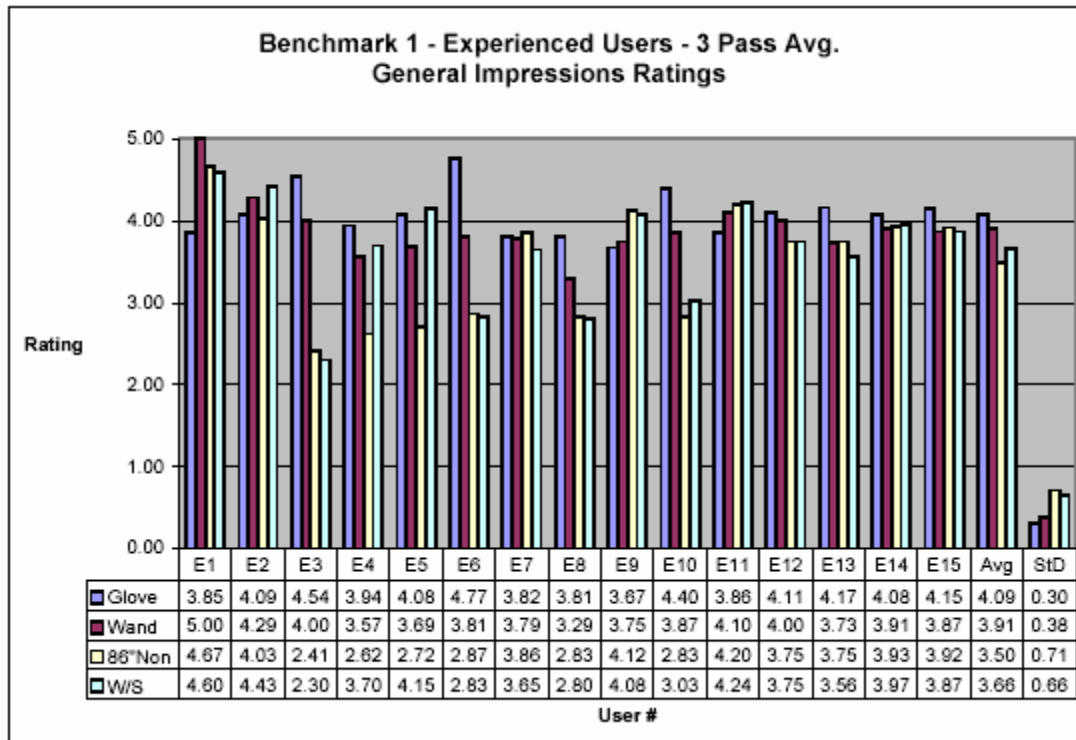




**Figure B- 22: B1E3pAvgMov Experienced User Manipulation Ratings**

**Table B- 22: B1E3pAvgMov Experienced User Manipulation Ratings Statistics**

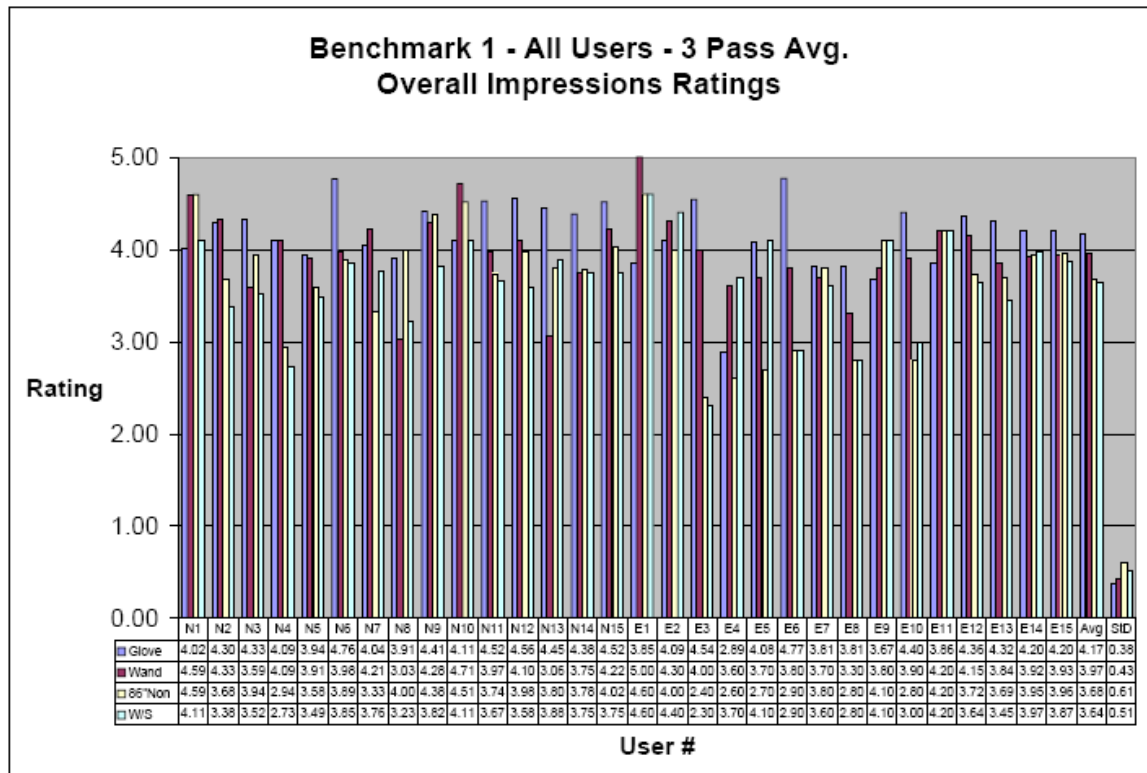
Benchmark 1, Experienced Users, 3 Pass Avg., Manipulation Ratings									
B1E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.	
Glove	15	4.199	0.373	3.73	4.93	> 0.1000	Yes	8.89%	
Wand	15	3.858	0.524	2.93	4.40	> 0.1000	Yes	13.57%	
86\"Non	15	3.368	0.561	2.40	4.20	0.0717	Yes	16.65%	
W/S	15	3.569	0.550	2.80	4.00	> 0.1000	Yes	15.40%	
B1E3pA		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	Pr > F
		Bartlett's Test		Levene's Test		df	F(1,14)	Pr > F	Equal Means ?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F				
Glove vs Wand		1.5226	0.2172			Yes 14	3.77	0.0727	No
Glove vs 86\"Non		2.1836	0.1395			Yes 14	18.77	0.0007	No
Glove vs W/S		1.9751	0.1599			Yes 14	12.00	0.0038	No
Wand vs 86\"Non		0.0635	0.8010			Yes 14	13.04	0.0028	No
Wand vs W/S		0.0313	0.8595			Yes 14	6.83	0.0205	No
86\"Non vs W/S		0.0056	0.9401			Yes 14	2.95	0.1077	Yes
								Statistically Significant	
								Statistically Better	
								Glove 8.84%	
								Glove 24.67%	
								Glove 17.65%	
								Wand 14.55%	
								Wand 8.10%	
								Neither	



**Figure B- 23: B1E3pAvgGen Experienced User General Impressions Ratings**

**Table B- 23: B1E3pAvgGen Experienced User General Impressions Ratings Statistics**

Benchmark 1, Experienced Users, 3 Pass Avg., General Impressions Ratings										
B1E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	15	4.089	0.297	3.67	4.78	> 0.1000	Yes	7.26%		
Wand	15	3.911	0.380	3.29	5.00	0.081	Yes	9.71%		
86*Non	15	3.501	0.707	2.41	4.67	0.0233	No	20.21%		
W/S	15	3.664	0.658	2.30	4.60	> 0.1000	Yes	17.96%		
B1E3pA	Homogeneity of Variance			Roy's Greatest Root: F(3,12) 2.31 Pr > F 0.1282						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						Statistically Better
Glove vs Wand	0.8152	0.3666			Yes	14	2.00	0.1718	Yes	Neither
Glove vs 86*Non			14.440	0.0010	No	14	6.42	0.0239	No	Var/Means Glove 16.80%
Glove vs W/S	7.8023	0.0052			No	14	3.57	0.0797	No	Var/Means Glove 11.60%
Wand vs 86*Non			6.7800	0.0146	No	14	8.03	0.0133	No	Var/Means Wand 11.71%
Wand vs W/S	3.8974	0.0484			No	14	2.84	0.1143	Yes	Var Wand 6.74%
86*Non vs W/S			0.1500	0.7044	Yes	14	1.80	0.2015	Yes	Neither



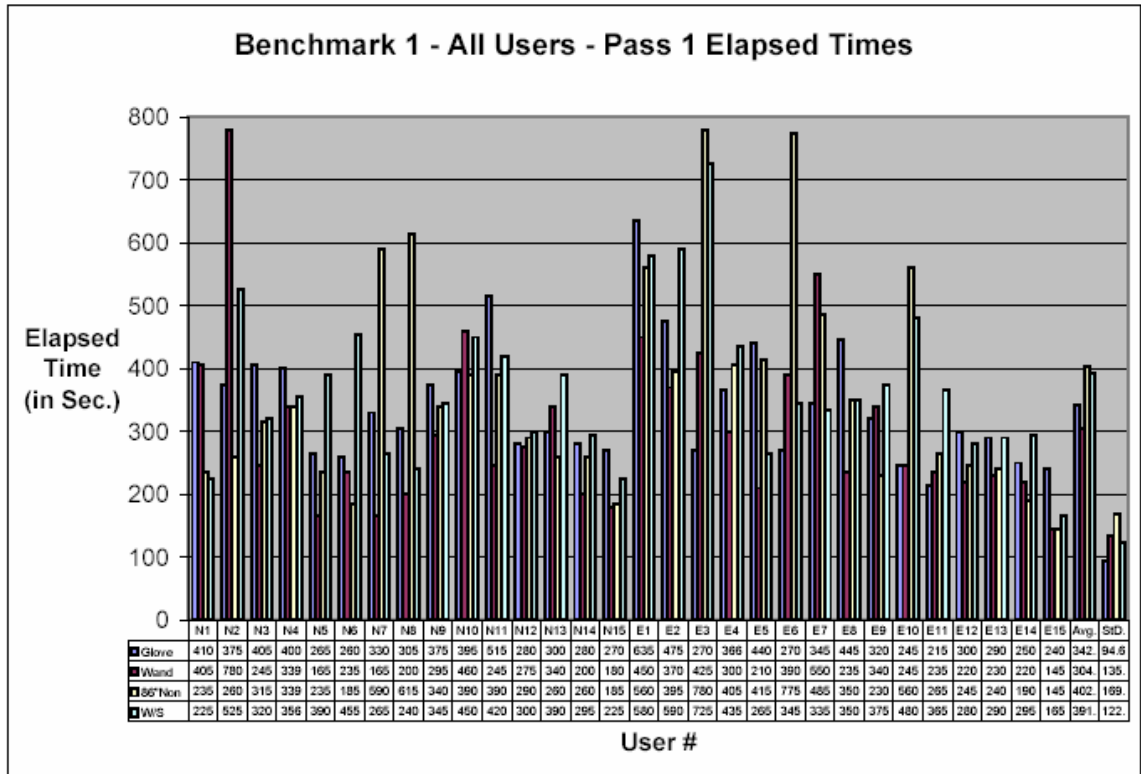
**Figure B- 24: B1E3pAvgOvr Experienced User Overall Impressions Ratings**

**Table B- 24: B1E3pAvgOvr Experienced User Overall Impressions Ratings Statistics**

Benchmark 1, Experienced Users, 3 Pass Avg., Overall Impressions Ratings											
B1E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	15	4.057	0.446	2.89	4.77	> 0.1000	Yes	10.98%			
Wand	15	3.943	0.384	3.30	5.00	> 0.1000	Yes	9.73%			
86*Non	15	3.481	0.702	2.40	4.60	0.0559	Yes	20.16%			
W/S	15	3.642	0.649	2.30	4.60	> 0.1000	Yes	17.82%			
B1E3pA	Homogeneity of Variance				Roy's Greatest Root: F(3,12)			3.00 Pr > F 0.0703			
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.3011	0.5832			Yes	14	0.66	0.4301	Yes	Neither	Neither
Glove vs 86*Non	2.6999	0.1004			Yes	14	6.63	0.0220	No	Means	Glove 16.55%
Glove vs W/S	1.8712	0.1713			Yes	14	3.01	0.1049	Yes	Neither	Neither
Wand vs 86*Non	4.6564	0.0309			No	14	10.65	0.0057	No	Var/Means	Wand 13.27%
Wand vs W/S	3.5777	0.0586			Yes	14	4.45	0.0533	No	Means	Wand 8.26%
86*Non vs W/S	0.0822	0.7743			Yes	14	1.75	0.2076	Yes	Neither	Neither

## **Appendix C**

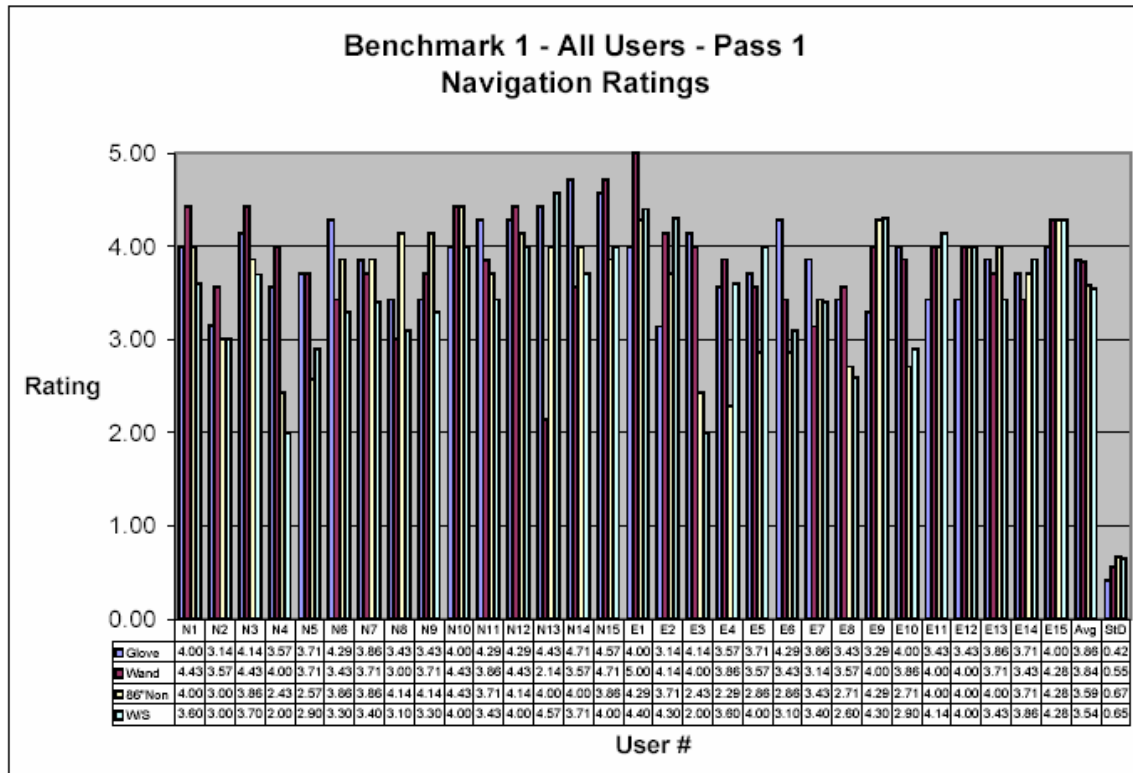
### **All Users Benchmark 1 (Navigation) Detail**



**Figure C- 1: B1ALLp1Tim All Users Elapsed Times**

**Table C- 1: B1ALLp1Tim All Users Elapsed Times Statistics**

Benchmark 1, All Users, Pass 1, Elapsed Times											
B1Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	30	269.90	122.60	155.00	685.00	< 0.0001	Yes	45.42%			
Wand	30	190.10	57.19	115.00	395.00	0.0311	No	30.08%			
86\"Non	30	257.90	126.40	95.00	750.00	< 0.0001	Yes	49.01%			
W/S	30	258.70	120.10	125.00	735.00	0.0012	Yes	46.42%			
B1Ap1	Homogeneity of Variance					Roy's Greatest Root:		F(3,27)	2.61	Pr > F	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			1.2200	0.2736	Yes	29	2.4600	0.1273	Yes	Neither	Neither
Glove vs 86\"Non	9.1000	0.0026			No	29	0.4500	0.5062	Yes	Var	86\"Non 4.65%
Glove vs W/S	1.8760	0.1708			Yes	29	1.2700	0.2686	Yes	Neither	Neither
Wand vs 86\"Non			0.8600	0.3585	Yes	29	3.0700	0.0905	No	Means	Wand 35.67%
Wand vs W/S			0.1200	0.7268	Yes	29	8.3500	0.0072	No	Means	Wand 36.09%
86\"Non vs W/S	2.9095	0.0881			Yes	28	0.0300	0.8697	Yes	Neither	Neither

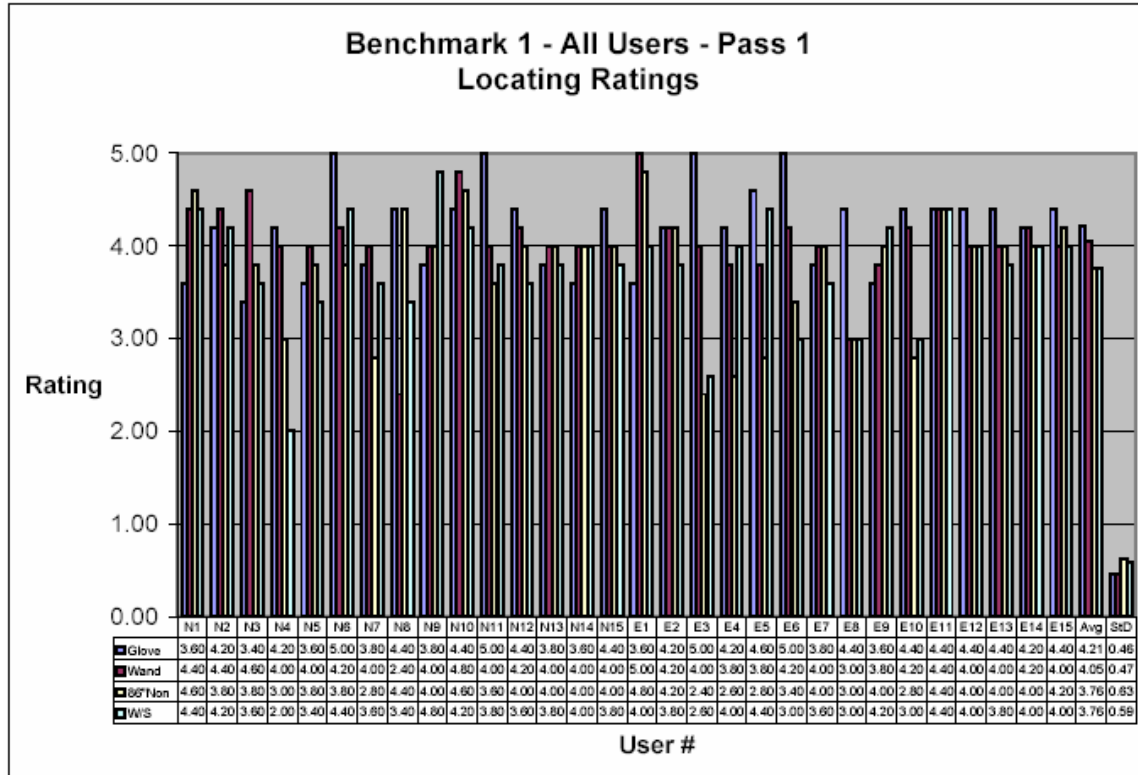


**Figure C- 2: B1ALLp1Nav All Users Navigation Ratings**

**Table C- 2: B1ALLp1Nav All Users Navigation Ratings Statistics**

Benchmark 1, All Users, Pass 1, Navigation Ratings											
B1Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	30	3.857	0.418	3.140	4.710	> 0.1000	Yes	10.84%			
Wand	30	3.837	0.553	2.140	5.000	> 0.1000	Yes	14.40%			
86\"Non	30	3.586	0.665	2.290	4.430	0.0001	No	18.54%			
W/S	30	3.544	0.654	2.000	4.570	> 0.1000	Yes	18.44%			
B1Ap1	Homogeneity of Variance					Roy's Greatest Root:		F(3,27)		1.97	Pr > F
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	2.1932	0.1368			Yes	29	0.03	0.8720	Yes	Neither	Wand 0.52%
Glove vs 86\"Non			8.8800	0.0042	No	29	4.21	0.0494	No	Var/Means	Glove 7.56%
Glove vs W/S	5.5148	0.0198			No	29	5.38	0.6300	Yes	Var	Neither
Wand vs 86\"Non			1.0100	0.3183	Yes	29	3.11	0.0882	No	Means	Wand 7.00%
Wand vs W/S	0.7989	0.3714			Yes	29	4.17	0.0503	No	Means	Wand 8.27%
86\"Non vs W/S			0.0100	0.9156	Yes	29	0.20	0.6566	Yes	Neither	Neither

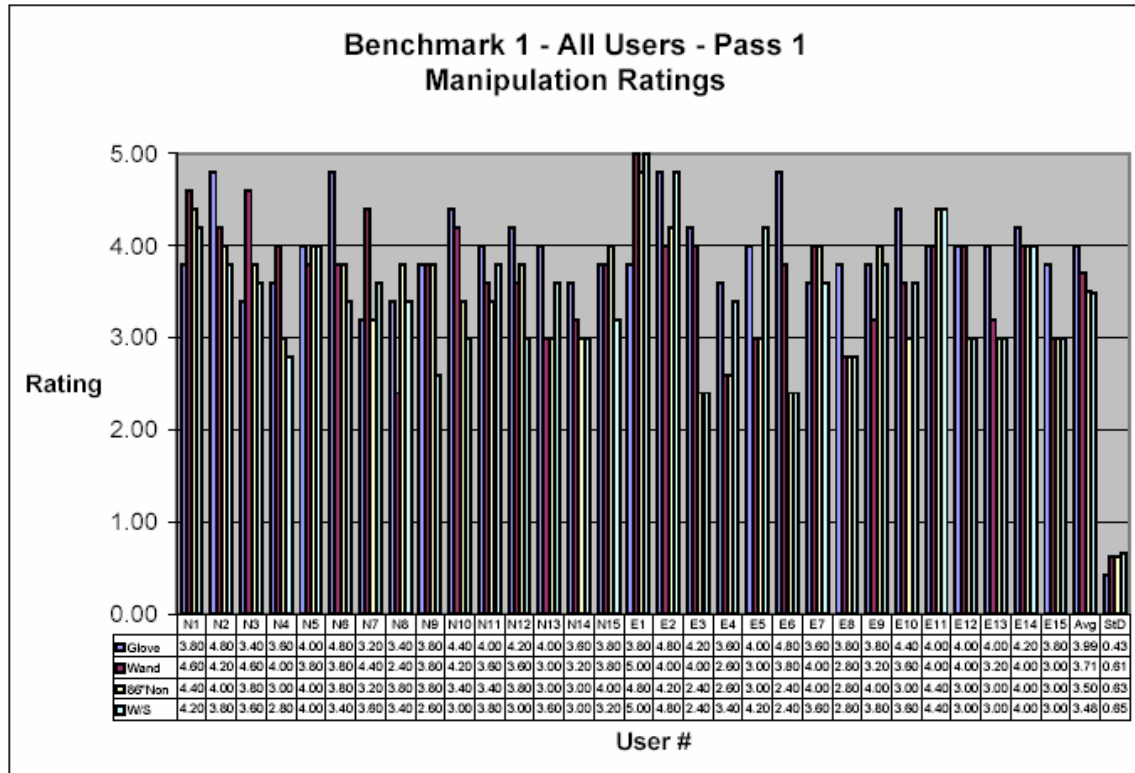




**Figure C- 3: B1ALLp1Loc All Users Locating Ratings**

**Table C- 3: B1ALLp1Loc All Users Locating Ratings Statistics**

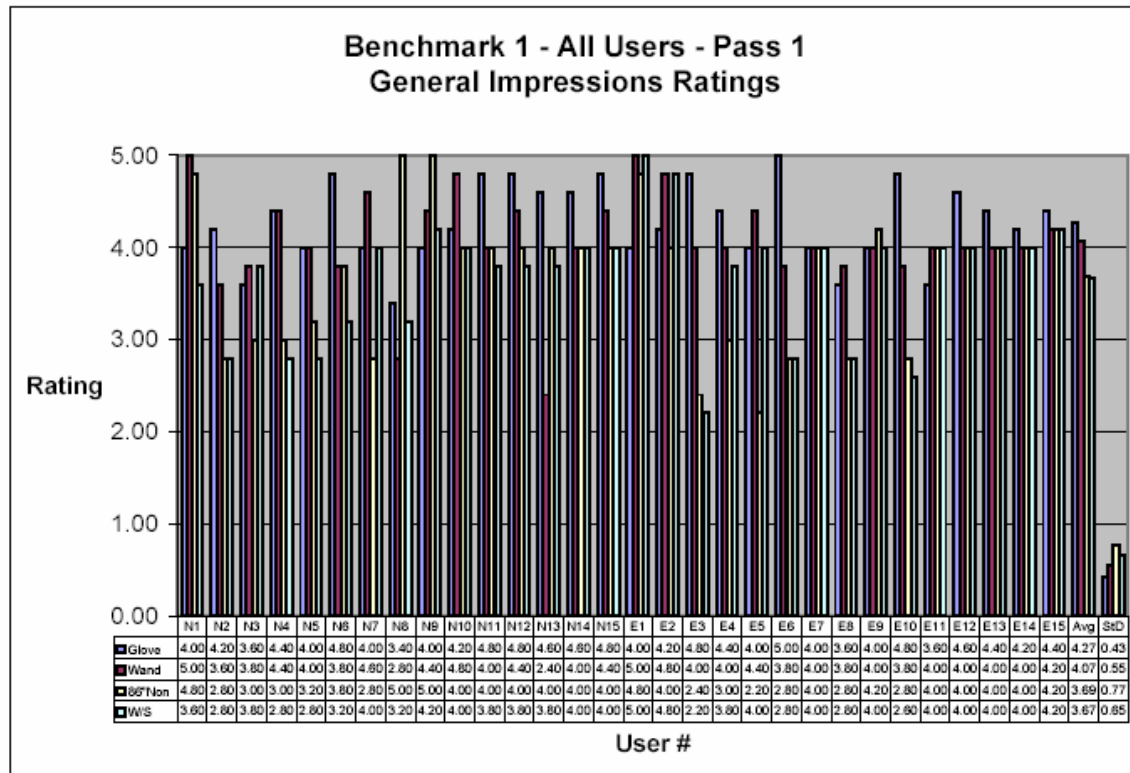
Benchmark 1, All Users, Pass 1, Locating Ratings										
B1Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	30	4.207	0.462	3.40	5.00	0.0249	No	10.99%		
Wand	30	4.053	0.467	2.40	5.00	< 0.0001	No	11.51%		
86\"Non	30	3.760	0.631	2.40	4.80	0.0005	No	16.78%		
W/S	30	3.760	0.593	2.00	4.80	0.0476	No	15.77%		
B1Ep1	Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	3.44	Pr > F	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand			0.0000	0.9716	Yes	14	1.38	0.2502	Yes	Neither
Glove vs 86\"Non			3.3900	0.0708	Yes	14	7.39	0.0110	No	Means
Glove vs W/S			1.2300	0.2715	Yes	14	8.76	0.0061	No	Means
Wand vs 86\"Non			1.7400	0.1926	Yes	14	5.71	0.0235	No	Means
Wand vs W/S			0.7400	0.3926	Yes	14	6.09	0.0198	No	Means
86\"Non vs W/S			0.1000	0.7509	Yes	14	0.00	1.0000	Yes	Neither



**Figure C- 4: B1ALLp1Mov All Users Manipulation Ratings**

**Table C- 4: B1ALLp1Mov All Users Manipulation Ratings Statistics**

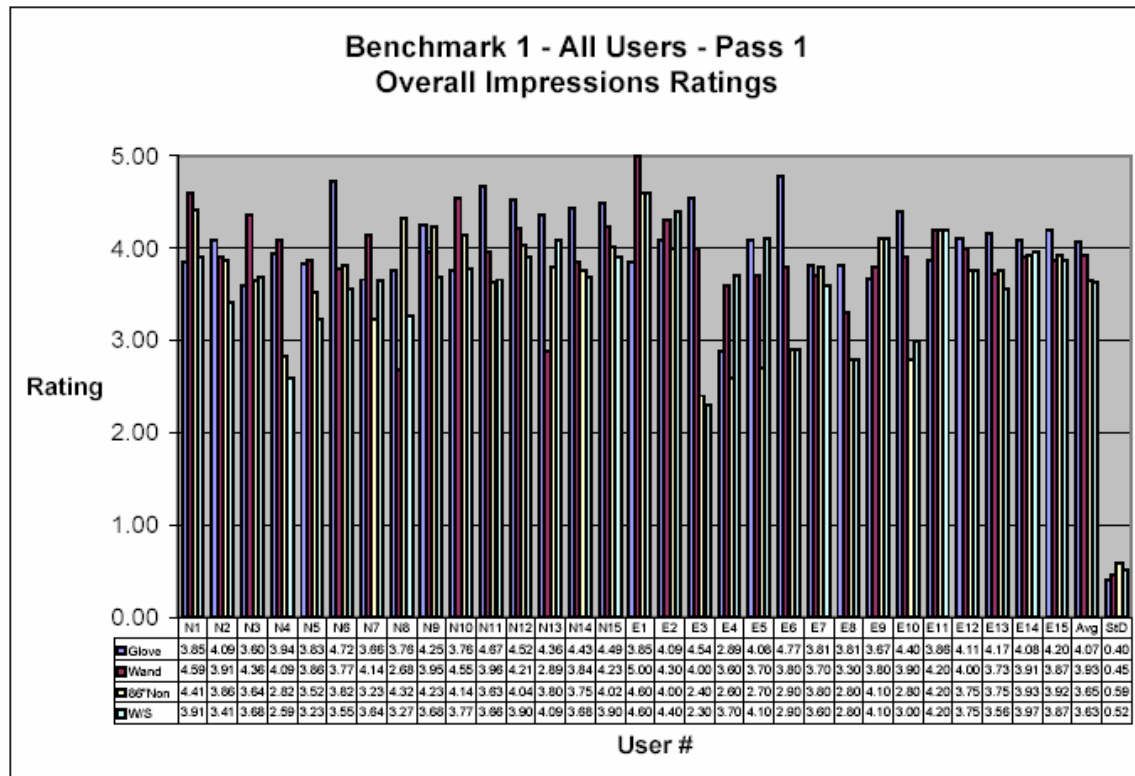
Benchmark 1, All Users, Pass 1, Manipulation Ratings											
B1Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	30	3.987	0.427	3.20	4.80	0.0086	No	10.70%			
Wand	30	3.707	0.614	2.40	5.00	0.0437	No	16.57%			
86\"Non	30	3.500	0.632	2.40	4.80	0.0098	No	18.05%			
W/S	30	3.480	0.653	2.40	5.00	> 0.1000	Yes	18.77%			
B1Ap1	Homogeneity of Variance					Roy's Greatest Root:			F(3,27)		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant	Pr > F
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			3.8200	0.0556	Yes	29	4.46	0.0357	No	Means	Wand 7.55%
Glove vs 86\"Non			6.1100	0.0164	No	29	12.23	0.0015	No	Var/Mean	Neither 13.91%
Glove vs W/S			4.6800	0.0346	No	29	13.12	0.0011	No	Var/Mean	W/S 14.57%
Wand vs 86\"Non			0.0400	0.8513	Yes	29	3.24	0.0824	No	Means	Wand 5.91%
Wand vs W/S			0.1300	0.7207	Yes	29	2.78	0.1060	Yes	Neither	Wand 6.52%
86\"Non vs W/S			0.0400	0.8335	Yes	29	0.05	0.8260	Yes	Neither	Neither



**Figure C- 5: B1ALLp1Gen All Users General Impressions Ratings**

**Table C- 5: B1ALLp1Gen All Users General Impressions Ratings Statistics**

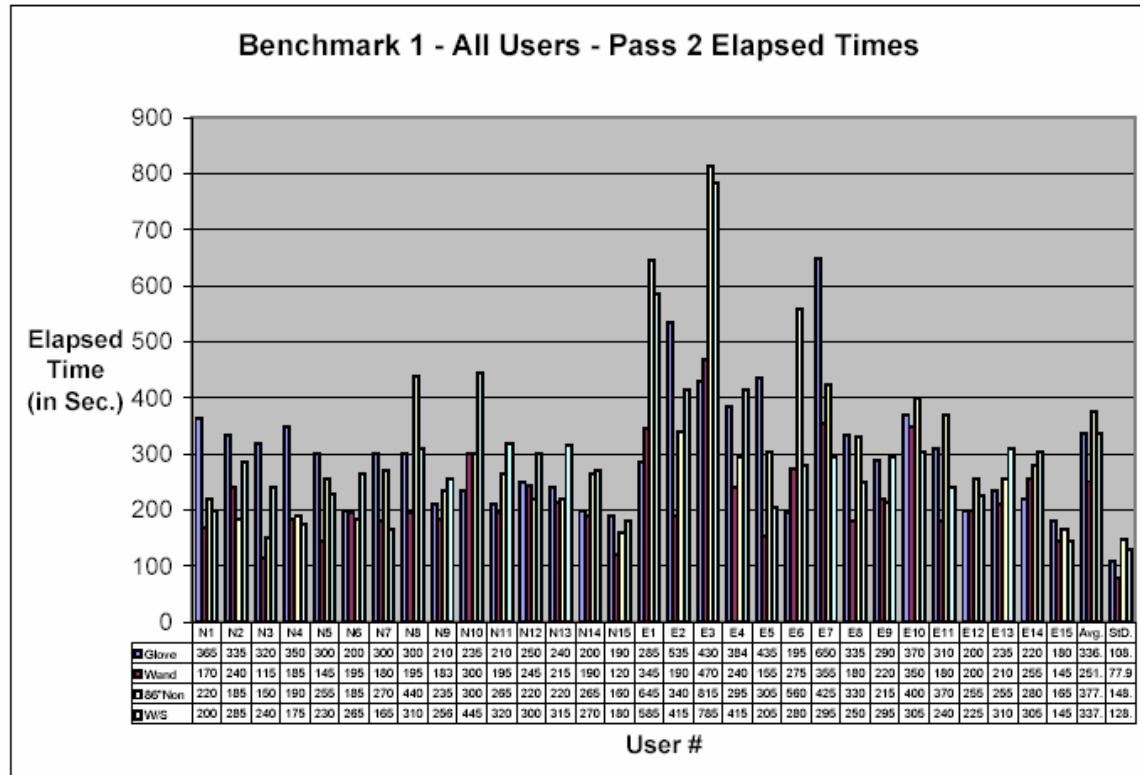
Benchmark 1, All Users, Pass 1, General Impressions Ratings											
B1Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	30	4.273	0.428	3.40	5.00	> 0.1000	Yes	10.02%			
Wand	30	4.073	0.547	2.40	5.00	0.0018	No	13.43%			
86*Non	30	3.687	0.768	2.20	5.00	< 0.0001	No	20.82%			
W/S	30	3.667	0.655	2.20	5.00	< 0.0001	No	17.85%			
B1Ap1		Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	5.41	Pr > F	0.0048
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant
Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.9700	0.3294	Yes	29	2.46	0.1278	Yes	Neither	Neither
Glove vs 86*Non			11.1200	0.0015	No	29	11.59	0.0020	No	Var/Mean	Glove 15.89%
Glove vs W/S			4.9500	0.0301	No	29	15.25	0.0005	No	Var/Mean	Glove 16.53%
Wand vs 86*Non			3.2300	0.0774	Yes	29	5.53	0.0257	No	Means	Wand 10.47%
Wand vs W/S			0.7100	0.4014	Yes	29	11.42	0.0021	No	Means	Wand 11.07%
86*Non vs W/S			1.0700	0.3063	Yes	29	0.03	0.8699	Yes	Neither	Neither



**Figure C- 6: B1ALLp1Ovr All Users Pass 1 Overall Impressions Ratings**

**Table C- 6: B1ALLp1Ovr All Users Pass 1 Overall Impressions Ratings Statistics**

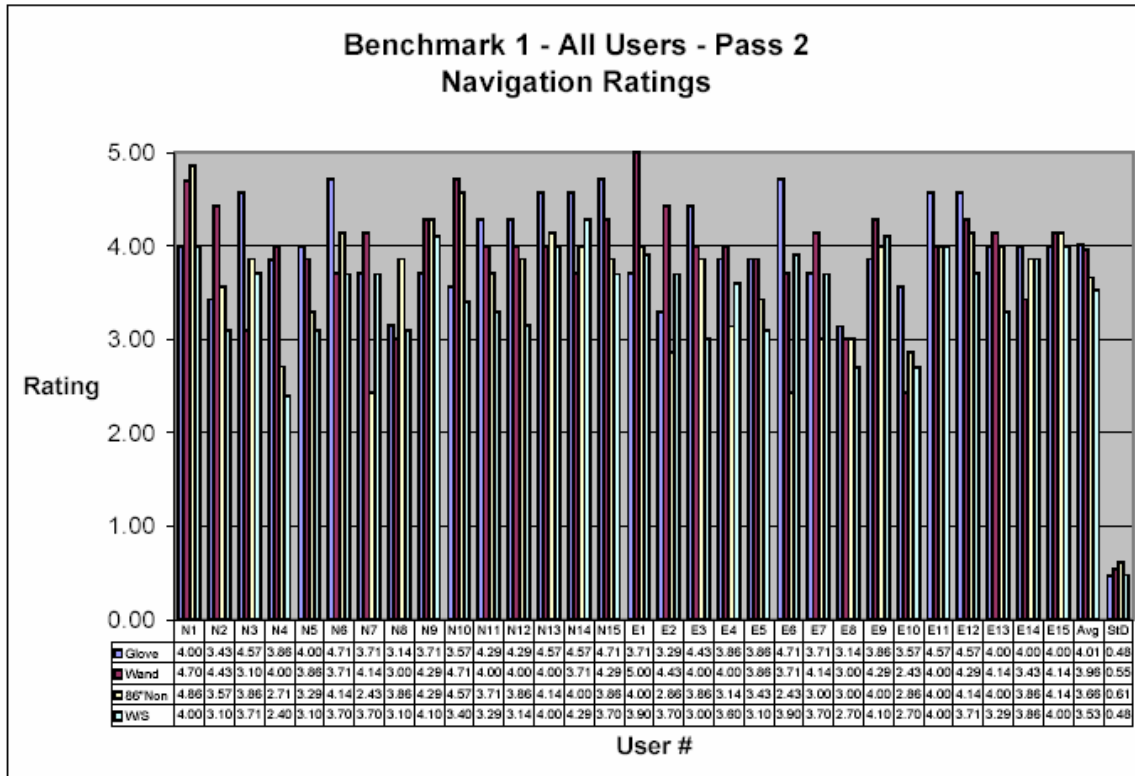
Benchmark 1, All Users, Pass 1, Overall Impressions Ratings											
B1Ap1	# Users	Mean	St. Dev.		Low	High	P Value	Normal?	Coeff. of Var.		
Glove	30	4.075	0.404		2.68	4.77	> 0.1000	Yes	9.92%		
Wand	30	3.928	0.454		2.40	5.00	0.0203	No	11.56%		
86"Non	30	3.649	0.593		2.30	4.60	0.0033	No	16.24%		
W/S	30	3.627	0.518		2.30	4.60	0.0207	No	14.28%		
B1Ap1	Homogeneity of Variance					Roy's Greatest Root:		F(3,27)	4.35	Pr > F 0.0126	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.2300	0.6302	Yes	29	1.72	0.1997	Yes	Neither	Neither
Glove vs 86"Non			4.3000	0.0426	No	29	10.55	0.0029	No	Var/Means	Glove 11.67%
Glove vs W/S			1.3100	0.2565	Yes	29	11.88	0.0018	No	Means	Glove 12.35%
Wand vs 86"Non			1.9100	0.1721	Yes	29	5.64	0.0244	No	Means	Wand 7.65%
Wand vs W/S			0.3400	0.5597	Yes	29	8.31	0.0074	No	Means	Wand 8.30%
86"Non vs W/S			0.6000	0.4424	Yes	29	0.01	0.7889	Yes	Neither	Neither



**Figure C- 7: B1ALLp2Tim All Users Elapsed Times**

**Table C- 7: B1ALLp2Tim All Users Elapsed Times Statistics**

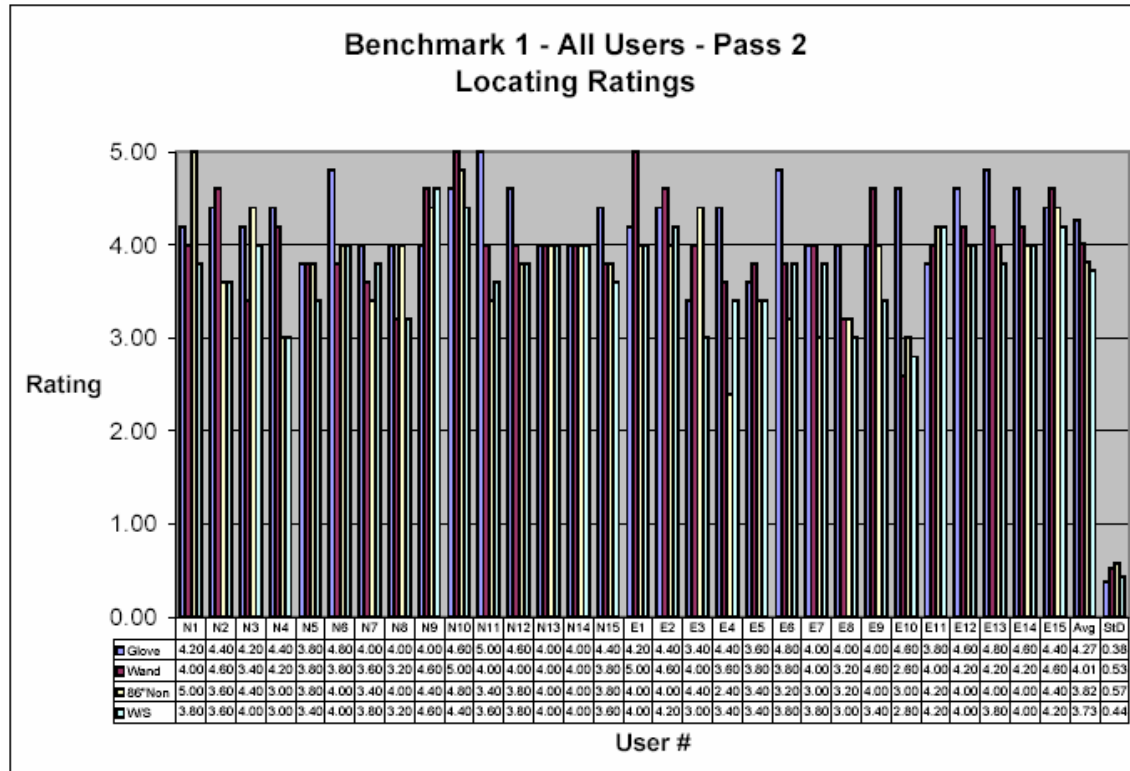
Benchmark 1, All Users, Pass 2, Elapsed Times									
B1Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.	
Glove	30	302.0	108.20	180.0	650.0	> 0.1000	Yes	35.83%	
Wand	30	221.4	77.87	115.0	470.0	0.0196	No	35.17%	
86°Non	30	307.4	148.80	150.0	815.0	0.0023	No	48.41%	
W/S	30	300.4	128.80	145.0	785.0	< 0.0001	No	42.88%	
B1Ap2		Homogeneity of Variance			Roy's Greatest Root:			F(3,27) 17.59	
		Bartlett's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant
		Chi-Sq	Pr > ChiSq						
Glove vs Wand				Yes	29	16.21	0.0004	No	Means
Glove vs 86°Non				Yes	29	0.03	0.8533	Yes	Neither
Glove vs W/S				Yes	29	0.00	0.9524	Yes	Neither
Wand vs 86°Non				Yes	29	21.87	0.0001	No	Means
Wand vs W/S				Yes	29	27.69	0.0001	No	Means
86°Non vs W/S				Yes	28	0.15	0.7003	Yes	Neither



**Figure C- 8: B1ALLp2Nav All Users Navigation Ratings**

**Table C- 8: B1ALLp2Nav All Users Navigation Ratings Statistics**

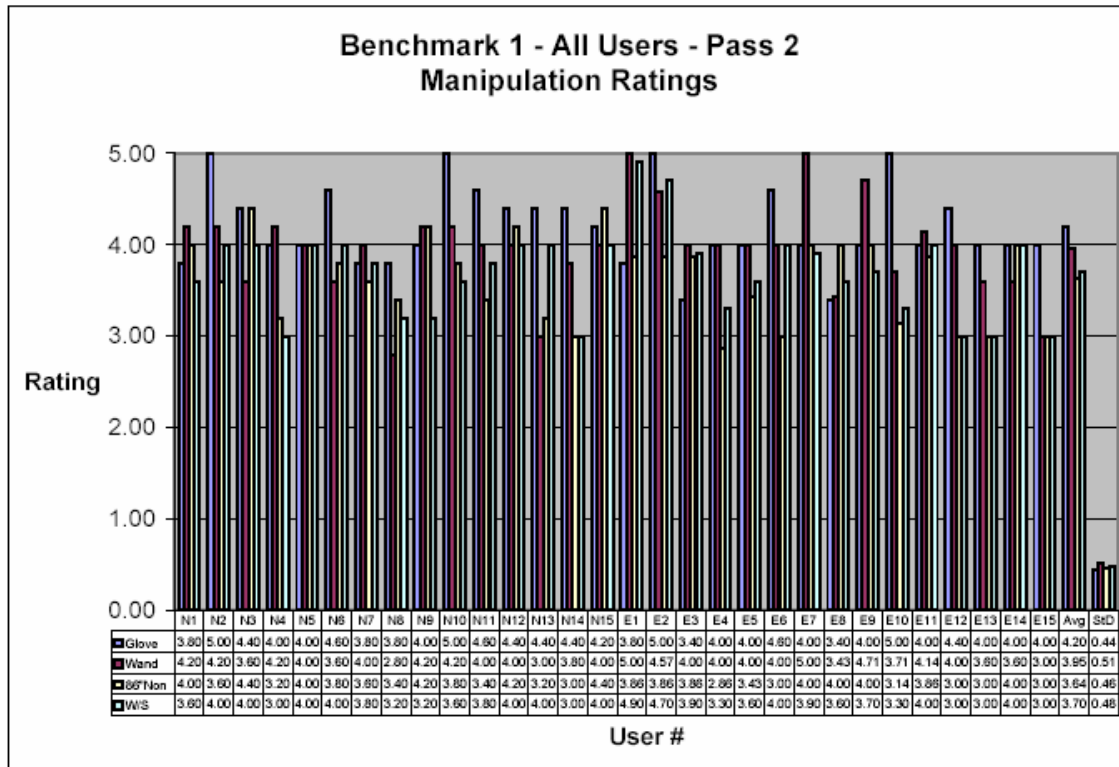
Benchmark 1, All Users, Pass 2, Navigation Ratings											
B1Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	30	4.014	0.476	3.140	4.710	> 0.1000	Yes	11.86%			
Wand	30	3.960	0.545	2.430	5.000	0.0047	No	13.77%			
86°Non	30	3.662	0.615	2.430	4.860	0.0004	No	16.78%			
W/S	30	3.533	0.480	2.400	4.290	0.0028	No	13.57%			
B1Ap2	Homogeneity of Variance				Roy's Greatest Root:		F(3,27)		11.99	Pr > F <0.0001	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.4600	0.4988	Yes	29	0.17	0.6805	Yes	Neither	Neither
Glove vs 86°Non			2.5100	0.1184	Yes	29	8.23	0.0076	No	Means	Glove 9.61%
Glove vs W/S			0.0000	0.9631	Yes	29	25.27	0.0001	No	Means	W/S 13.61%
Wand vs 86°Non			0.4000	0.5302	Yes	29	6.03	0.0203	No	Means	Wand 8.14%
Wand vs W/S			0.3900	0.5339	Yes	29	16.96	0.0003	No	Means	Wand 12.09%
86°Non vs W/S			2.2000	0.1435	Yes	29	1.39	0.0203	No	Means	W/S 3.65%



**Figure C- 9: B1ALLp2Loc All Users Locating Ratings**

**Table C- 9: B1ALLp2Loc All Users Locating Ratings Statistics**

Benchmark 1, All Users, Pass 2, Locating Ratings												
B1Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.				
Glove	30	4.267	0.384	3.40	5.00	0.0587	Yes	8.99%				
Wand	30	4.013	0.528	2.60	5.00	> 0.1000	Yes	13.15%				
86\"Non	30	3.820	0.574	2.40	5.00	0.0074	No	15.02%				
W/S	30	3.727	0.441	2.80	4.60	0.0341	No	11.83%				
B1Ap2	Homogeneity of Variance					Roy's Greatest Root:			F(3,27)		12.16	Pr > F <0.0001
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand	2.8523	0.0912			Yes	14	4.92	0.0346	No	Means	Wand	6.33%
Glove vs 86\"Non			3.8500	0.0546	Yes	14	11.14	0.0023	No	Means	Glove	11.70%
Glove vs W/S			0.7300	0.3971	Yes	14	31.66	0.0001	No	Means	W/S	14.49%
Wand vs 86\"Non			0.1800	0.6710	Yes	14	3.35	0.0777	No	Means	Wand	5.05%
Wand vs W/S			0.8200	0.3694	Yes	14	12.82	0.0012	No	Means	Wand	7.67%
86\"Non vs W/S			1.9100	0.1722	Yes	14	1.07	0.3100	Yes	Neither	Neither	

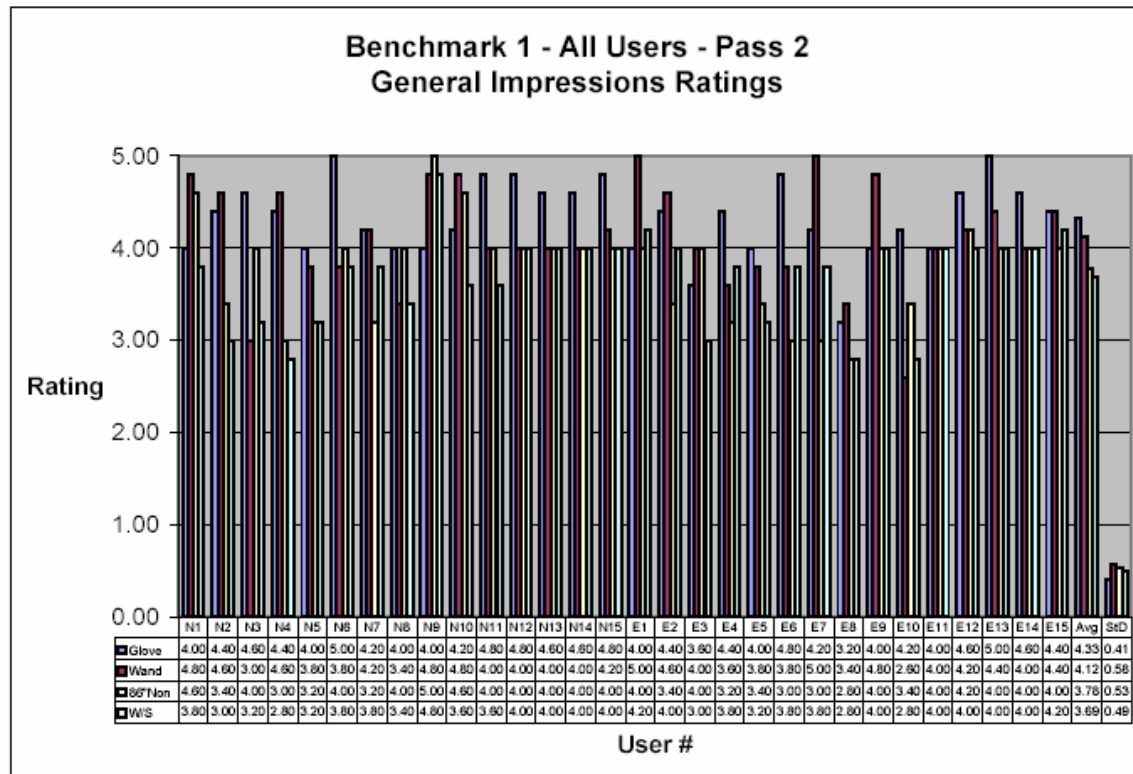


**Figure C- 10: B1ALLp2Mov All Users Manipulation Ratings**

**Table C- 10: B1ALLp2Mov All Users Manipulation Ratings Statistics**

Benchmark 1, All Users, Pass 2, Manipulation Ratings										
B1Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	30	4.200	0.439	3.40	5.00	< 0.0001	No	10.46%		
Wand	30	3.952	0.509	2.80	5.00	0.0025	No	12.87%		
86\"Non	30	3.636	0.465	2.86	4.40	0.0245	No	12.78%		
W/S	30	3.703	0.482	3.00	4.90	0.0029	No	13.03%		
B1Ap2	Homogeneity of Variance				Roy's Greatest Root:			F(3,27) 7.9 Pr > F <0.0001		
	Bartlett's Test		Levene's Test		df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand			0.5700	0.4538	Yes	29	4.37	0.0455	No	Means Wand 6.28%
Glove vs 86\"Non			0.1600	0.6905	Yes	29	20.04	0.0001	No	Means Glove 15.51%
Glove vs W/S			0.2800	0.6010	Yes	29	20.19	0.0001	No	Means W/S 13.42%
Wand vs 86\"Non			0.2700	0.6057	Yes	29	9.46	0.0045	No	Means Wand 8.69%
Wand vs W/S			0.2700	0.6057	Yes	29	6.56	0.0159	No	Means Wand 6.72%
86\"Non vs W/S			0.0600	0.8112	Yes	29	0.68	0.4164	Yes	Neither Neither

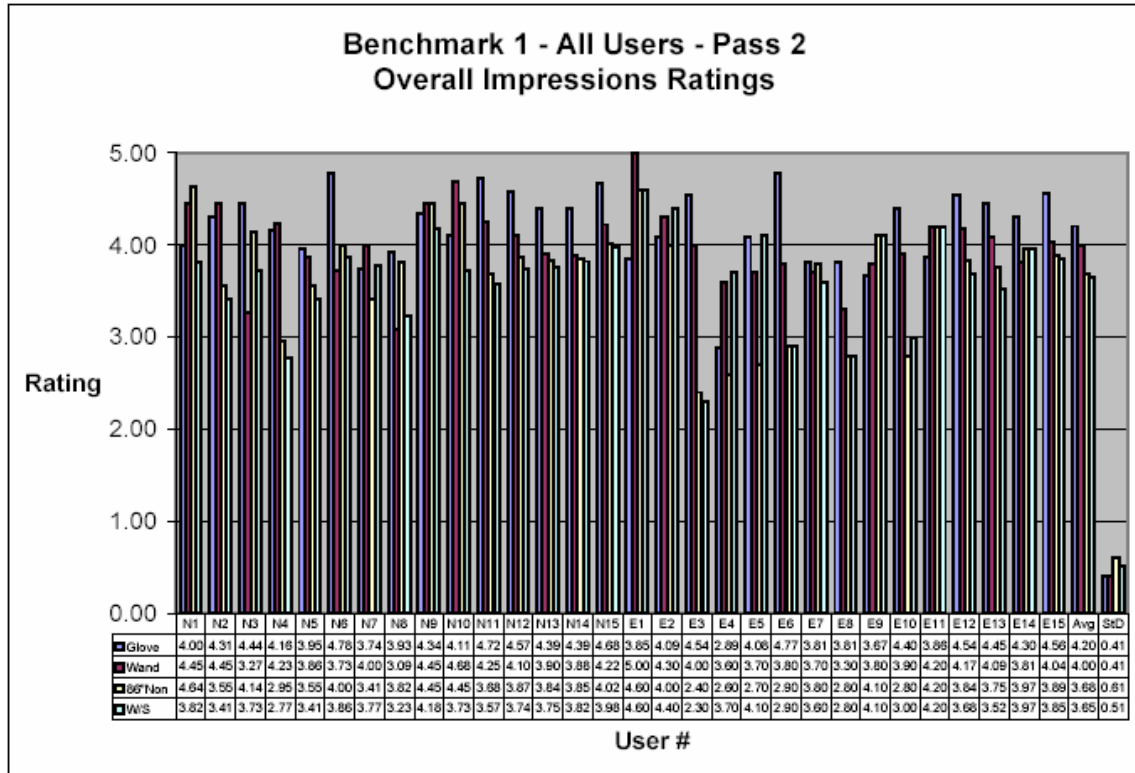




**Figure C- 11: B1ALLp2Gen All Users General Impressions Ratings**

**Table C- 11: B1ALLp2Gen All Users General Impressions Ratings Statistics**

Benchmark 1, All Users, Pass 2, General Impressions Ratings											
B1Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	30	4.327	0.412	3.20	5.00	0.096	Yes	9.52%			
Wand	30	4.120	0.577	2.60	5.00	> 0.1000	Yes	14.00%			
86°Non	30	3.780	0.529	2.80	5.00	< 0.0001	No	13.99%			
W/S	30	3.687	0.489	2.80	4.80	0.0005	No	13.26%			
B1Ap2	Homogeneity of Variance				Roy's Greatest Root:			F(3,27)	18.80	Pr > F	<0.0001
	Chi-Sq	Pr > ChiSq	F Value	Pr > F	Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
Glove vs Wand	3.1755	0.0747			Yes	29	2.50	0.1244	Yes	Neither	Neither
Glove vs 86°Non			1.8700	0.1771	Yes	29	23.06	0.0001	No	Means	Glove 14.47%
Glove vs W/S			0.8800	0.3518	Yes	29	45.27	0.0001	No	Means	Glove 17.36%
Wand vs 86°Non			0.2300	0.6362	Yes	29	7.85	0.0090	No	Means	Wand 8.99%
Wand vs W/S			0.7600	0.3861	Yes	29	19.25	0.0001	No	Means	Wand 11.74%
86°Non vs W/S			0.2200	0.6420	Yes	29	1.13	0.2963	Yes	Neither	Neither



**Figure C- 12: B1ALLp2Ovr All Users Overall Impressions Ratings**

**Table C- 12: B1ALLp2Ovr All Users Overall Impressions Ratings Statistics**

Benchmark 1, All Users, Pass 2, Overall Impressions Ratings												
B1Ap2	# Users	Mean	St. Dev.		Low	High	P Value	Normal?	Coeff. of Var.			
Glove	30	4.204	0.411		2.89	4.78	> 0.1000	Yes	9.77%			
Wand	30	3.999	0.409		3.09	5.00	> 0.1000	Yes	10.22%			
86"Non	30	3.686	0.610		2.40	4.64	0.0193	No	16.54%			
W/S	30	3.650	0.508		2.30	4.60	0.0567	Yes	13.91%			
B1Ap2	Homogeneity of Variance					Roy's Greatest Root:			F(3,27)	6.56	Pr > F	0.0018
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand	0.0007	0.9784			Yes	29	4.37	0.0453	No	Means	Glove	5.13%
Glove vs 86"Non			4.0200	0.0495	No	29	16.55	0.0003	No	Var/Mean	Glove	14.05%
Glove vs W/S	1.2687	0.2600			Yes	29	18.52	0.0002	No	Means	Glove	15.18%
Wand vs 86"Non			4.6300	0.0355	No	29	9.21	0.0050	No	Var/Mean	Wand	8.49%
Wand vs W/S	1.3298	0.2488			Yes	29	13.39	0.0010	No	Means	Wand	9.56%
86"Non vs W/S			1.0500	0.3097	Yes	29	0.20	0.6541	Yes	Neither	Neither	

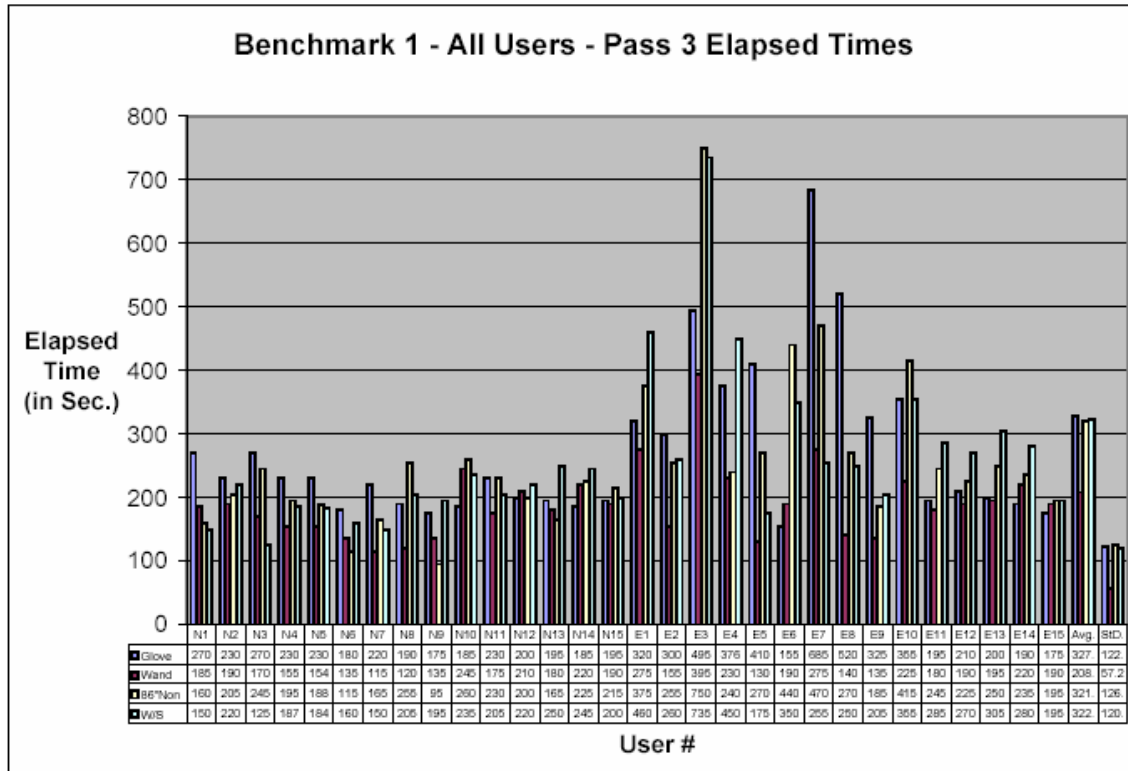
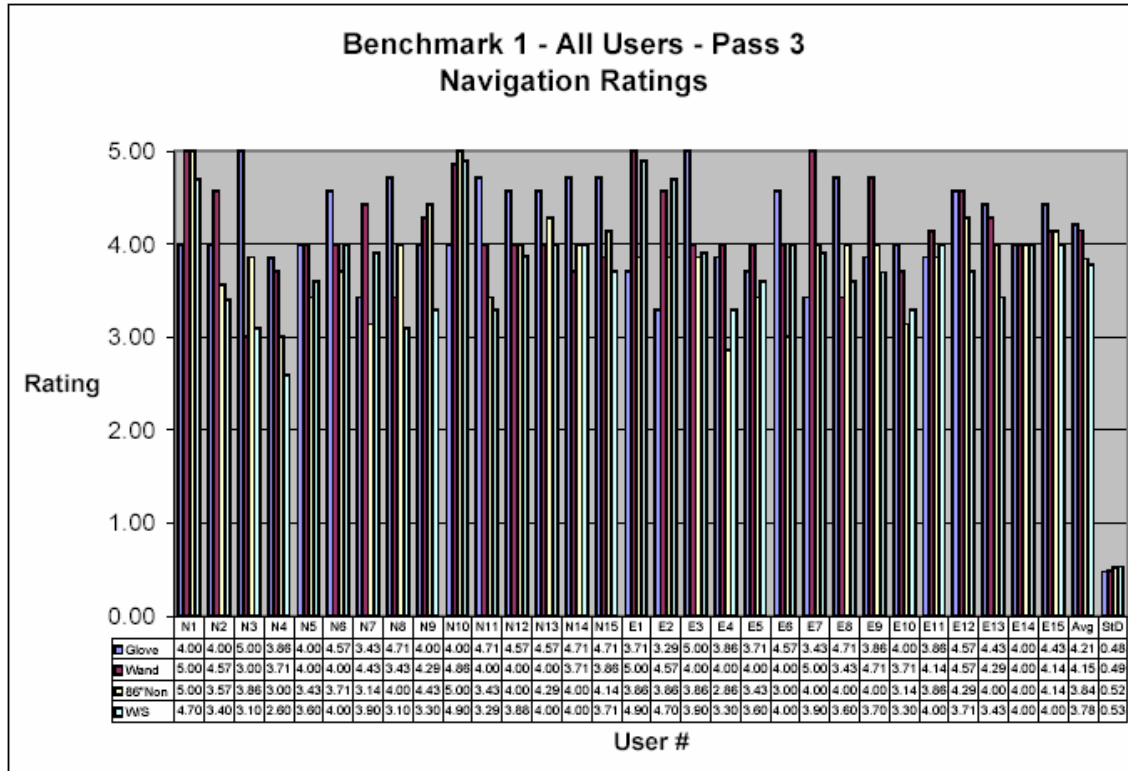


Figure C- 13: B1ALLp3Tim All Users Elapsed Times

Table C- 13: B1ALLp3Tim All Users Elapsed Times Statistics

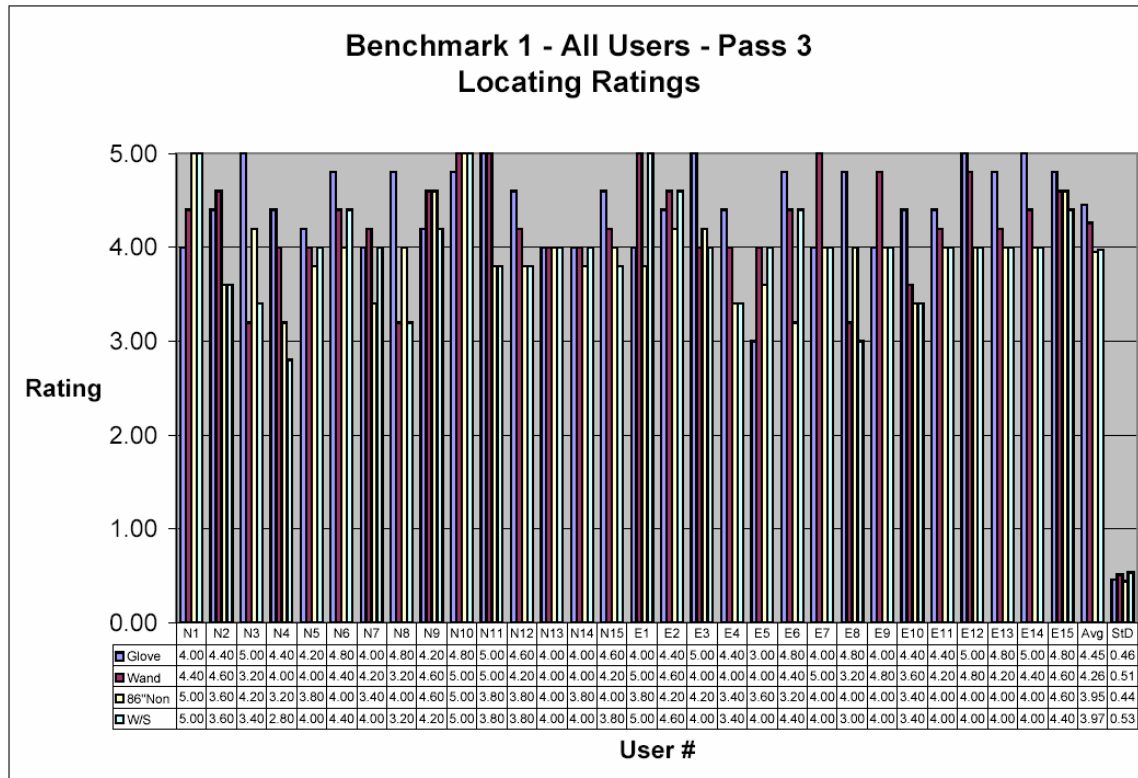
Benchmark 1, All Users, Pass 3, Elapsed Times										
B1Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	30	269.90	122.60	155.00	685.00	< 0.0001	No	45.42%		
Wand	30	190.10	57.19	115.00	395.00	0.0311	No	30.08%		
86*Non	30	257.90	126.40	95.00	750.00	< 0.0001	No	49.01%		
W/S	30	258.70	120.10	125.00	735.00	0.0012	No	46.42%		
B1Ap3		Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	10.89	Pr > F < 0.0001
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					
Glove vs Wand				3.4100	0.0697	Yes	29	15.01	0.0006	No
Glove vs 86*Non				0.0100	0.9278	Yes	29	0.34	0.5659	Yes
Glove vs W/S				0.0000	0.9519	Yes	29	0.20	0.6552	Yes
Wand vs 86*Non				2.2400	0.1400	Yes	29	17.84	0.0002	No
Wand vs W/S				1.9700	0.1661	Yes	29	21.67	0.0001	No
86*Non vs W/S				0.0200	0.8926	Yes	28	0.00	0.9563	Yes
										Statistically Significant
										Statistically Better
										Wand 41.98%
										Neither
										Neither
										Wand 35.67%
										Wand 36.09%
										Neither



**Figure C- 14: B1ALLp3Nav All Users Navigation Ratings**

**Table C- 14: B1ALLp3Nav All Users Navigation Ratings Statistics**

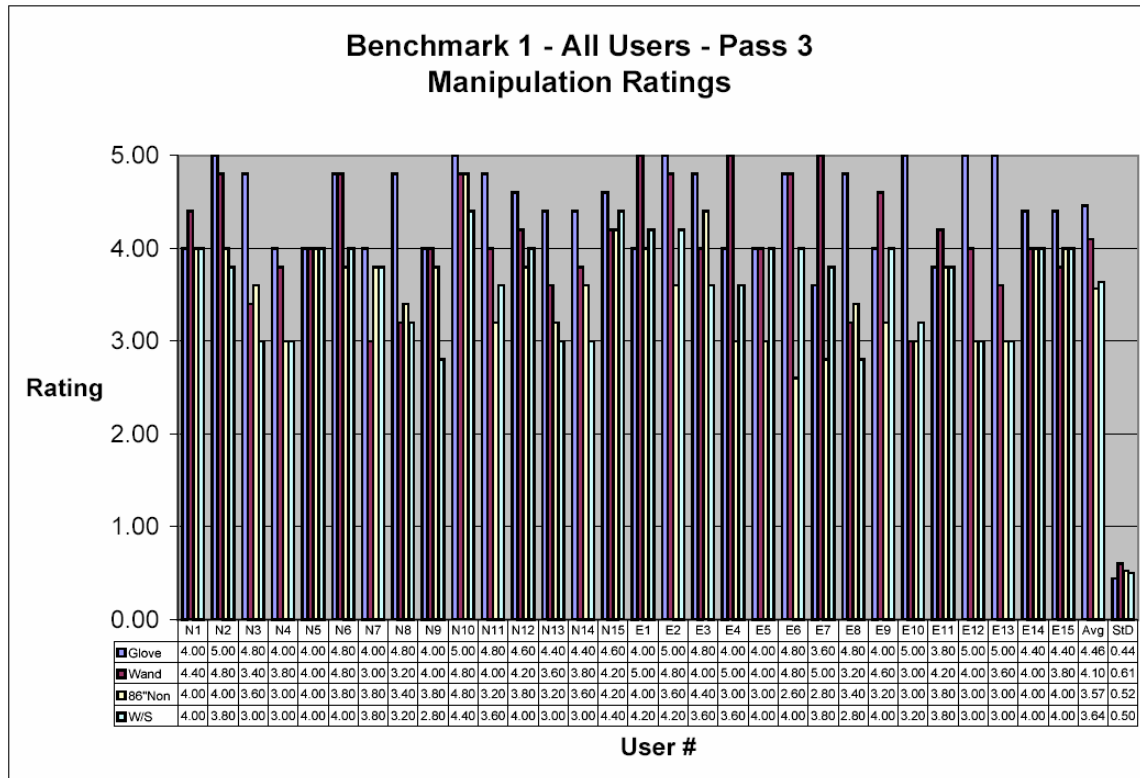
Benchmark 1, All Users, Pass 3, Navigation Ratings														
B1Ap3	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	Coeff. of Var.		
Glove	30	4.209		0.477		3.29		5.00		0.0029	No	11.34%		
Wand	30	4.147		0.487		3.00		5.00		0.0099	No	11.75%		
86"Non	30	3.843		0.518		2.86		5.00		0.0148	No	13.47%		
W/S	30	3.784		0.533		2.60		4.90		0.0017	No	14.07%		
B1Ap3		Homogeneity of Variance				Roy's Greatest Root:				F(3,27)		10.50	Pr > F	<0.0001
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better		
		Chi-Sq	Pr > ChiSq	F Value	Pr > F									
Glove vs Wand				0.0200	0.8954	Yes	29	0.15	0.6992	Yes	Neither	Neither		
Glove vs 86"Non				0.2300	0.6310	Yes	29	9.66	0.0042	No	Means	86" Non	9.52%	
Glove vs W/S				0.4000	0.5296	Yes	29	8.39	0.0071	No	Means	W/S	11.23%	
Wand vs 86"Non				0.1100	0.7434	Yes	29	8.62	0.0064	No	Means	Wand	7.91%	
Wand vs W/S				0.2200	0.6386	Yes	29	20.77	0.0001	No	Means	Wand	9.59%	
86"Non vs W/S				0.0200	0.8860	Yes	29	0.39	0.5368	Yes	Neither	Neither		



**Figure C- 15: B1ALLp3Loc All Users Locating Ratings**

**Table C- 15: B1ALLp3Loc All Users Locating Ratings Statistics**

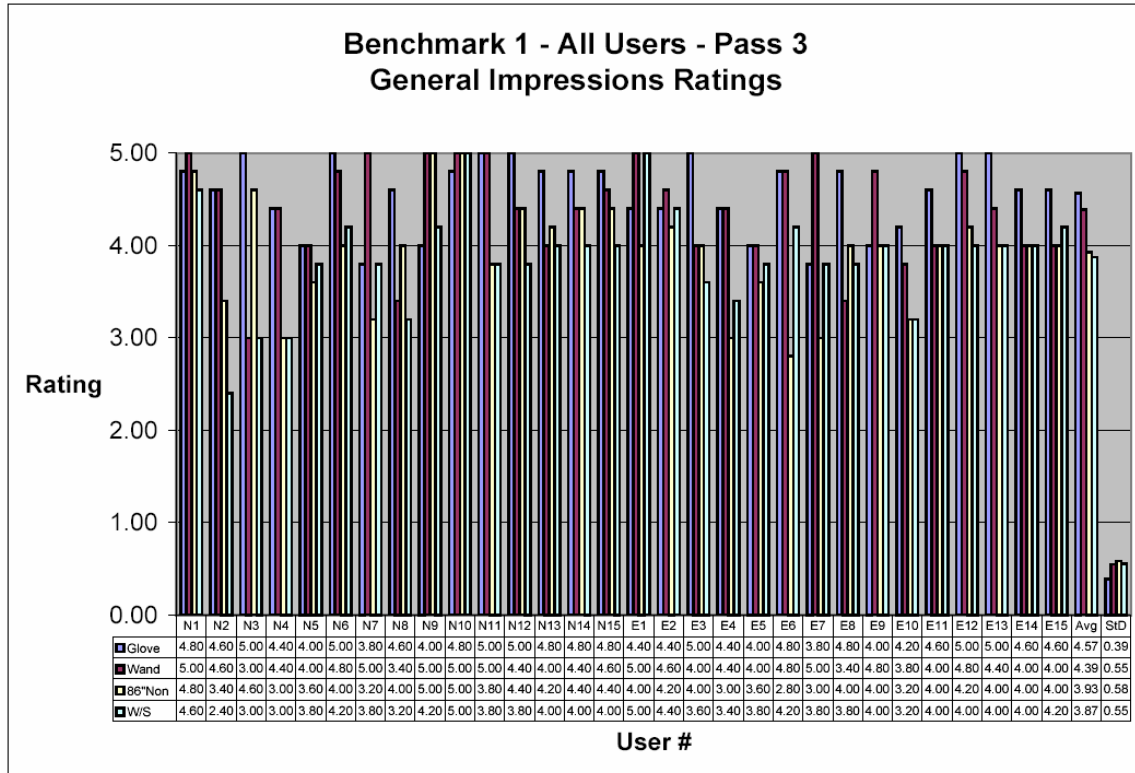
Benchmark 1, All Users, Pass 3, Locating Ratings													
B1Ap3	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	Coeff. of Var.	
Glove	30	4.453		0.458		3.00		5.00		0.019	No	10.28%	
Wand	30	4.260		0.512		3.20		5.00		0.0229	No	12.03%	
86"Non	30	3.953		0.442		3.20		5.00		0.0005	No	11.17%	
W/S	30	3.973		0.535		2.80		5.00		0.0012	No	13.46%	
B1Ap3	Homogeneity of Variance					Roy's Greatest Root:			F(3,27)		6.36	Pr > F	0.0021
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better		
	Chi-Sq	Pr > ChiSq	F Value	Pr > F									
Glove vs Wand			0.3000	0.5853	Yes	14	2.19	0.1496	Yes	Neither	Neither		
Glove vs 86"Non			0.0300	0.8742	Yes	14	22.15	0.0001	No	Means	86" Non	12.65%	
Glove vs W/S			0.5400	0.4659	Yes	14	11.95	0.0017	No	Means	W/S	12.08%	
Wand vs 86"Non			0.6000	0.4412	Yes	14	8.07	0.0081	No	Means	Wand	7.77%	
Wand vs W/S			0.0500	0.8155	Yes	14	13.40	0.0010	No	Means	Wand	7.22%	
86"Non vs W/S			0.9100	0.3447	Yes	14	0.05	0.8188	Yes	Neither	Neither		



**Figure C- 16: B1ALLp3Mov All Users Manipulation Ratings**

**Table C- 16: B1ALLp3Mov All Users Manipulation Ratings Statistics**

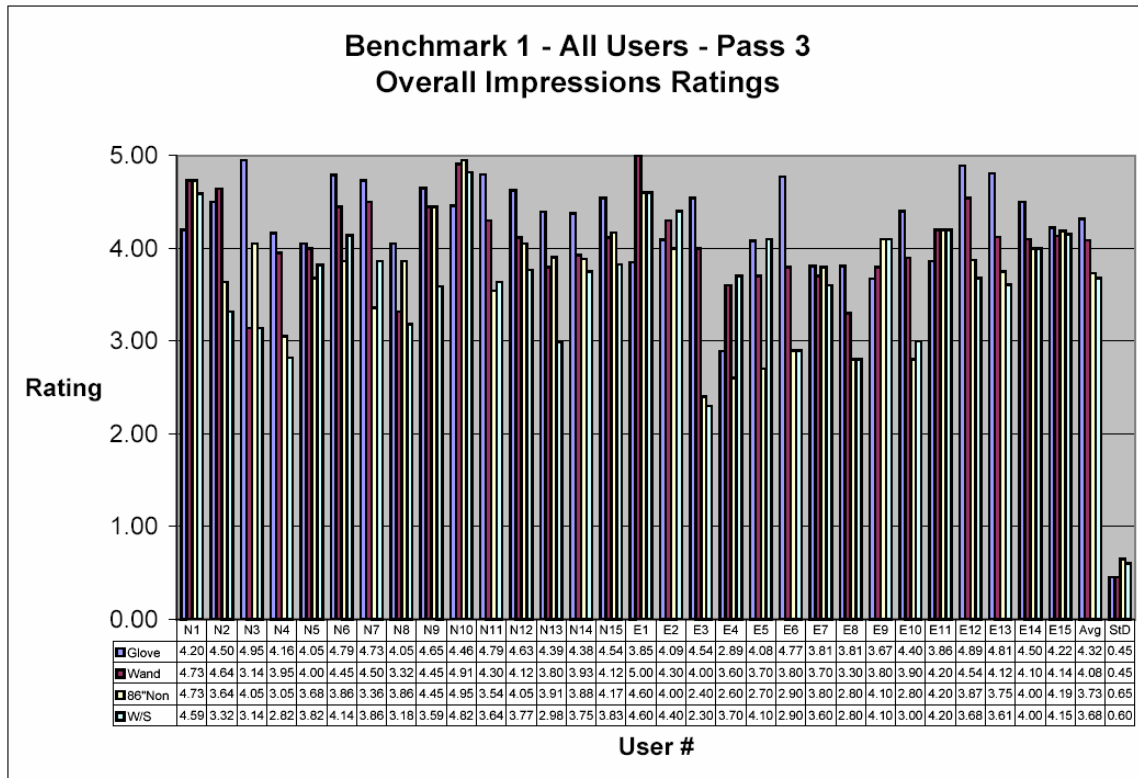
Benchmark 1. All Users, Pass 3, Manipulation Ratings										
B1Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	30	4.460	0.443	3.60	5.00	0.0009	No	9.93%		
Wand	30	4.100	0.605	3.00	5.00	> 0.1000	Yes	14.76%		
86\"Non	30	3.576	0.523	2.60	4.80	> 0.1000	Yes	14.62%		
W/S	30	3.640	0.502	2.80	4.40	0.0045	No	13.79%		
B1Ap3		Homogeneity of Variance				Roy's Greatest Root:			F(3,27)	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					
Glove vs Wand				4.9200	0.0305	No	29	5.91	0.0214	No
Glove vs 86\"Non				1.3200	0.2546	Yes	29	55.37	0.0001	No
Glove vs W/S				1.2800	0.2620	Yes	29	39.45	0.0001	No
Wand vs 86\"Non		0.6070	0.4357			Yes	29	14.99	0.0006	No
Wand vs W/S				1.9500	0.1679	Yes	29	26.10	0.0001	No
86\"Non vs W/S				0.0900	0.7716	Yes	29	0.55	0.4646	Yes
										Statistically Significant
										Statistically Better
										Var/Mean
										Wand 8.78%
										86\"Non 24.72%
										W/S 22.53%
										Wand 14.65%
										Wand 12.64%
										Neither



**Figure C- 17: B1ALLp3Gen All Users General Impressions Ratings**

**Table C- 17: B1ALLp3Gen All Users General Impressions Ratings Statistics**

Benchmark 1, All Users, Pass 3, General Impressions Ratings											
B1Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	30	4.567	0.390	3.80	5.00	0.0063	No	8.54%			
Wand	30	4.387	0.548	3.00	5.00	> 0.1000	Yes	12.50%			
86*Non	30	3.927	0.584	2.80	5.00	0.0009	No	14.86%			
W/S	30	3.873	0.552	2.40	5.00	0.0012	No	14.26%			
B1Ap3	Homogeneity of Variance				Roy's Greatest Root:				F(3,27)		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			3.4500	0.0682	Yes	29	1.84	0.1853	Yes	Neither	Neither
Glove vs 86*Non			5.1900	0.0265	Yes	29	39.63	0.0001	No	Means	Glove 16.30%
Glove vs W/S			2.3400	0.1314	Yes	29	33.35	0.0001	No	Means	Glove 17.92%
Wand vs 86*Non			0.1400	0.7086	Yes	29	9.83	0.0039	No	Means	Wand 11.71%
Wand vs W/S			0.0000	0.9718	Yes	29	25.66	0.0001	No	Means	Wand 13.27%
86*Non vs W/S			0.0900	0.7708	Yes	29	0.25	0.6239	Yes	Neither	Neither

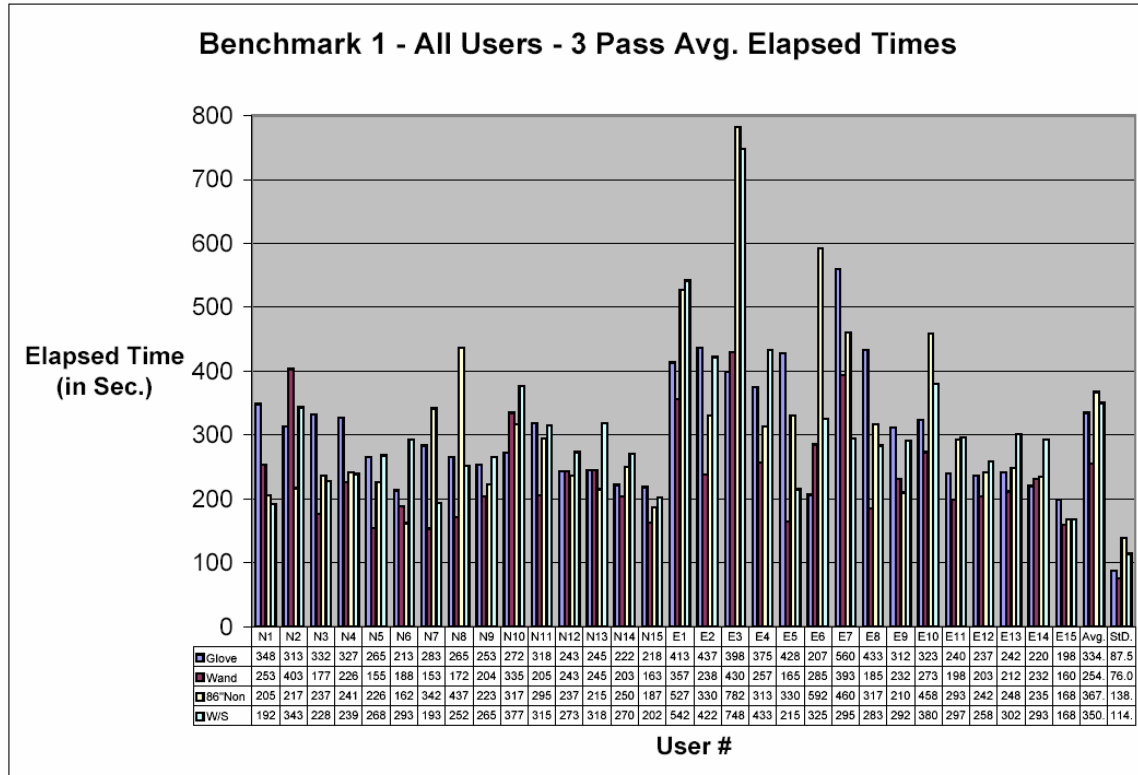


**Figure C- 18: B1ALLp3Ovr All Users Overall Impressions Ratings**

**Table C- 18: B1ALLp3Ovr All Users Overall Impressions Ratings Statistics**

Benchmark 1, All Users, Pass 3, Overall Impressions Ratings											
B1Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	30	4.315	0.451	2.89	4.95	> 0.1000	Yes	10.45%			
Wand	30	4.085	0.453	3.14	5.00	> 0.1000	Yes	11.10%			
86\"Non	30	3.730	0.648	2.40	4.59	> 0.1000	Yes	17.36%			
W/S	30	3.680	0.599	2.30	4.82	0.0998	Yes	16.28%			
B1Ap3	Homogeneity of Variance				Roy's Greatest Root:			F(3,27)		7.77	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.8466	0.3375			Yes	29	5.19	0.0303	No	Means	Glove 5.63%
Glove vs 86\"Non	1.2766	0.2585			Yes	29	19.36	0.0001	No	Means	Glove 15.68%
Glove vs W/S	0.6266	0.4286			Yes	29	18.50	0.0002	No	Means	Glove 17.26%
Wand vs 86\"Non	2.1347	0.1440			Yes	29	11.88	0.0018	No	Means	Wand 9.52%
Wand vs W/S	2.6999	0.1004			Yes	29	20.56	0.0001	No	Means	Wand 11.01%
86\"Non vs W/S	0.1183	0.7309			Yes	29	0.31	0.5831	Yes	Neither	Neither

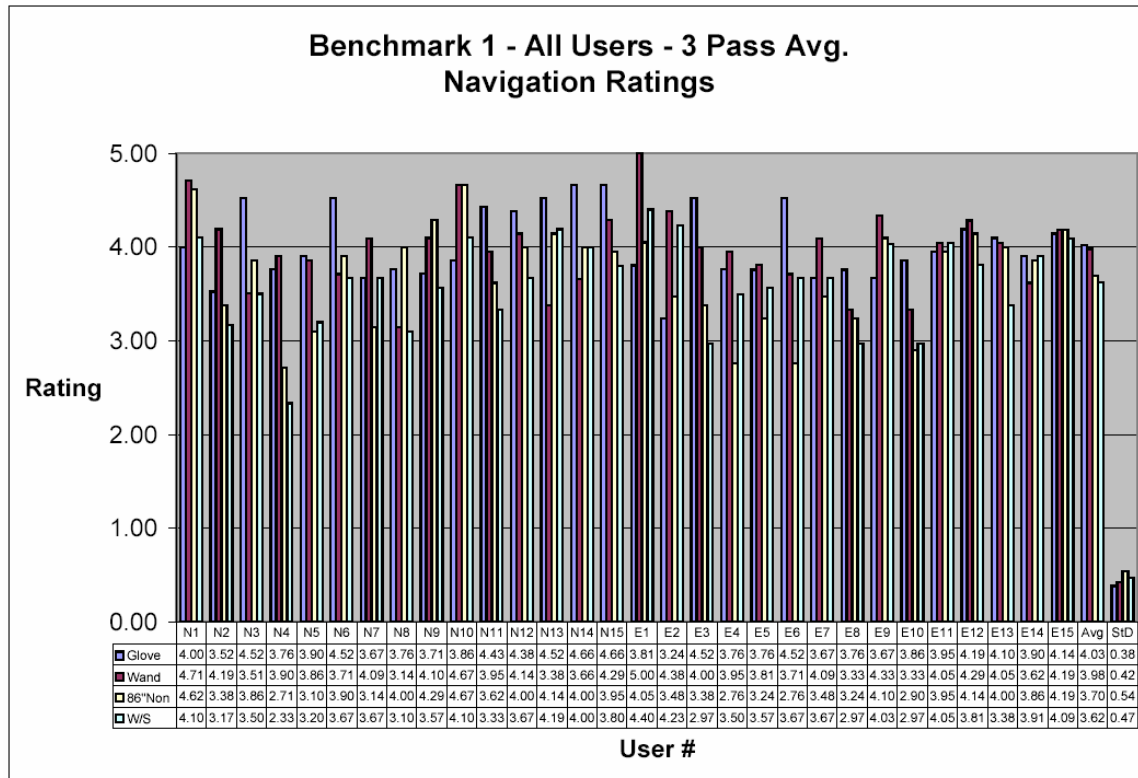




**Figure C- 19: B1ALL3pAvgTim All Users Elapsed Times**

**Table C- 19: B1ALL3pAvgTim All Users Elapsed Times Statistics**

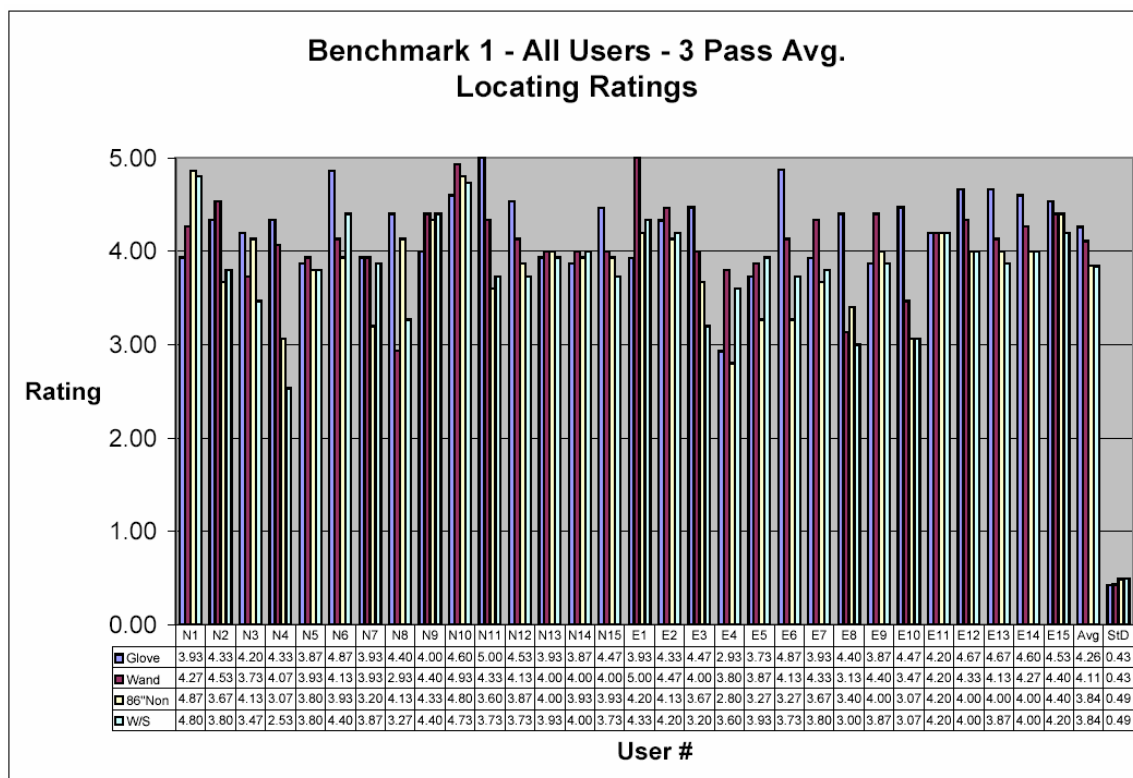
Benchmark 1, All Users, 3 Pass Avg., Elapsed Times												
B1A3pA	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	Coeff. of Var.
Glove	30	304.70		87.49		198.30		560.00		0.0965	Yes	28.71%
Wand	30	238.20		76.02		153.30		430.00		0.0258	No	31.91%
86*Non	30	309.80		138.40		161.70		781.70		0.0018	No	44.67%
W/S	30	309.40		114.40		168.30		748.30		0.0013	No	36.97%
B1A3pA		Homogeneity of Variance					Roy's Greatest Root:			F(3,27)	11.68	Pr > F <0.0001
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.4400	0.5092	Yes	29	17.78	0.0002	No	Means	Wand 27.92%
Glove vs 86*Non				1.9500	0.1674	Yes	29	0.05	0.8300	Yes	Neither	Neither
Glove vs W/S				0.5800	0.4478	Yes	29	0.05	0.8267	Yes	Neither	Neither
Wand vs 86*Non				2.7400	0.1033	Yes	29	12.21	0.0015	No	Means	Wand 30.06%
Wand vs W/S				1.1100	0.2965	Yes	29	24.05	0.0001	No	Means	Wand 29.89%
86*Non vs W/S				0.3400	0.5604	Yes	28	0.00	0.9819	Yes	Neither	Neither



**Figure C- 20: B1ALL3pAvgNav All Users Navigation Ratings**

**Table C- 20: B1ALL3pAvgNav All Users Navigation Ratings Statistics**

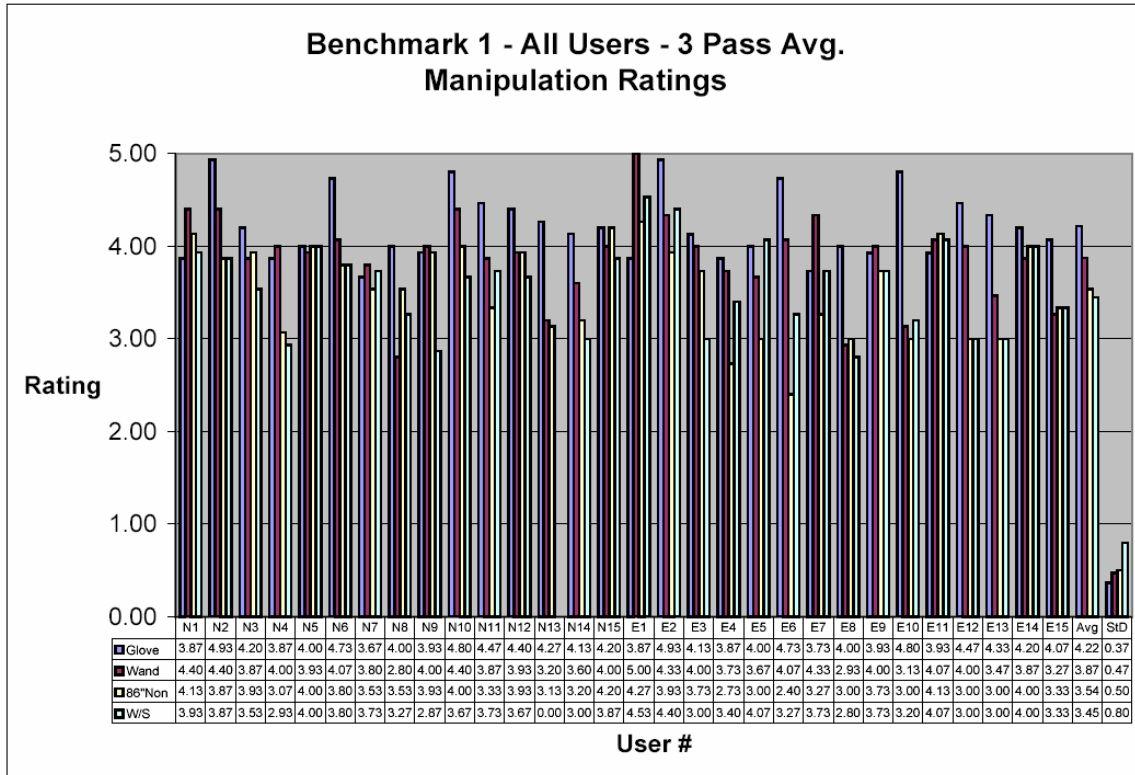
Benchmark 1, All Users, 3 Pass Avg., Navigation Ratings										
B1A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	30	4.024	0.381	3.24	4.67	0.0458	No	9.46%		
Wand	30	3.981	0.425	3.14	5.00	> 0.1000	Yes	10.69%		
86\"Non	30	3.697	0.539	2.71	4.67	0.0098	No	14.57%		
W/S	30	3.621	0.470	2.33	4.40	> 0.1000	Yes	12.98%		
B1A3pA	Homogeneity of Variance				Roy's Greatest Root:			F(3,27)		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand			0.4300	0.5130	Yes	29	0.14	0.7117	Yes	Neither
Glove vs 86\"Non			4.9600	0.0298	No	29	9.13	0.0052	No	Var/Mean
Glove vs W/S			1.3100	0.2570	Yes	29	14.38	0.0007	No	Means
Wand vs 86\"Non			2.0800	0.1542	Yes	29	8.29	0.0074	No	Means
Wand vs W/S			0.2700	0.6040	Yes	29	19.39	0.0001	No	Means
86\"Non vs W/S			0.6700	0.4158	Yes	29	0.91	0.3493	Yes	Neither



**Figure C- 21: B1ALL3pLOC All Users Locating Ratings**

**Table C- 21: B1ALL3pLOC All Users Locating Ratings Statistics**

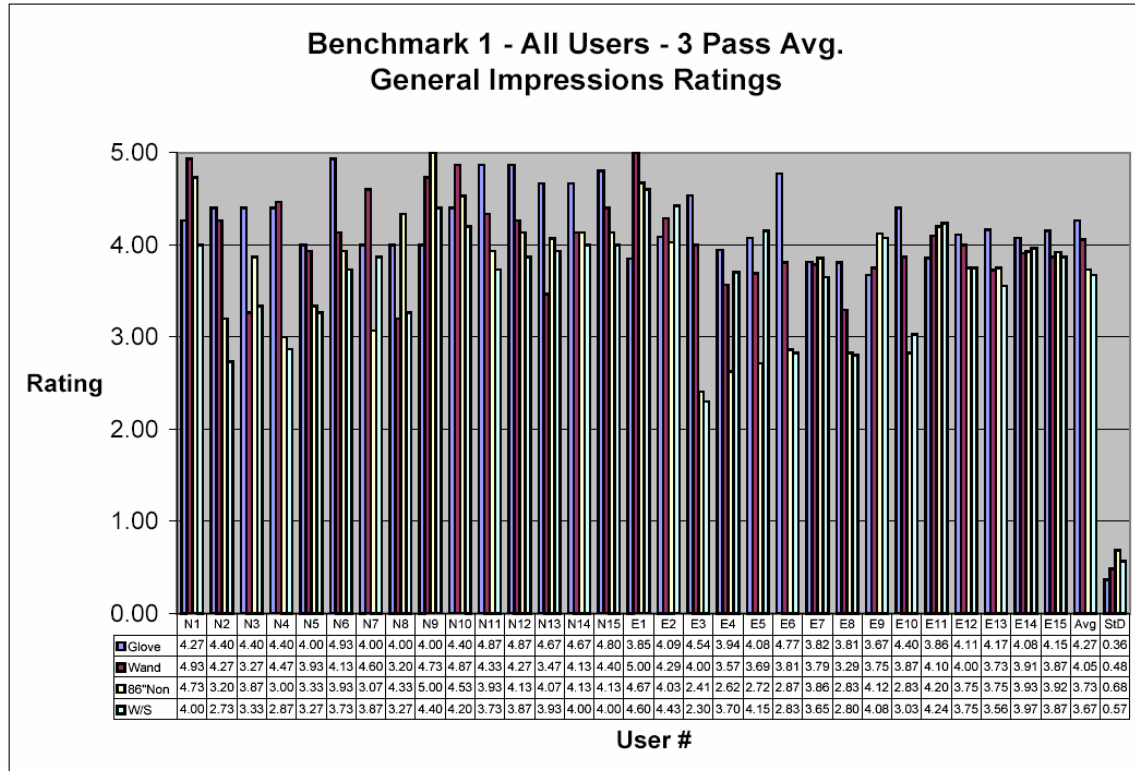
Benchmark 1, All Users, 3 Pass Avg., Locating Ratings										
B1A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	30	4.262	0.428	2.93	5.00	> 0.1000	Yes	10.05%		
Wand	30	4.108	0.432	2.93	5.00	> 0.1000	Yes	10.51%		
86\"Non	30	3.845	0.487	2.80	4.87	> 0.1000	Yes	12.66%		
W/S	30	3.840	0.494	2.53	4.80	0.0155	No	12.85%		
B1A3pA	Homogeneity of Variance				Roy's Greatest Root:				F(3,27)	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand	0.0019	0.9652			Yes	14	2.09	0.1588	Yes	Neither
Glove vs 86\"Non	0.4653	0.4952			Yes	14	15.33	0.0005	No	Means
Glove vs W/S			0.4000	0.5277	Yes	14	11.91	0.0017	No	Means
Wand vs 86\"Non	0.4079	0.5231			Yes	14	8.83	0.0059	No	Means
Wand vs W/S			0.3600	0.5525	Yes	14	14.46	0.0007	No	Means
86\"Non vs W/S			0.0100	0.9428	Yes	14	0.01	0.9420	Yes	Neither



**Figure C- 22: B1ALL3pAvgMov All Users Manipulation Ratings**

**Table C- 22 B1ALL3pAvgMov All Users Manipulation Ratings Statistics**

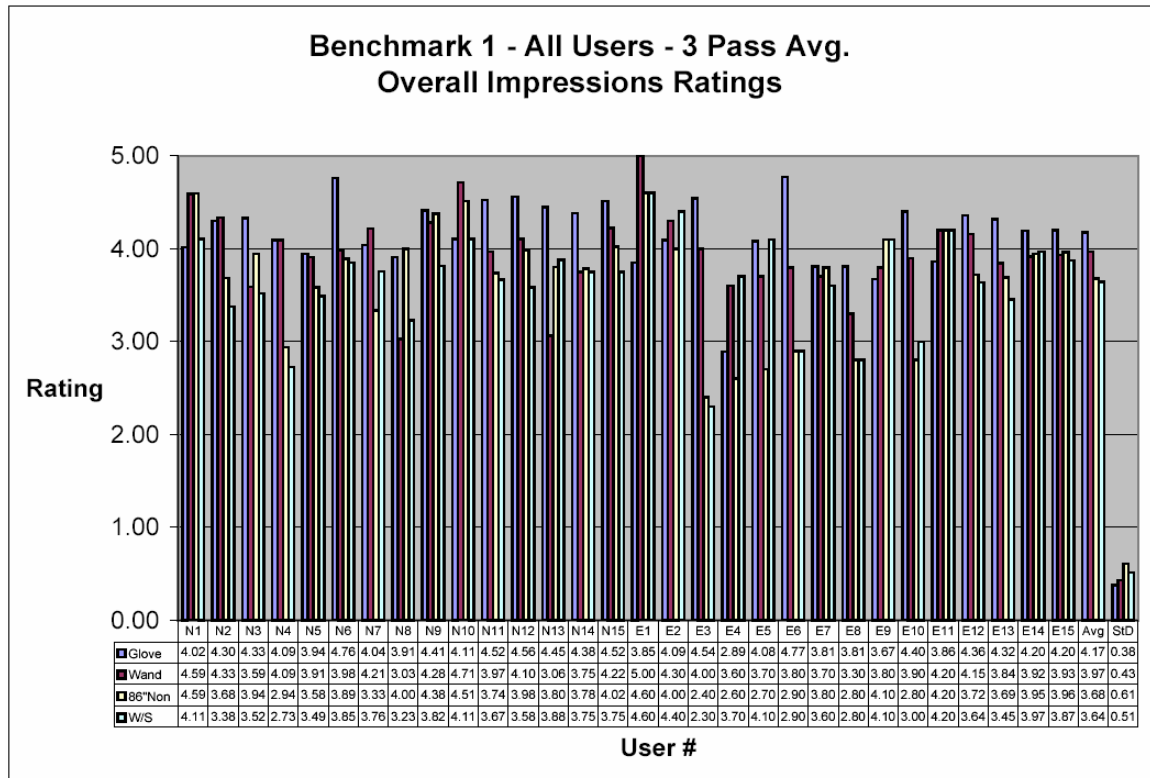
Benchmark 1, All Users, 3 Pass Avg., Manipulation Ratings										
B1A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.		
Glove	30	4.215	0.366	3.67	5.00	0.0637	Yes	8.69%		
Wand	30	3.871	0.472	2.80	4.27	0.0652	No	12.19%		
86\"Non	30	3.537	0.500	2.40	4.53	0.0784	Yes	14.12%		
W/S	30	3.584	0.461	2.80	376.70	> 0.1000	Yes	12.86%		
B1A3pA		Homogeneity of Variance				Roy's Greatest Root:			F(3,27)	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					
Glove vs Wand				1.6000	0.2109	Yes	29	10.85	0.0026	No
Glove vs 86\"Non		4.3700	0.0409			Yes	29	34.50	0.0001	No
Glove vs W/S		15.761	< 0.0001			No	29	38.86	0.0001	No
Wand vs 86\"Non				0.1100	0.7367	Yes	29	14.60	0.0006	No
Wand vs W/S				1.0200	0.3169	Yes	29	15.04	0.0006	No
86\"Non vs W/S		6.0294	0.0141			No	29	0.12	0.7313	Yes
										Statistically Significant
										Statistically Better
										Wand 8.89%
										Glove 19.17%
										Glove 17.61%
										Wand 9.44%
										Wand 8.01%
										Neither



**Figure C- 23: B1ALL3pAvgGen All Users General Impressions Ratings**

**Table C- 23: B1ALL3pAvgGen All Users General Impressions Ratings Statistics**

Benchmark 1, All Users, 3 Pass Avg., General Impressions Ratings											
B1A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	30	4.267	0.364	3.67	4.93	> 0.1000	Yes	8.53%			
Wand	30	4.056	0.480	3.20	5.00	> 0.1000	Yes	11.84%			
86\"Non	30	3.730	0.683	2.71	5.00	0.0157	No	18.30%			
W/S	30	3.672	0.567	2.30	4.60	0.0709	Yes	15.44%			
B1A3pA	Homogeneity of Variance					Roy's Greatest Root:			F(3,27)		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Statistically Significant	Pr > F
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	2.1701	0.1407			Yes	29	4.27	0.0478	No	Means	Wand 5.20%
Glove vs 86\"Non			12.0400	0.0010	No	29	14.18	0.0008	No	Var/Mean	Glove 14.40%
Glove vs W/S	5.4395	0.0197			No	29	19.61	0.0001	No	Var/Mean	Glove 16.20%
Wand vs 86\"Non			4.8900	0.0309	No	29	7.53	0.0103	No	Var/Mean	Wand 8.74%
Wand vs W/S	0.7818	0.3766			Yes	29	13.58	0.0009	No	Means	Wand 10.46%
86\"Non vs W/S			1.4300	0.2364	Yes	29	0.42	0.5229	Yes	Neither	Neither



**Figure C- 24: B1ALL3pAvgOvr All Users Overall Impressions Rating**

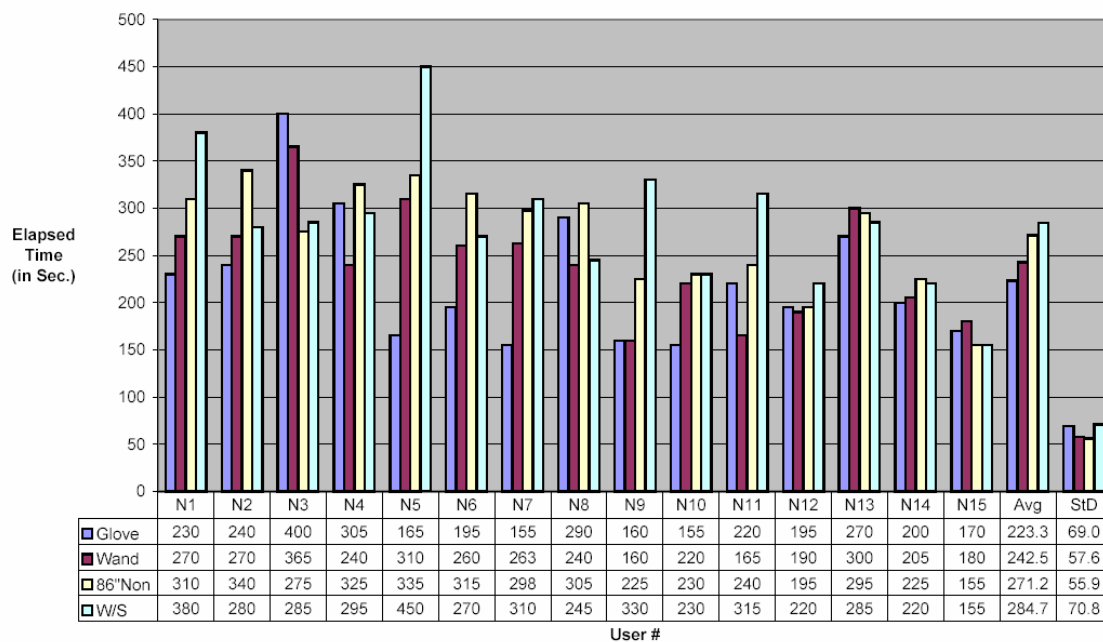
**Table C- 24: B1ALL3pAvgOvr All Users Overall Impressions Rating Statistics**

Benchmark 1, All Users, 3 Pass Avg., Overall Impressions Ratings											
B1A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	Coeff. of Var.			
Glove	30	4.173	0.376	2.89	4.77	> 0.1000	Yes	9.02%			
Wand	30	3.965	0.427	3.03	5.00	> 0.1000	Yes	10.77%			
86\"Non	30	3.679	0.605	2.40	4.60	0.0034	No	16.45%			
W/S	30	3.642	0.513	2.30	4.60	> 0.1000	Yes	14.07%			
Homogeneity of Variance				Roy's Greatest Root:		F(3,27)		6.00		Pr > F	
B1A3pA	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Statistically Significant	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.2700	0.6066	Yes	29	4.35	0.0460	No	Means	Wand 5.25%
Glove vs 86\"Non	5.1400	0.0271			No	29	15.31	0.0005	No	Var/Means	Glove 13.43%
Glove vs W/S			1.6800	0.2003	Yes	29	17.02	0.0003	No	Means	Glove 14.58%
Wand vs 86\"Non	3.3976	0.0653			Yes	29	7.62	0.0099	No	Means	Wand 7.77%
Wand vs W/S			0.7600	0.3868	Yes	29	11.77	0.0018	No	Means	Wand 8.87%
86\"Non vs W/S	0.7837	0.3760			Yes	29	0.22	0.6421	Yes	Neither	Neither

## **Appendix D**

### Novice User Benchmark 2 (Find/Repair) Detail

**Benchmark 2 - Novice Users  
Pass 1 Elapsed Times**

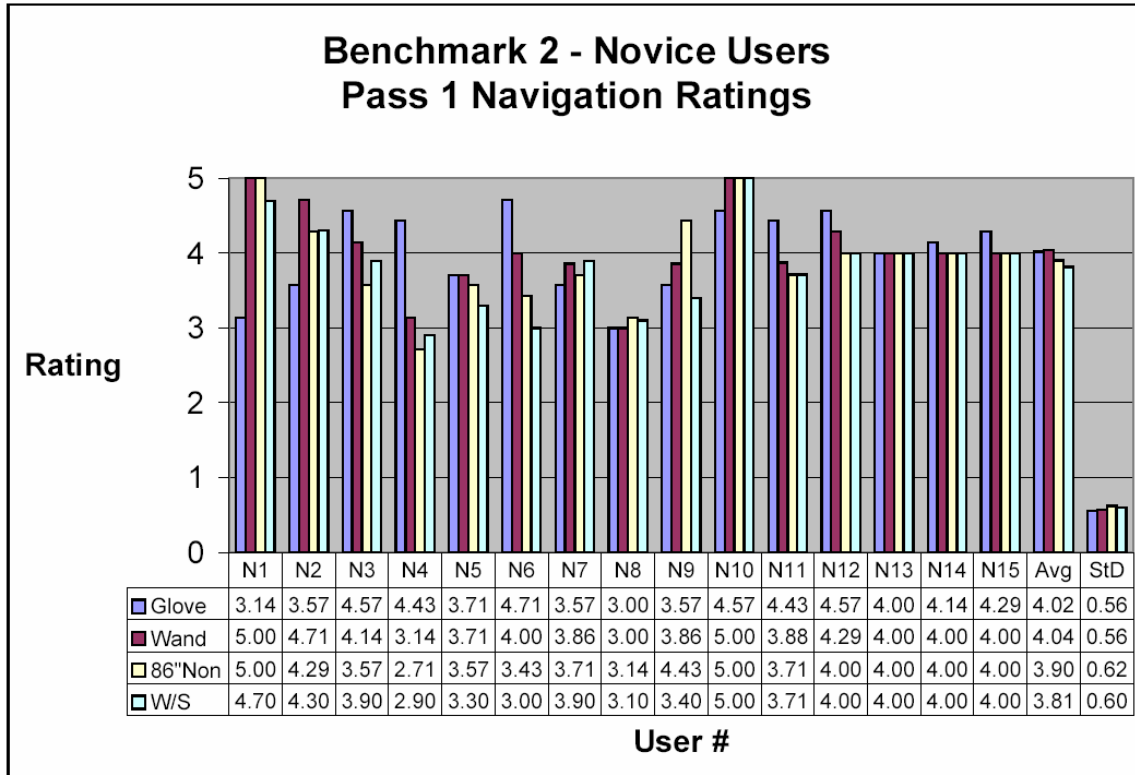


**Figure D- 1: B2Np1Tim Novice User Elapsed Times**

**Table D- 1: B2Np1Tim Novice User Elapsed Times Statistics**

Benchmark 2, Novice Users, Pass 1, Elapsed Times											
B2Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	223.3	68.96	155.0	400.0	> 0.1000	Yes	30.88%			
Wand	15	242.5	57.64	160.0	365.0	> 0.1000	Yes	23.77%			
86\"Non	15	271.2	55.94	155.0	340.0	>0.1000	Yes	20.63%			
W/S	15	284.7	70.77	155.0	450.0	> 0.1000	Yes	24.86%			
B2Np1	Homogeneity of Variance				Roy's Greatest Root: F(3,12) 2.38 Pr > F 0.1206						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.4328	0.5106			Yes	14	1.54	0.2344	Yes	Neither	Neither
Glove vs 86\"Non	0.5877	0.4433			Yes	14	6.52	0.0229	No	Means	Glove 17.66%
Glove vs W/S	0.0091	0.9241			Yes	14	5.69	0.0318	No	Means	Glove 21.57%
Wand vs 86\"Non	0.0120	0.9126			Yes	14	5.83	0.0301	No	Means	Wand 10.58%
Wand vs W/S	0.5659	0.4519			Yes	14	5.29	0.0373	No	Means	Wand 14.82%
86\"Non vs W/S	0.7408	0.3864			Yes	14	0.88	0.3635	Yes	Neither	Neither

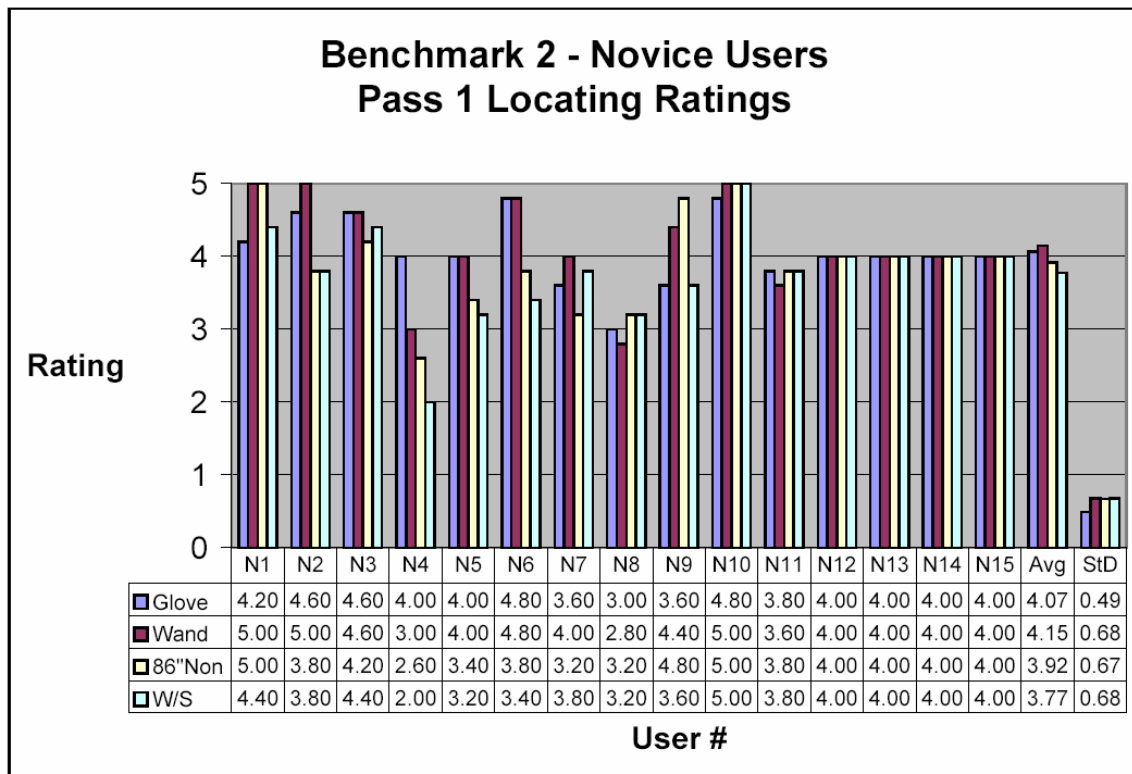




**Figure D- 2: B2Np1Nav Novice User Navigation Ratings**

**Table D- 2: B2Np1Nav Novice User Navigation Ratings Statistics**

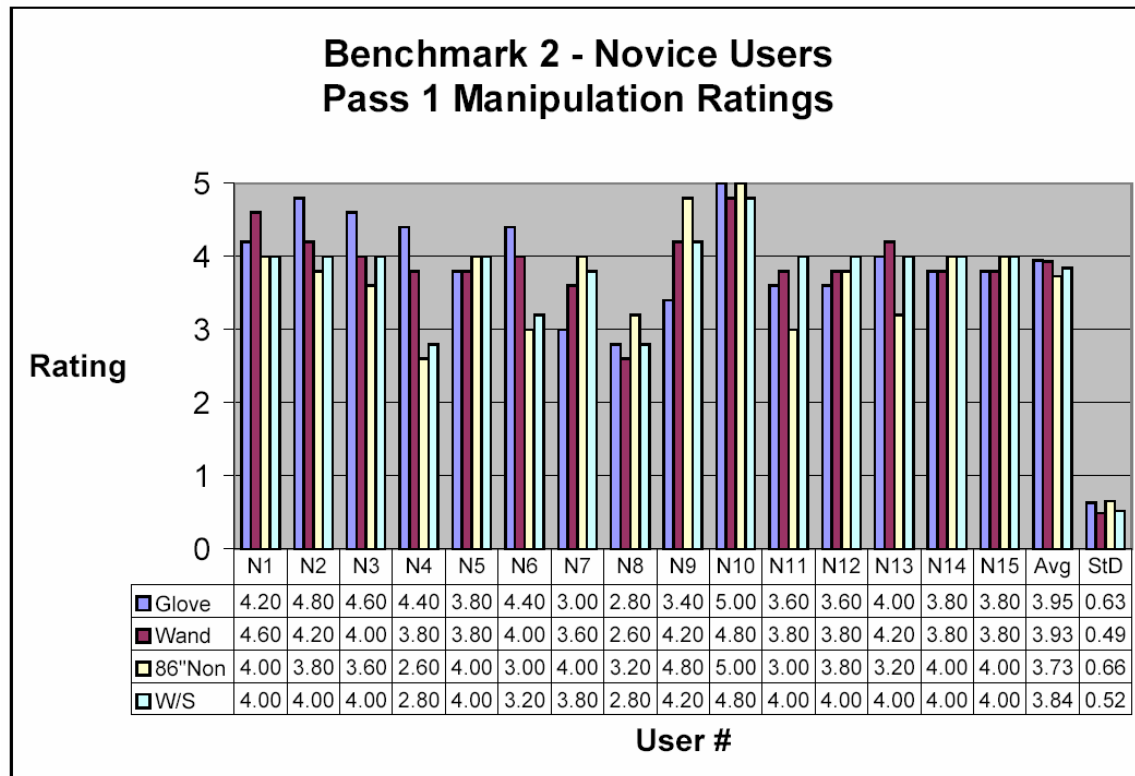
Benchmark 2, Novice Users, Pass 1, Navigation Ratings											
B2Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.02	0.56	3.00	4.71	> 0.1000	Yes	13.84%			
Wand	15	4.04	0.56	3.00	5.00	> 0.1000	Yes	13.98%			
86\"Non	15	3.90	0.62	2.71	5.00	> 0.1000	Yes	15.90%			
W/S	15	3.81	0.56	2.90	5.00	> 0.1000	Yes	14.67%			
		Homogeneity of Variance			Roy's Greatest Root:		F(3,12)	2.95	Pr > F		0.0755
B2Np1	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		0.0032	0.9546		Yes	14	0.01	0.9142	Yes	Neither	Neither
Glove vs 86\"Non		0.1652	0.6844		Yes	14	0.24	0.6309	Yes	Neither	Neither
Glove vs W/S		0.0778	0.7803		Yes	14	0.92	0.3539	Yes	Neither	Neither
Wand vs 86\"Non		0.1222	0.7266		Yes	14	3.11	0.0995	No	Means	Wand 3.46%
Wand vs W/S		0.0493	0.8242		Yes	14	9.55	0.0080	No	Means	Wand 5.90%
86\"Non vs W/S		0.0163	0.8984		Yes	14	1.16	0.3022	Yes	Neither	Neither



**Figure D- 3: B2Np1Loc Novice User Locating Ratings**

**Table D- 3: B2Np1Loc Novice User Locating Ratings Statistics**

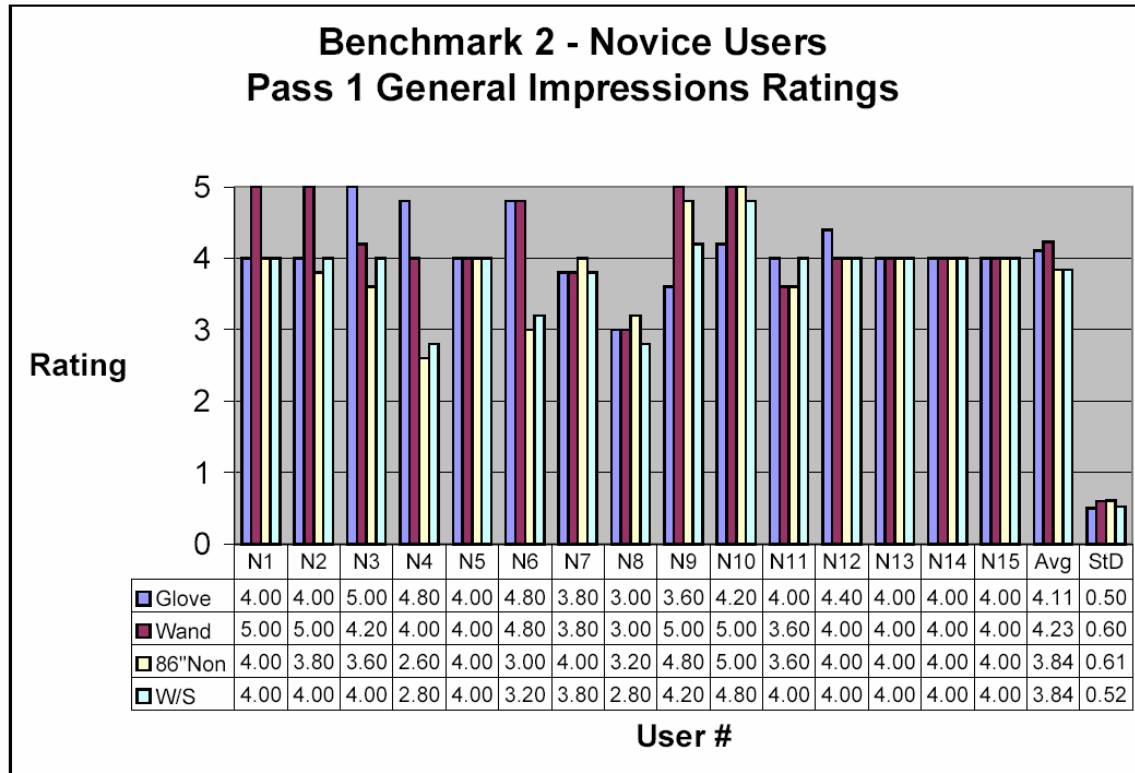
Benchmark 2, Novice Users, Pass 1, Locating Ratings												
B2Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.07	0.49	3.00	4.80	0.0471	No	12.00%				
Wand	15	4.15	0.68	2.80	5.00	0.062	Yes	16.35%				
86"Non	15	3.92	0.67	2.60	5.00	> 0.1000	Yes	17.12%				
W/S	15	3.77	0.68	2.00	5.00	> 0.1000	Yes	18.02%				
B2Np1		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 2.03 Pr > F 0.1629						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				1.6200	0.2140	Yes	14	0.51	0.4860	Yes	Neither	Neither
Glove vs 86"Non				1.4400	0.2402	Yes	14	0.74	0.4046	Yes	Neither	Neither
Glove vs W/S				0.8100	0.3754	Yes	14	2.94	0.1085	Yes	Neither	Neither
Wand vs 86"Non		0.0015	0.9693			Yes	14	3.22	0.0943	No	Means	Wand 5.79%
Wand vs W/S		0.0001	0.9927			Yes	14	6.92	0.0198	No	Means	Wand 9.91%
86"Non vs W/S		0.0023	0.9620			Yes	14	1.85	0.1949	Yes	Neither	Neither



**Figure D- 4: B2Np1Mov Novice User Manipulation Ratings**

**Table D- 4: B2Np1Mov Novice User Manipulation Ratings Statistics**

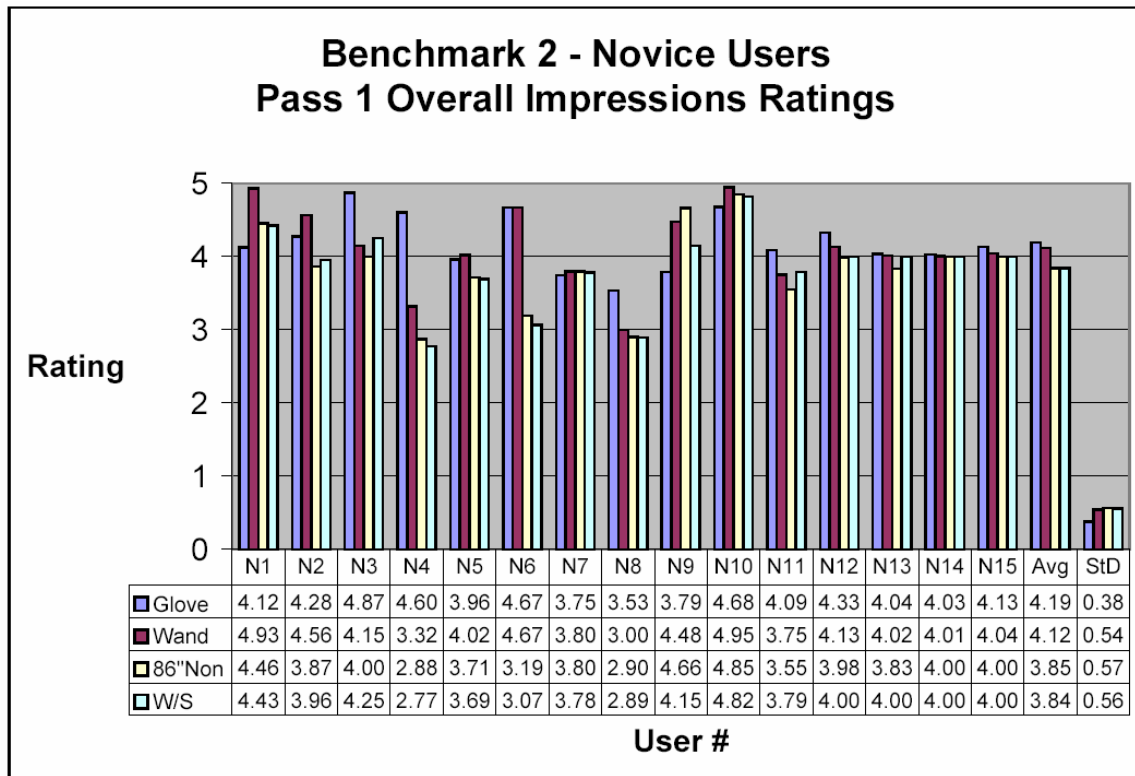
Benchmark 2, Novice Users, Pass 1, Manipulation Ratings												
B2Np1	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	CV
Glove	15	3.95		0.63		2.80		5.00		> 0.1000	Yes	15.96%
Wand	15	3.93		0.49		2.60		4.80		0.0073	No	12.56%
86"Non	15	3.73		0.66		2.60		5.00		0.0763	Yes	17.62%
W/S	15	3.84		0.52		2.80		4.80		< 0.0001	No	13.67%
B2Np1		Homogeneity of Variance					Roy's Greatest Root: F(3,12)			0.52	0.6776	0.0561
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.7400	0.3970	Yes	14	0.01	0.9063	Yes	Neither	Neither
Glove vs 86"Non		0.7951	0.3726			Yes	14	0.88	0.3636	Yes	Neither	Neither
Glove vs W/S				0.0400	0.8519	Yes	14	0.36	0.5579	Yes	Neither	Neither
Wand vs 86"Non				0.5400	0.4690	Yes	14	1.62	0.2244	Yes	Neither	Neither
Wand vs W/S				0.9300	0.3439	Yes	14	0.82	0.3803	Yes	Neither	Neither
86"Non vs W/S				0.0300	0.8574	Yes	14	1.00	0.3343	Yes	Neither	Neither



**Figure D- 5: B2Np1Gen Novice User General Impressions Ratings**

**Table D- 5: B2Np1Gen Novice User General Impressions Ratings Statistics**

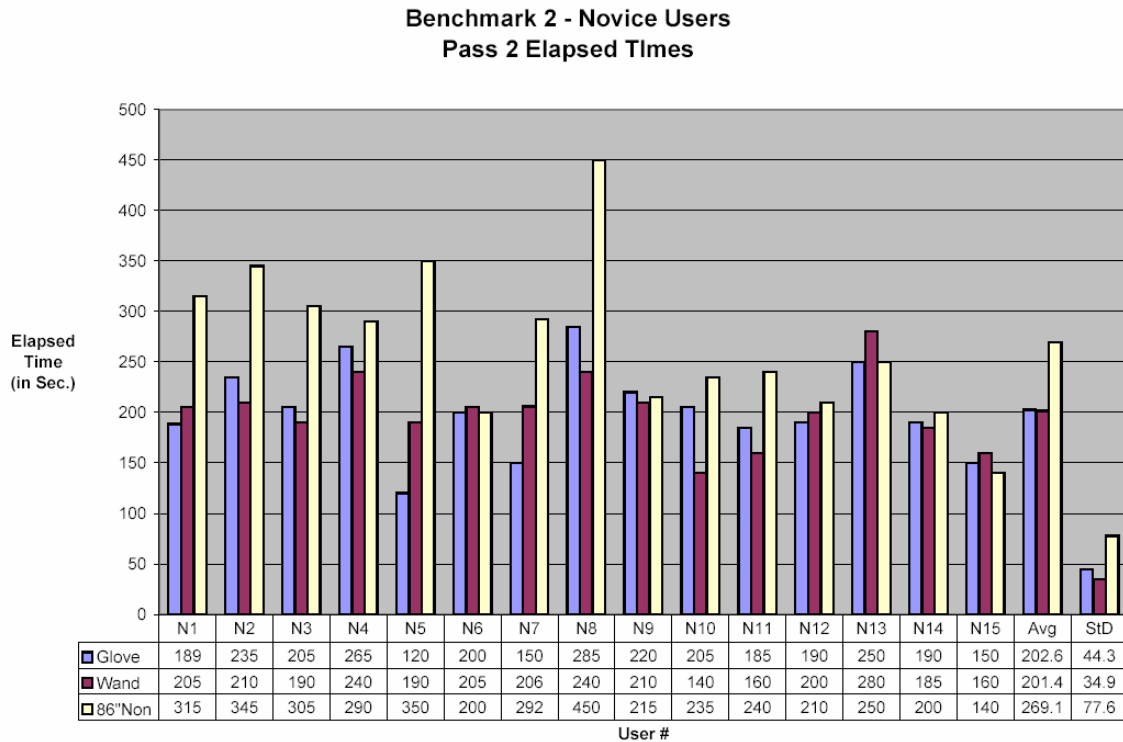
Benchmark 2, Novice Users, Pass 1, General Impressions Ratings											
B2Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.11	0.50	3.00	5.00	0.0118	No	12.19%			
Wand	15	4.23	0.60	3.00	5.00	0.0149	No	14.29%			
86\"Non	15	3.84	0.61	2.60	5.00	0.0062	No	15.90%			
W/S	15	3.84	0.52	2.80	4.80	< 0.0001	No	13.67%			
B2Np1	Homogeneity of Variance				Roy's Greatest Root: F(3,12)			1.93	Pr > F	0.1785	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.5800	0.4351	Yes	14	0.51	0.4888	Yes	Neither	Neither
Glove vs 86\"Non			0.4700	0.4973	Yes	14	1.30	0.2735	Yes	Neither	Neither
Glove vs W/S			0.0300	0.8721	Yes	14	2.01	0.1777	Yes	Neither	Glove 6.95%
Wand vs 86\"Non			0.0000	0.9672	Yes	14	5.34	0.0366	No	Means	Wand 10.08%
Wand vs W/S			0.3200	0.5791	Yes	14	6.71	0.0214	No	Means	Wand 10.08%
86\"Non vs W/S			0.2800	0.6041	Yes	14	0.00	1.0000	Yes	Neither	Either



**Figure D- 6: B2Np1Ovr Novice User Pass 1 Overall Impressions Ratings**

**Table D- 6: B2Np1Ovr Novice User Pass 1 Overall Impressions Ratings Statistics**

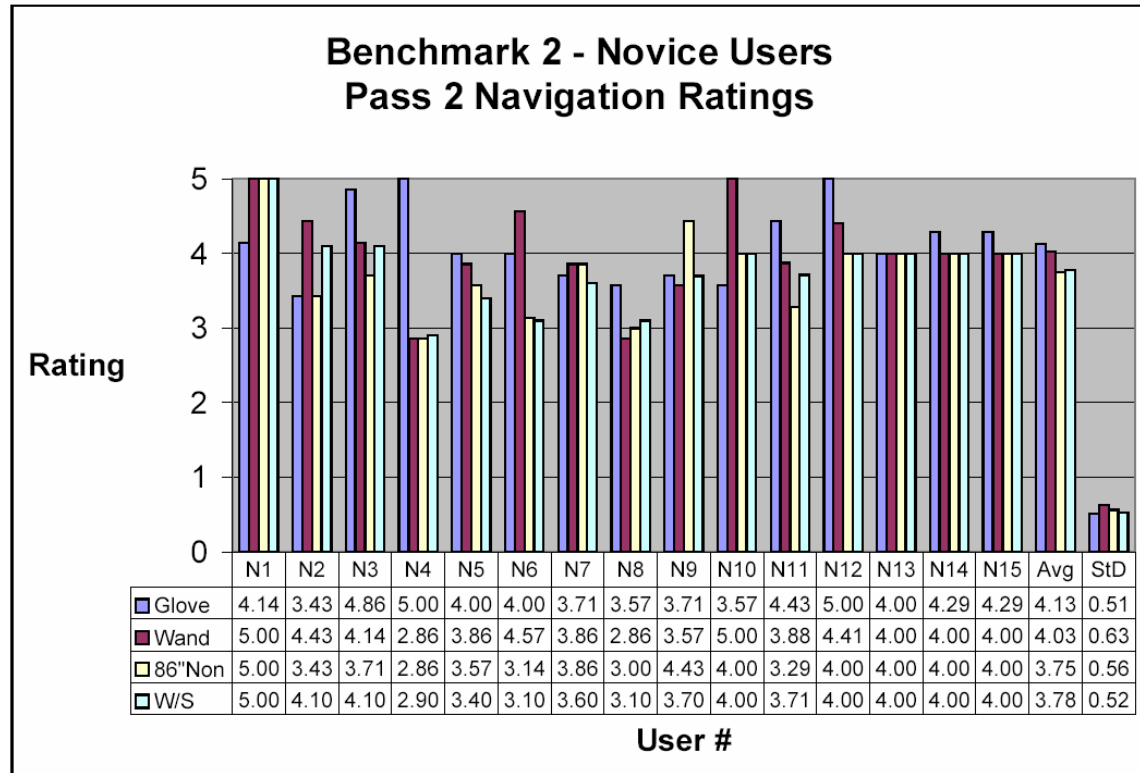
Benchmark 2, Novice Users, Pass 1, Overall Impressions Ratings												
B2Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.19	0.38	3.53	4.87	> 0.1000	Yes	9.10%				
Wand	15	4.12	0.54	3.00	4.95	> 0.1000	Yes	13.20%				
86"Non	15	3.85	0.57	2.88	4.85	> 0.1000	Yes	14.75%				
W/S	15	3.84	0.56	2.77	4.82	> 0.1000	Yes	14.53%				
B2Np1		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 2.21 Pr > F 0.1402						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		1.6687	0.1964			Yes	14	0.26	0.6152	Yes	Neither	Neither
Glove vs 86"Non		2.0734	0.1499			Yes	14	4.11	0.0621	No	Means	Glove 9.00%
Glove vs W/S		1.9088	0.1671			Yes	14	4.72	0.0475	No	Means	Glove 9.14%
Wand vs 86"Non		0.0235	0.8782			Yes	14	7.29	0.0172	No	Means	Wand 7.20%
Wand vs W/S		0.0086	0.9261			Yes	14	6.53	0.0229	No	Means	Wand 7.34%
86"Non vs W/S		0.0037	0.9518			Yes	14	0.01	0.9095	Yes	Neither	Neither



**Figure D- 7: B2Np2Tim Novice User Elapsed Times**

**Table D- 7: B2Np2Tim Novice User Elapsed Times Statistics**

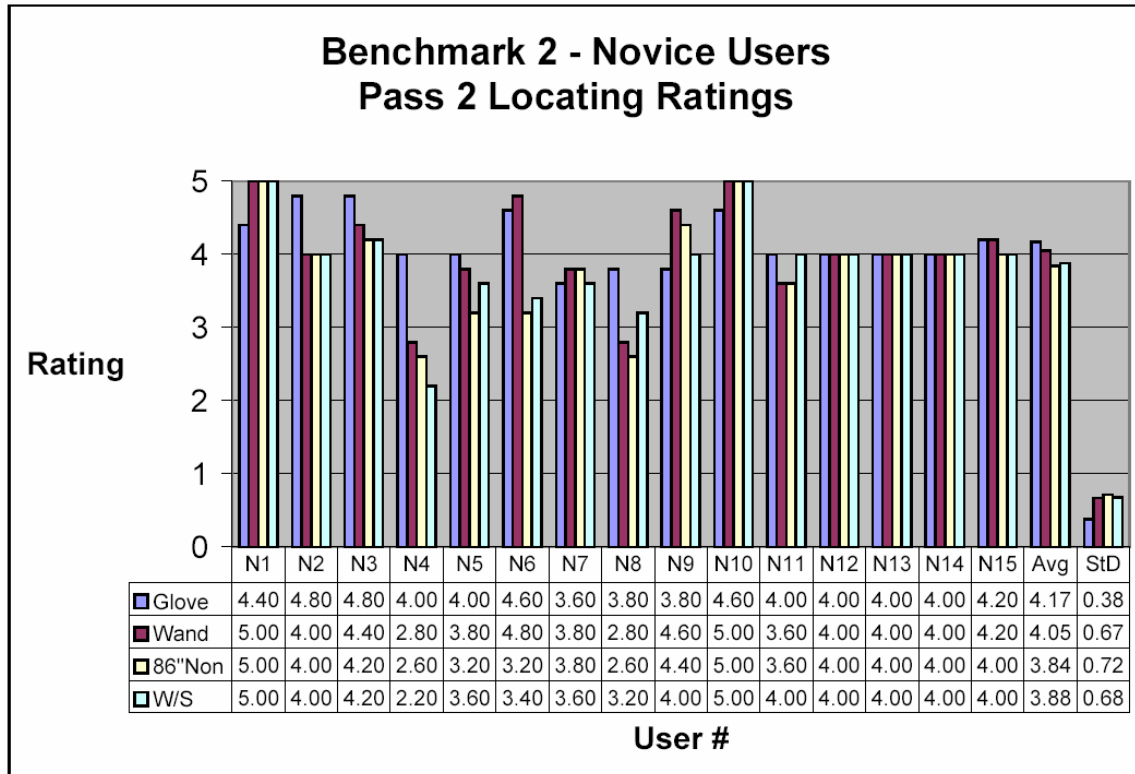
Benchmark 2, Novice Users, Pass 2, Elapsed Times											
B2Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	206.6	44.30	120.0	285.0	> 0.1000	Yes	21.44%			
Wand	15	201.4	34.93	140.0	280.0	0.0979	Yes	17.34%			
86\"Non	15	269.1	77.64	140.0	450.0	>0.1000	Yes	28.85%			
W/S	15	243.0	62.01	140.0	415.0	0.0346	No	25.52%			
B2Np2		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	5.55	Pr > F	0.0127
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand		0.7556	0.3847			Yes	14	0.02	0.8984	Yes	Neither
Glove vs 86\"Non		12.0700	0.0037			No	14	12.07	0.0037	No	Var/Mean
Glove vs W/S				0.7200	0.4047	Yes	14	6.46	0.0235	No	Means
Wand vs 86\"Non		7.8383	0.0051			No	14	13.49	0.0025	No	Var/Mean
Wand vs W/S				1.4500	0.2389	Yes	14	10.08	0.0067	No	Means
86\"Non vs W/S				0.4700	0.4984	Yes	14	1.17	0.2125	Yes	Neither



**Figure D- 8: B2Np2Nav Novice User Navigation Ratings**

**Table D- 8: B2Np2Nav Novice User Navigation Ratings Statistics**

Benchmark 2, Novice Users, Pass 2, Navigation Ratings												
B2Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.13	0.51	3.43	5.00	> 0.1000	Yes	12.45%				
Wand	15	4.03	0.63	2.86	5.00	> 0.1000	Yes	15.59%				
86"Non	15	3.75	0.56	2.86	5.00	> 0.1000	Yes	14.98%				
W/S	15	3.78	0.52	2.90	5.00	0.0938	Yes	13.82%				
B2Np2		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 2.28			Pr > F 0.1318			
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		0.7150	0.3978			Yes	14	0.22	0.6479	Yes	Neither	Neither
Glove vs 86"Non		0.0822	0.7744			Yes	14	3.41	0.0858	No	Means	Glove 10.13%
Glove vs W/S		0.0043	0.9475			Yes	14	3.44	0.0847	No	Means	Glove 9.31%
Wand vs 86"Non		0.3137	0.5754			Yes	14	3.62	0.0779	No	Means	Wand 6.56%
Wand vs W/S		0.8298	0.3623			Yes	14	4.48	0.0526	No	Means	Wand 6.56%
86"Non vs W/S		0.1242	0.7246			Yes	14	0.12	0.7375	Yes	Neither	Neither

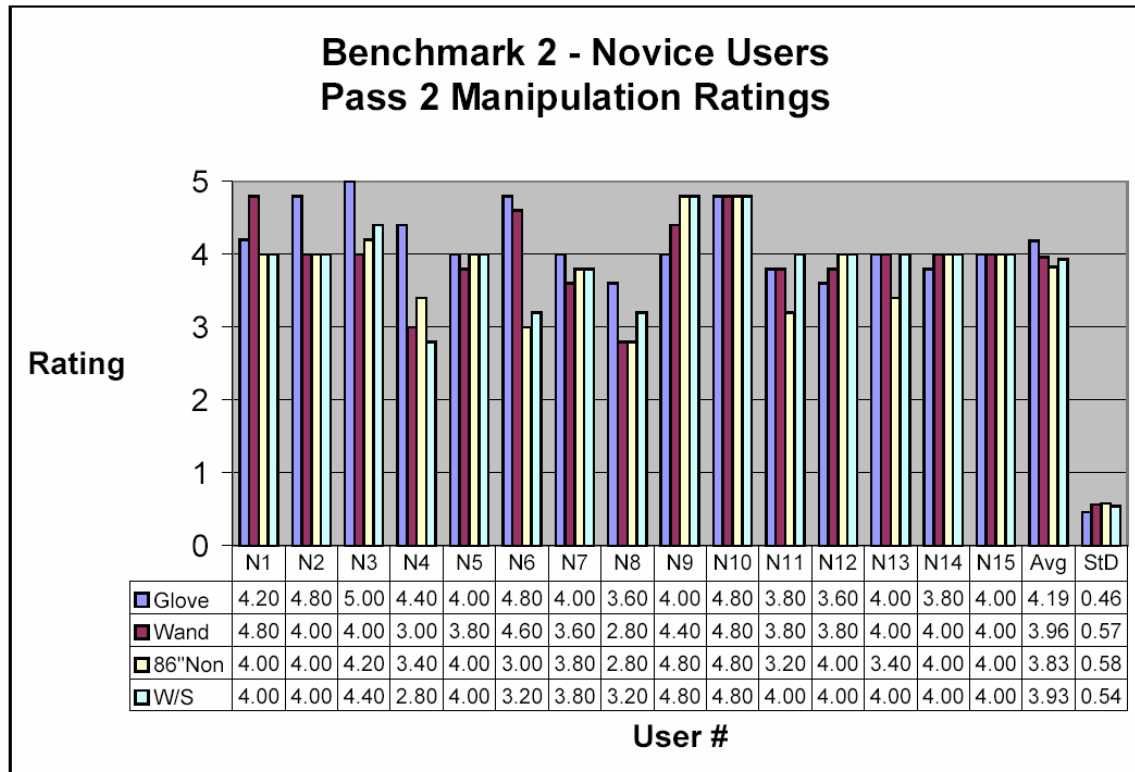


**Figure D- 9: B2Np2Loc Novice User Locating Ratings**

**Table D- 9: B2Np2Loc Novice User Locating Ratings Statistics**

Benchmark 2, Novice Users, Pass 2, Locating Ratings											
B2Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.17	0.38	3.60	4.80	0.0029	No	9.03%			
Wand	15	4.05	0.67	2.80	5.00	> 0.1000	Yes	16.52%			
86\"Non	15	3.84	0.72	2.60	5.00	> 0.1000	Yes	18.70%			
W/S	15	3.88	0.68	2.00	5.00	0.0229	No	17.40%			
Homogeneity of Variance				Roy's Greatest Root: F(3,12)				1.40	Pr > F	0.2913	
B2Np2	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			3.6900	0.0648	Yes	14	0.68	0.4243	Yes	Neither	Neither
Glove vs 86\"Non			4.5500	0.0418	Yes	14	3.82	0.0780	No	Means	Glove 8.67%
Glove vs W/S			2.0100	0.1670	Yes	14	3.28	0.0916	No	Means	Glove 7.55%
Wand vs 86\"Non	0.0656	0.7979			Yes	14	3.92	0.0667	No	Means	Wand 5.55%
Wand vs W/S			0.0000	0.9775	Yes	14	2.33	0.0493	No	Means	Wand 4.46%
86\"Non vs W/S			0.0500	0.8316	Yes	14	0.32	0.5816	Yes	Neither	Neither

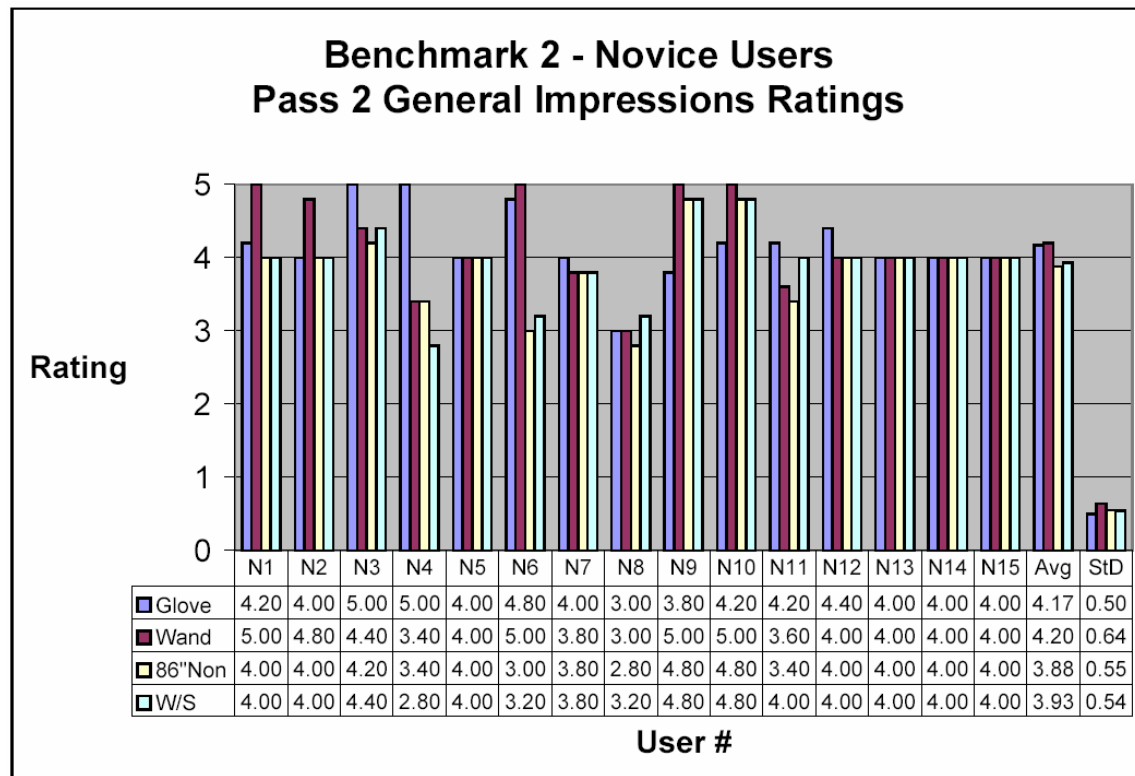




**Figure D- 10: B2Np2Mov Novice User Manipulation Ratings**

**Table D- 10: B2Np2Mov Novice User Manipulation Ratings Statistics**

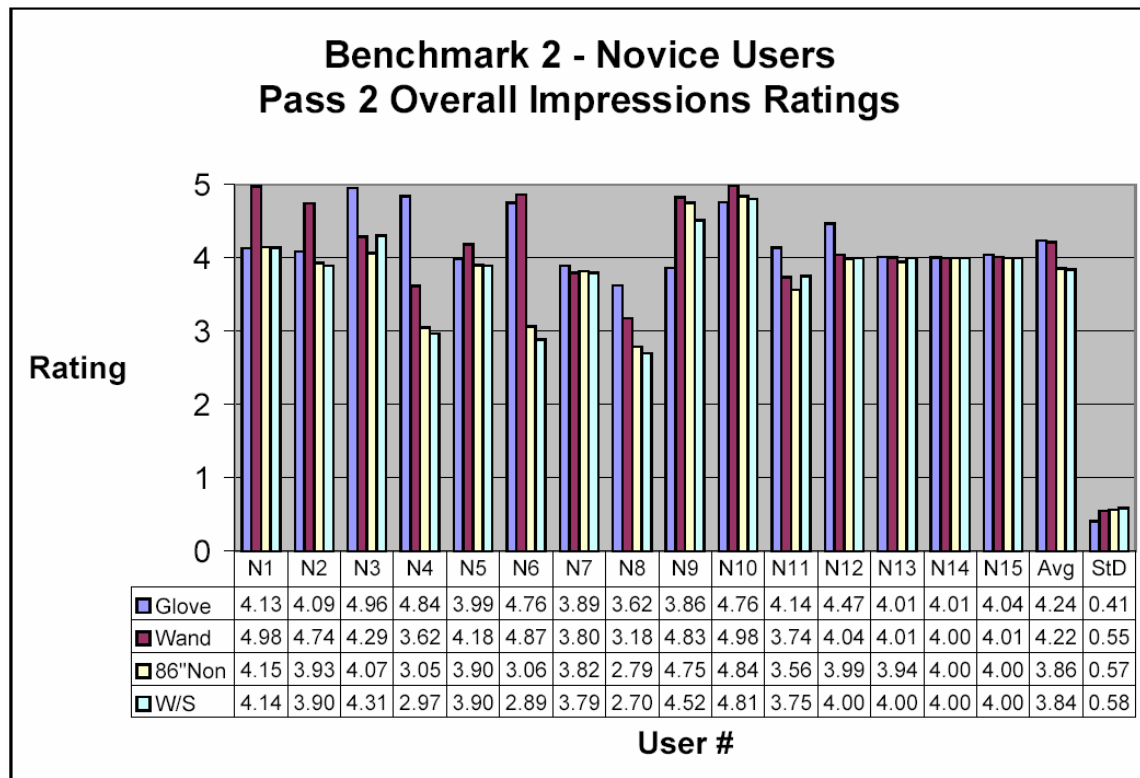
Benchmark 2, Novice Users, Pass 2, Manipulation Ratings											
B2Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.19	0.46	3.60	5.00	0.0088	No	11.05%			
Wand	15	3.96	0.57	2.80	4.80	0.0892	Yes	14.31%			
86\"Non	15	3.83	0.58	2.80	4.80	0.0546	Yes	15.16%			
W/S	15	3.93	0.54	2.80	4.80	0.0022	No	13.82%			
B2Np2		Homogeneity of Variance				Roy's Greatest Root: F(3,12)			1.80	Pr > F	0.2003
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						Statistically Better
Glove vs Wand				0.8800	0.4187	Yes	14	2.50	0.1359	Yes	Neither
Glove vs 86\"Non				1.0200	0.3221	Yes	14	4.65	0.0490	No	Means
Glove vs W/S				0.4300	0.5186	Yes	14	2.15	0.1649	Yes	Neither
Wand vs 86\"Non		0.0073	0.9321			Yes	14	0.90	0.3580	Yes	Neither
Wand vs W/S				0.0300	0.8741	Yes	14	0.05	0.8338	Yes	Neither
86\"Non vs W/S				0.0700	0.7929	Yes	14	1.67	0.2170	Yes	Neither



**Figure D- 11: B2Np2Gen Novice User General Impressions Ratings**

**Table D- 11: B2Np2Gen Novice User General Impressions Ratings Statistics**

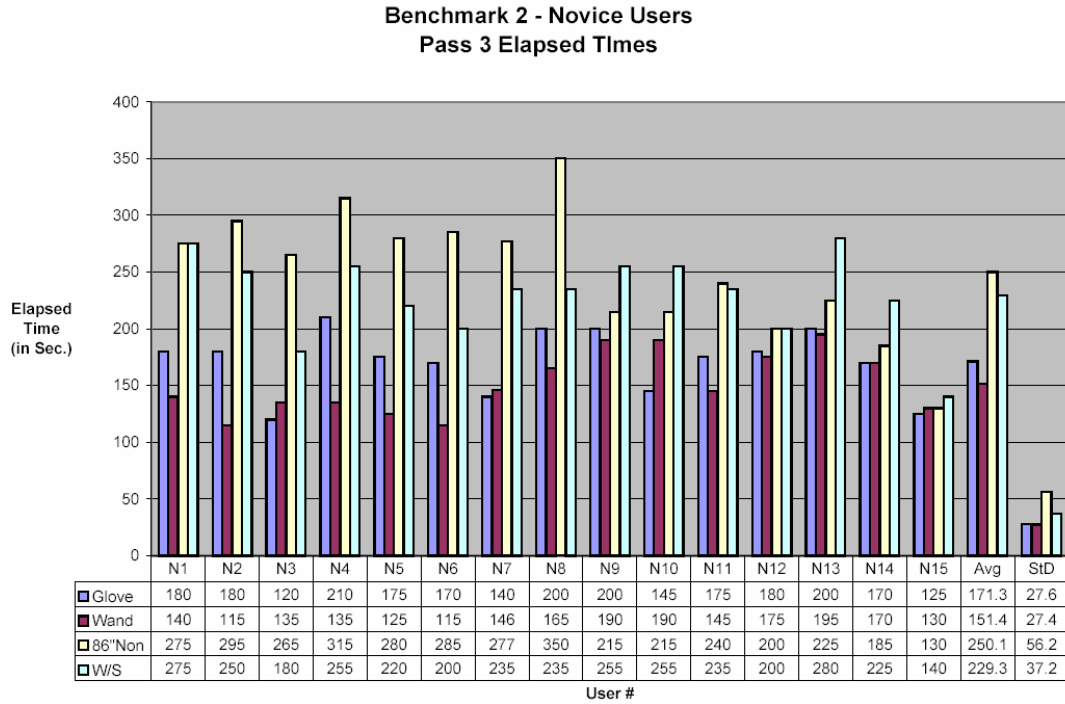
Benchmark 2, Novice Users, Pass 2, General Impressions Ratings												
B2Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.17	0.50	3.00	5.00	0.0300	No	12.00%				
Wand	15	4.20	0.64	3.00	5.00	0.0443	No	15.27%				
86"Non	15	3.88	0.55	2.80	4.80	0.0110	No	14.29%				
W/S	15	3.93	0.54	2.80	4.80	0.0022	No	13.82%				
B2Np2		Homogeneity of Variance				Roy's Greatest Root: F(3,12)			1.62	Pr > F 0.2368		
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				1.0800	0.9000	Yes	14	0.02	0.8849	Yes	Neither	Neither
Glove vs 86"Non				0.1300	0.1397	Yes	14	2.43	0.1412	Yes	Neither	Neither
Glove vs W/S				0.0800	0.7773	Yes	14	1.41	0.2547	Yes	Neither	Neither
Wand vs 86"Non				0.4400	0.5118	Yes	14	5.00	0.0422	No	Means	Wand 8.25%
Wand vs W/S				0.5500	0.4635	Yes	14	3.35	0.0884	No	Means	Wand 6.79%
86"Non vs W/S				0.0100	0.9389	Yes	14	0.65	0.4332	Yes	Neither	Neither



**Figure D- 12: B2Np2Ovr Novice User Overall Impressions Ratings**

**Table D- 12: B2Np2Ovr Novice User Overall Impressions Ratings Statistics**

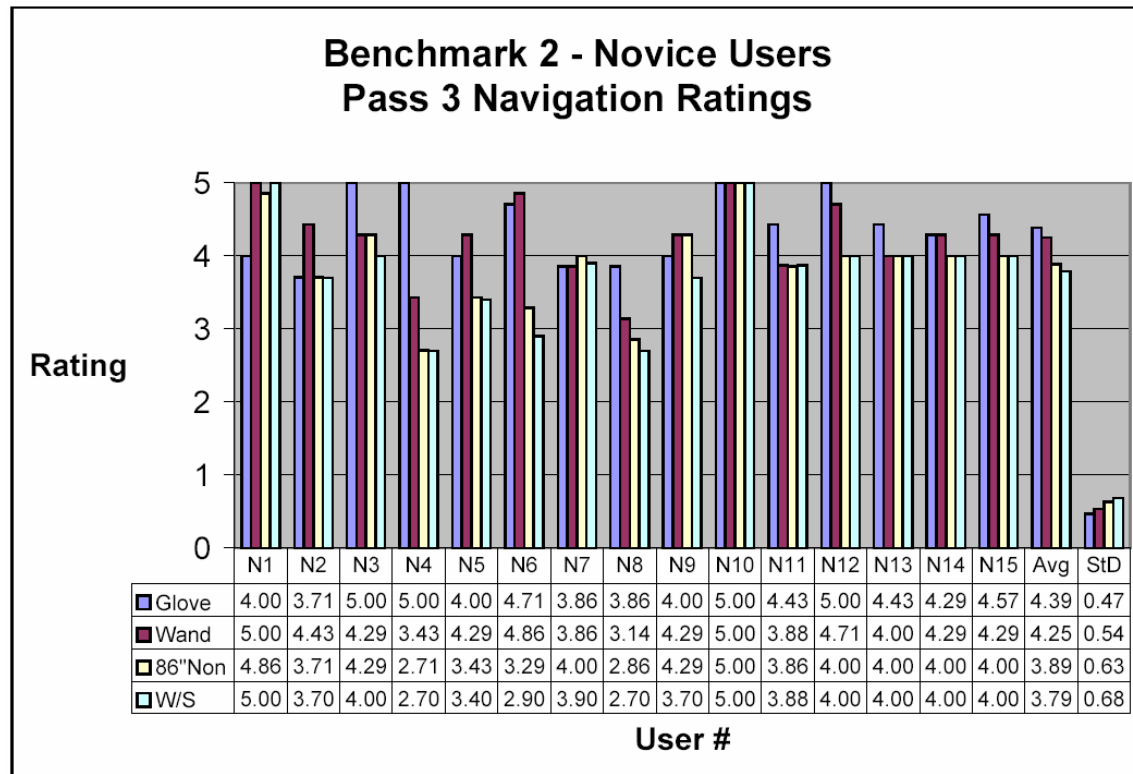
Benchmark 2, Novice Users, Pass 2, Overall Impressions Ratings												
B2Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.24	0.47	3.62	4.96	0.0072	No	11.12%				
Wand	15	4.22	0.55	3.18	4.98	> 0.1000	Yes	13.05%				
86"Non	15	3.86	0.57	2.79	4.84	0.0816	Yes	14.67%				
W/S	15	3.85	0.59	2.70	4.81	0.0250	No	15.22%				
B2Np2		Homogeneity of Variance				Roy's Greatest Root: F(3,12)		2.53	Pr > F		0.1089	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				2.0100	0.1670	Yes	14	0.02	0.8951	Yes	Neither	
Glove vs 86"Non				1.1500	0.2295	Yes	14	4.49	0.0525	No	Means	Glove 9.88%
Glove vs W/S				1.7500	0.1962	Yes	14	4.75	0.0469	No	Means	Glove 10.22%
Wand vs 86"Non		0.0101	0.9201			Yes	14	8.23	0.0124	No	Means	Wand 9.36%
Wand vs W/S				0.0700	0.7925	Yes	14	7.10	0.0185	No	Means	Wand 9.70%
86"Non vs W/S				0.0200	0.8953	Yes	14	0.14	0.7152	Yes	Neither	Neither



**Figure D- 13: B2Np3Tim Novice User Elapsed Times**

**Table D- 13: B2Np3Tim Novice User Elapsed Times Statistics**

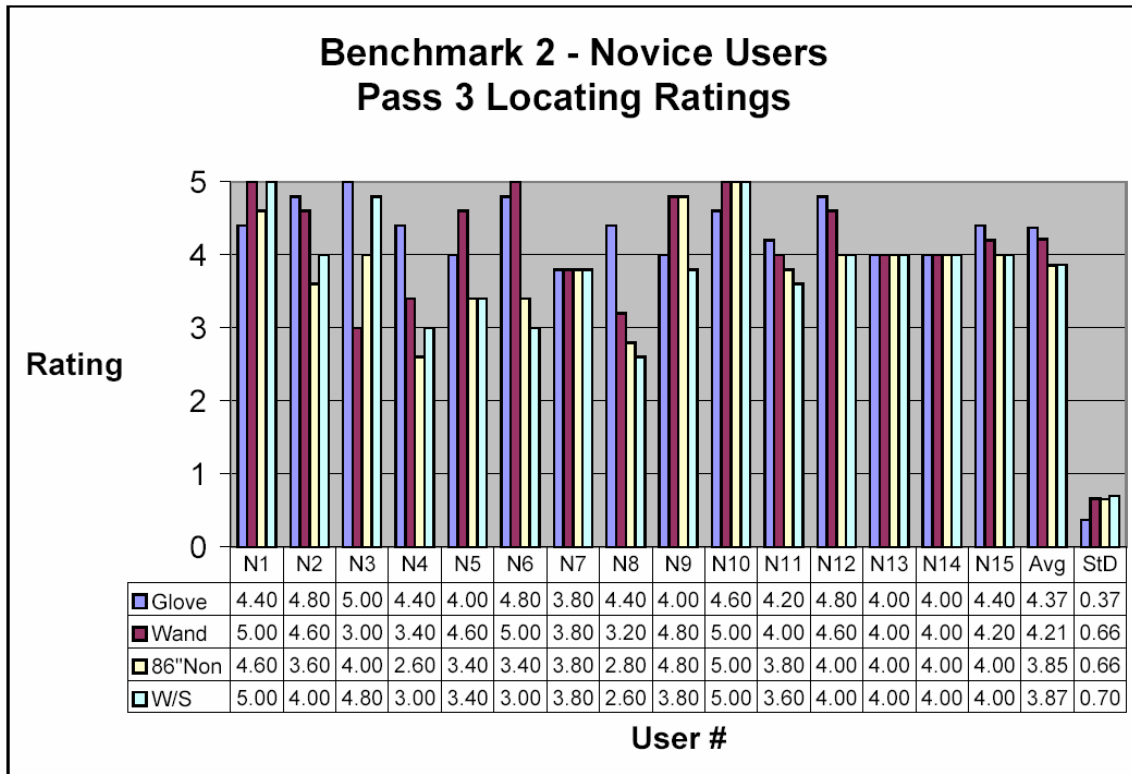
Benchmark 2, Novice Users, Pass 3, Elapsed Times												
B2Np3	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	CV
Glove	15	171.3		27.61		120.0		210.0		0.0628	Yes	16.12%
Wand	15	154.2		27.40		115.0		195.0		>0.1000	Yes	17.77%
86"Non	15	250.0		56.22		130.0		350.0		>0.1000	Yes	22.49%
W/S	15	229.3		37.22		140.0		280.0		> 0.1000	Yes	16.23%
B2Np3		Homogeneity of Variance				Roy's Greatest Root:				F(3,12) 27.26 Pr > F <0.0001		
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		0.0008	0.9776			Yes	14	5.35	0.0001	No	Means	Wand 9.98%
Glove vs 86"Non		6.3265	0.0119			No	14	34.96	0.0001	No	Var/Means	Glove 31.48%
Glove vs W/S		1.1887	0.2758			Yes	14	63.99	0.0001	No	Means	Glove 25.29%
Wand vs 86"Non		6.4531	0.0111			No	14	29.23	0.0001	No	Var/Means	Wand 31.48%
Wand vs W/S		1.2482	0.2639			Yes	14	69.44	0.0001	No	Means	Wand 32.75%
86"Non vs W/S		2.2377	0.1347			Yes	14	2.26	0.1546	Yes	Neither	Neither



**Figure D- 14: B2Np3Nav Novice User Navigation Ratings**

**Table D- 14: B2Np3Nav Novice User Navigation Ratings Statistics**

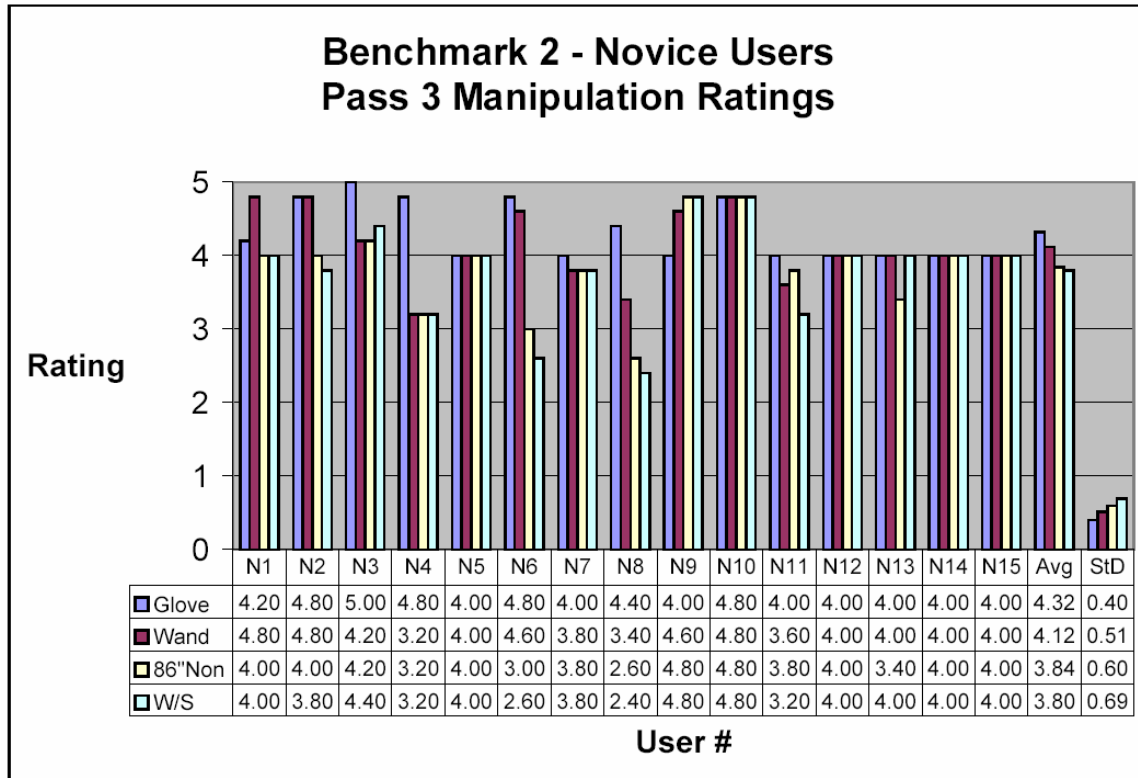
Benchmark 2, Novice Users, Pass 3, Navigation Ratings										
B2Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	15	4.39	0.47	3.71	5.00	> 0.1000	Yes	10.74%		
Wand	15	4.25	0.53	3.14	5.00	> 0.1000	Yes	12.58%		
86\"Non	15	3.89	0.63	2.71	5.00	> 0.1000	Yes	16.29%		
W/S	15	3.79	0.68	2.70	5.00	0.0145	No	17.98%		
B2Np3		Homogeneity of Variance				Roy's Greatest Root: F(3,12)		3.63	Pr > F	0.045
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					
Glove vs Wand				0.3500	0.5608	Yes	14	0.74	0.4037	Yes
Glove vs 86\"Non				1.4700	0.2352	Yes	14	6.63	0.0220	No
Glove vs W/S				2.0200	0.1663	Yes	14	8.47	0.0114	No
Wand vs 86\"Non		0.3826	0.5362			Yes	14	9.14	0.0091	No
Wand vs W/S		0.7875	0.3749			Yes	14	11.49	0.0044	No
86\"Non vs W/S		0.0738	0.7859			Yes	14	3.72	0.0745	No
										Significant?
										Statistically Better
										Neither
										Means
										Glove
										12.97%
										Means
										Glove
										15.80%
										Means
										Wand
										12.10%
										Means
										Wand
										12.10%
										Means
										86\"Non
										2.51%



**Figure D- 15: B2Np3Loc Novice User Locating Ratings**

**Table D- 15: B2Np3Loc Novice User Locating Ratings Statistics**

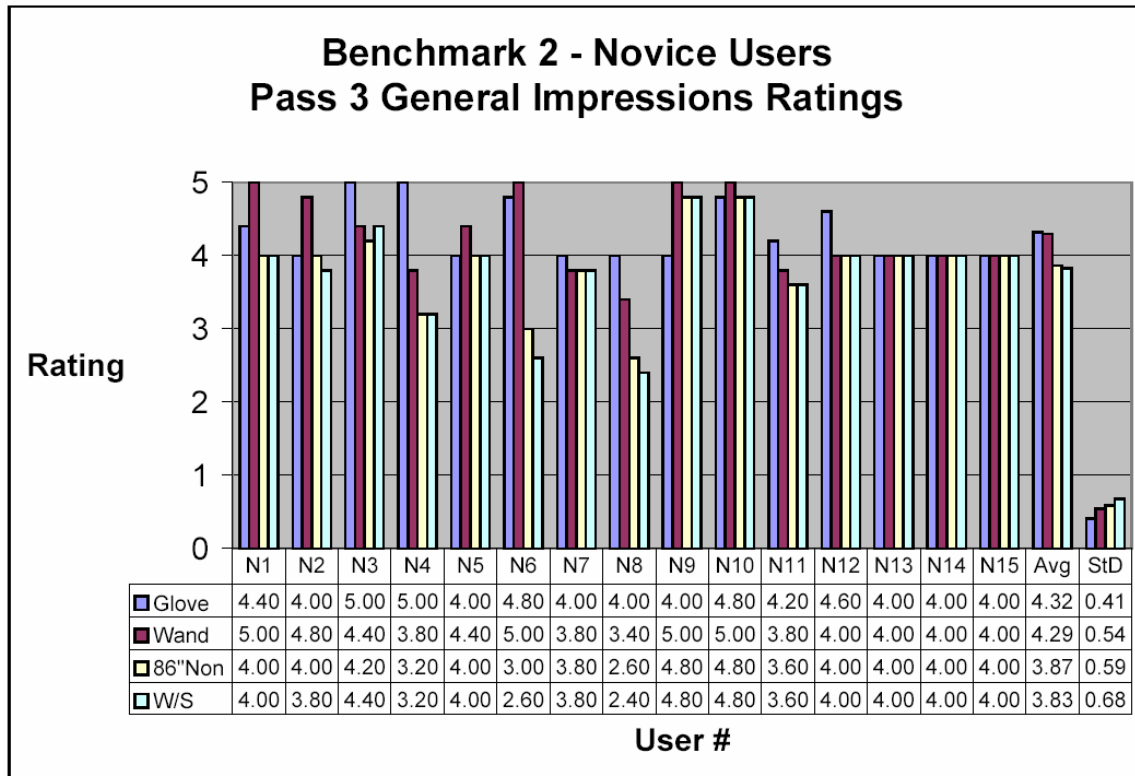
Benchmark 2, Novice Users, Pass 3, Locating Ratings											
B2Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.37	0.37	3.80	5.00	> 0.1000	Yes	8.45%			
Wand	15	4.21	0.66	3.00	5.00	> 0.1000	Yes	15.69%			
86"Non	15	3.85	0.66	2.60	5.00	0.0693	Yes	17.04%			
W/S	15	3.87	0.70	2.60	5.00	0.0399	No	18.20%			
B2Np3	Homogeneity of Variance				Roy's Greatest Root: F(3,12)				2.41	Pr > F 0.1174	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	4.3440	0.0371			No	14	0.69	0.4197	Yes	Var	Glove 3.80%
Glove vs 86"Non	4.2511	0.0392			No	14	6.84	0.0204	No	Var/Mean	Glove 13.50%
Glove vs W/S			4.9000	0.0352	No	14	7.34	0.0170	No	Var/Mean	Glove 13.09%
Wand vs 86"Non	0.0006	0.9807			Yes	14	4.99	0.0424	No	Mean	Neither
Wand vs W/S			0.0900	0.7702	Yes	14	2.71	0.1221	Yes	Neither	Wand 8.95%
86"Non vs W/S			0.0900	0.7727	Yes	14	0.02	0.9016	Yes	Neither	Neither



**Figure D- 16: B2Np3Mov Novice User Manipulation Ratings**

**Table D- 16: B2Np3Mov Novice User Manipulation Ratings Statistics**

Benchmark 2, Novice Users, Pass 3, Manipulation Ratings											
B2Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.32	0.40	4.00	5.00	0.0002	No	9.23%			
Wand	15	4.12	0.51	3.20	4.80	>0.1000	Yes	12.42%			
86\"Non	15	3.84	0.60	2.60	4.80	0.0846	Yes	15.53%			
W/S	15	3.80	0.69		4.80	0.0273	No	18.23%			
B2Np3		Homogeneity of Variance				Roy's Greatest Root: F(3,12)			1.85	Pr > F	0.1919
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						Statistically Better
Glove vs Wand				2.3400	0.1377	Yes	14	1.84	0.1962	Yes	Neither
Glove vs 86\"Non				2.1300	0.1554	Yes	14	6.06	0.0274	No	Means Glove 12.50%
Glove vs W/S				3.4400	0.0742	Yes	14	5.64	0.0324	No	Means Glove 13.68%
Wand vs 86\"Non		0.3153	0.5744			Yes	14	4.40	0.0546	No	Means Wand 7.29%
Wand vs W/S				1.4000	0.2466	Yes	14	4.02	0.0646	No	Means Wand 8.42%
86\"Non vs W/S				1.7500	0.1960	Yes	14	0.34	0.5667	Yes	Neither

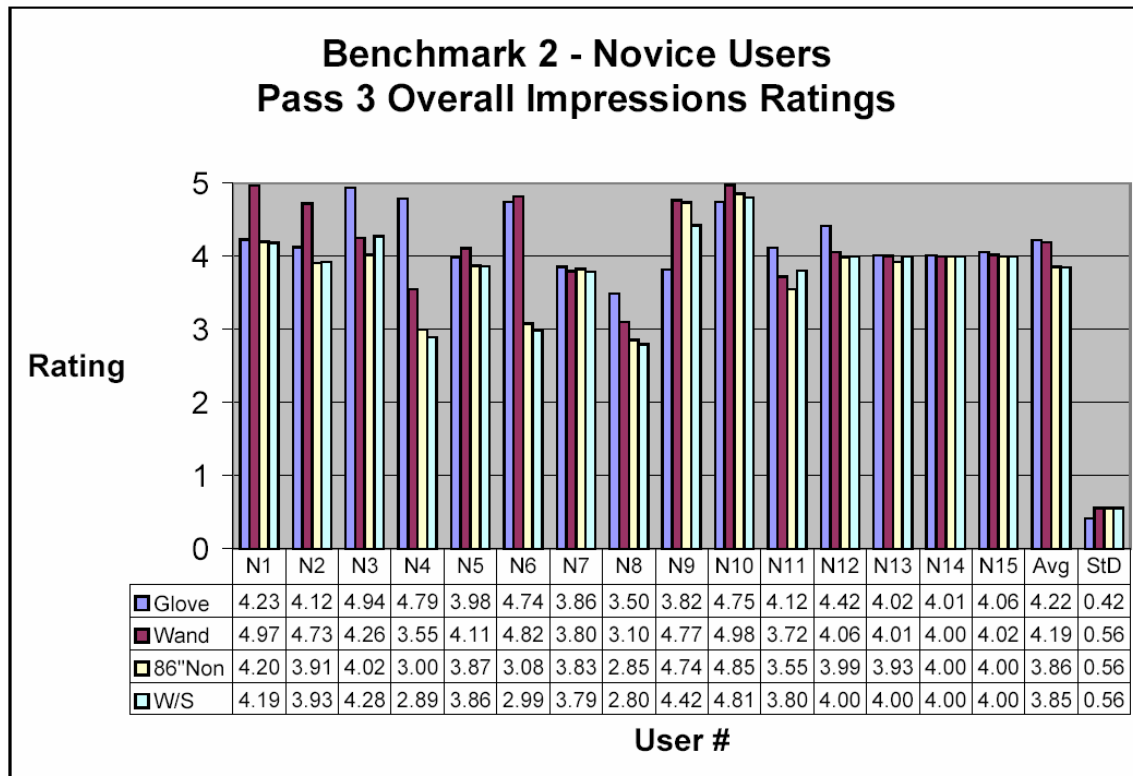


**Figure D- 17: B2Np3Gen Novice User General Impressions Ratings**

**Table D- 17: B2Np3Gen Novice User General Impressions Ratings Statistics**

Benchmark 2, Novice Users, Pass 3, General Impressions Ratings												
B2Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.32	0.41	4.00	5.00	0.0002	No	9.39%				
Wand	15	4.29	0.54	3.40	5.00	0.0217	No	12.68%				
86\"Non	15	3.87	0.59	2.60	4.80	0.009	No	15.23%				
W/S	15	3.83	0.68	2.40	4.80	0.0543	Yes	17.65%				
B2Np3		Homogeneity of Variance				Roy's Greatest Root: F(3,12)				3.80	Pr > F	0.0397
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				3.0800	0.0900	Yes	14	0.03	0.8646	Yes	Neither	Neither
Glove vs 86\"Non				1.7000	0.2032	Yes	14	5.71	0.6315	Yes	Neither	Neither
Glove vs W/S				2.5900	0.1189	Yes	14	5.71	0.0315	No	Means	Glove 12.88%
Wand vs 86\"Non				0.1100	0.7383	Yes	14	9.04	0.0094	No	Means	Wand 11.02%
Wand vs W/S				0.7200	0.4036	Yes	14	7.46	0.0163	No	Means	Wand 12.18%
86\"Non vs W/S				0.2400	0.6257	Yes	14	1.31	0.2711	Yes	Neither	Neither

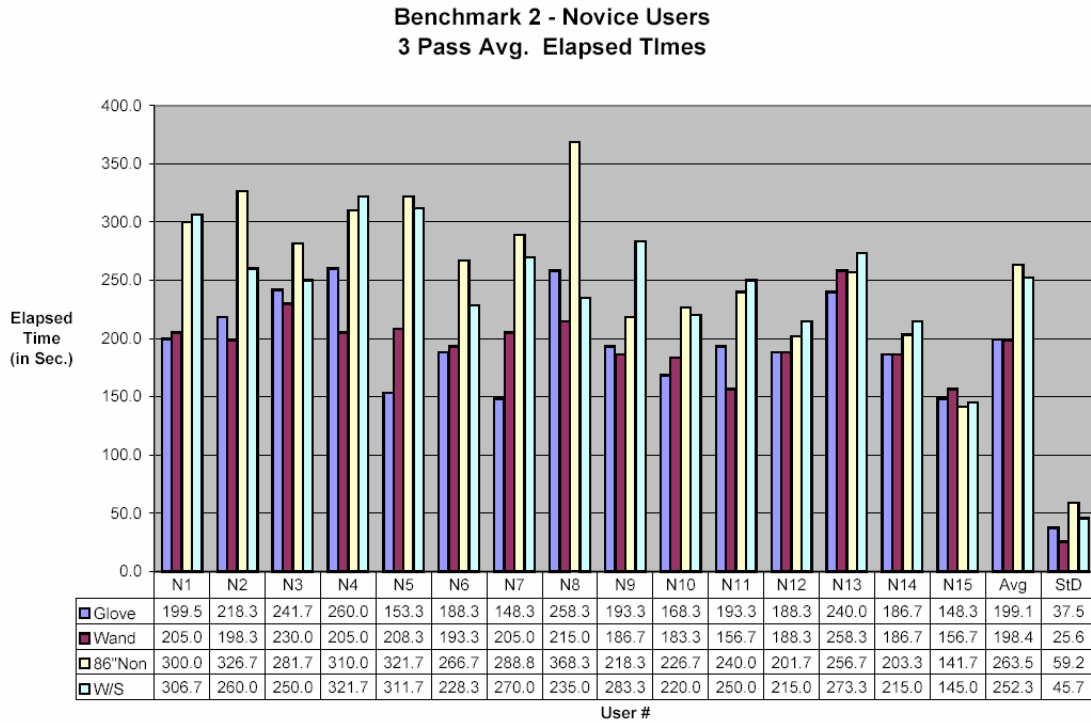




**Figure D- 18: B2Np3Ovr Novice User Overall Impressions Ratings**

**Table D- 18: B2Np3Ovr Novice User Overall Impressions Ratings Statistics**

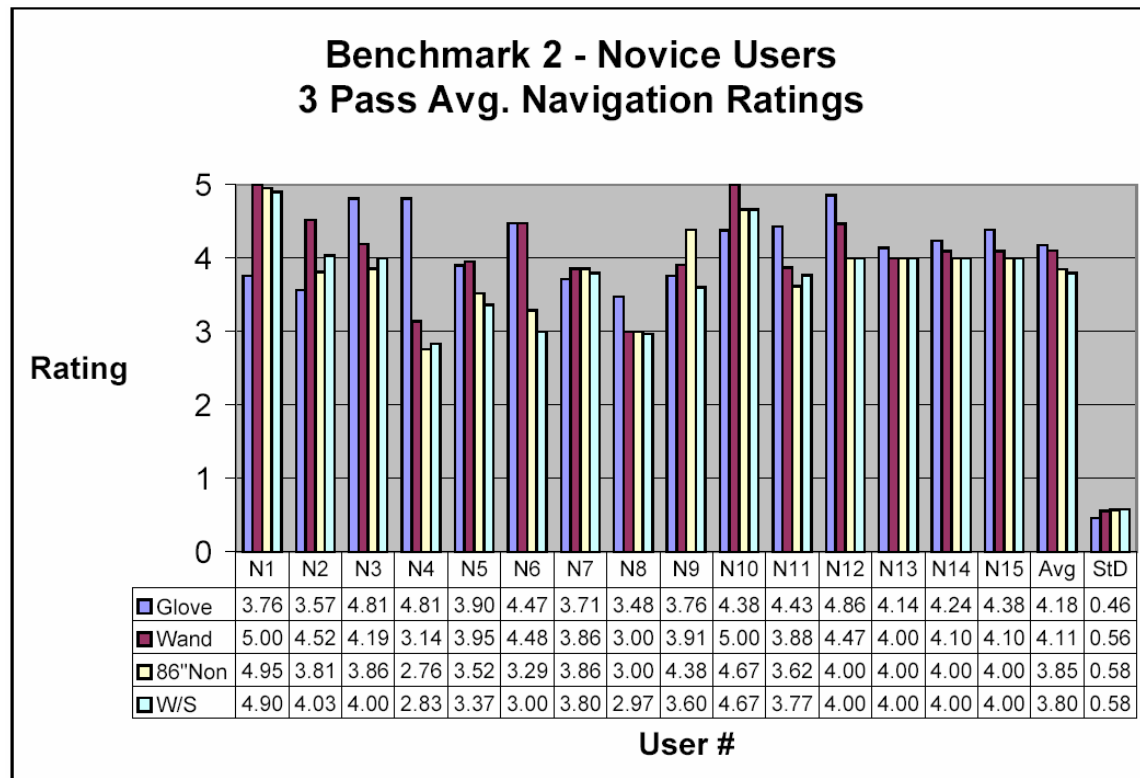
Benchmark 2, Novice Users, Pass 3, Overall Impressions Ratings											
B2Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.22	0.42	3.50	4.94	> 0.1000	Yes	9.86%			
Wand	15	4.19	0.56	3.10	4.98	>0.1000	Yes	13.30%			
86\"Non	15	3.86	0.56	2.85	4.85	0.0584	Yes	14.58%			
W/S	15	3.85	0.56	2.80	4.81	0.0087	No	14.58%			
B2Np3		Homogeneity of Variance				Roy's Greatest Root: F(3,12)			2.53	Pr > F	0.1069
		Bartlett's Test		Levene's Test		df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand		1.1334	0.2870			Yes	14	0.05	0.8341	Yes	Neither
Glove vs 86\"Non		1.1958	0.2742			Yes	14	4.33	0.0563	No	Means
Glove vs W/S				1.3300	0.2593	Yes	14	4.71	0.0476	No	Means
Wand vs 86\"Non		0.0009	0.9765			Yes	14	7.74	0.0147	No	Means
Wand vs W/S				0.0000	0.9767	Yes	14	6.88	0.0201	No	Means
86\"Non vs W/S				0.0000	0.9962	Yes	14	0.01	0.9117	Yes	Neither



**Figure D- 19: B2N3pAvgTim Novice Users Elapsed Times**

**Table D- 19: B2N3pAvgTim Novice Users Elapsed Times Statistics**

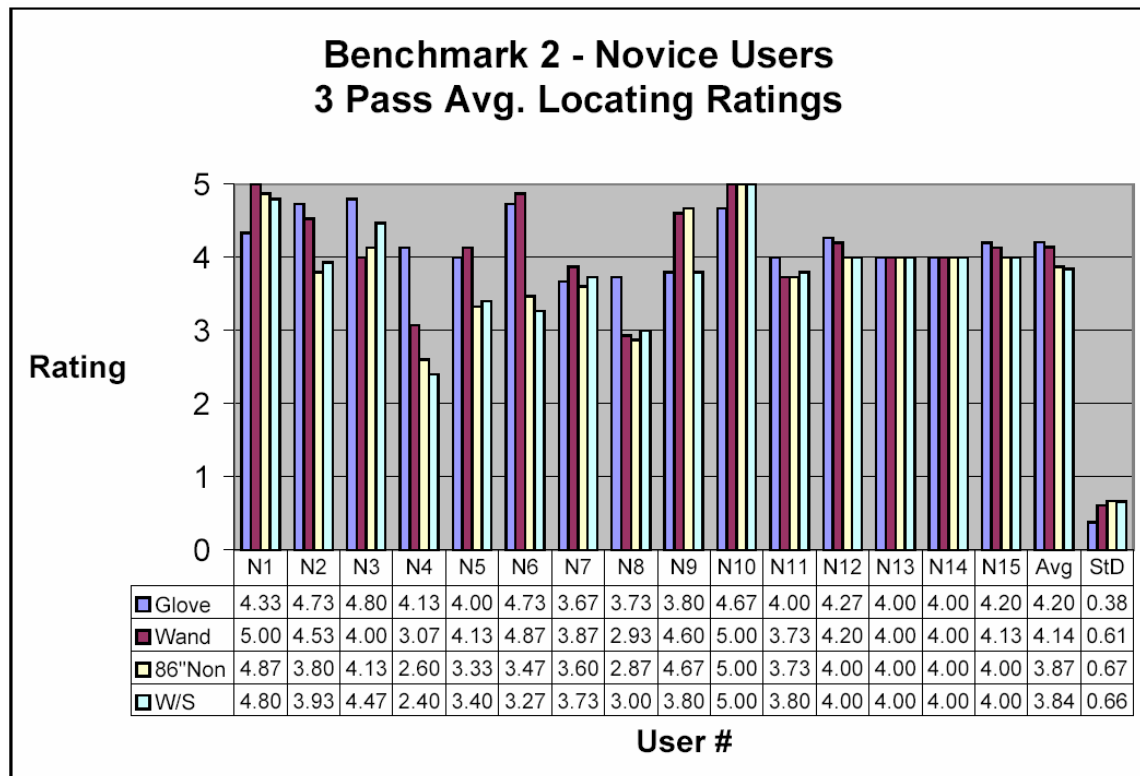
Benchmark 2, Novice Users, 3 Pass Avg., Elapsed Times											
B2N3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	199.1	37.54	148.3	260.0	> 0.1000	Yes	18.85%			
Wand	15	198.4	25.63	156.7	258.3	>0.1000	Yes	12.92%			
86\"Non	15	263.5	59.23	141.7	368.3	>0.1000	Yes	22.48%			
W/S	15	252.3	45.75	145.0	321.7	> 0.1000	Yes	18.13%			
B2N3pA		Homogeneity of Variance				Roy's Greatest Root:				F(3,12)	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand		1.9223	0.1656			Yes	14	0.01	0.9400	Yes	Neither
Glove vs 86\"Non		2.7180	0.0992			Yes	14	23.54	0.0003	No	Means
Glove vs W/S		0.5252	0.4687			Yes	14	17.95	0.0008	No	Means
Wand vs 86\"Non		8.5452	0.0035			No	14	25.37	0.0001	No	Var/Means
Wand vs W/S		4.3037	0.0380			No	14	27.23	0.3541	Yes	Var
86\"Non vs W/S		0.8915	0.3451			Yes	14	0.92	0.0213	No	Means
											Statistically Better
											Neither
											Glove 24.44%
											Wand 21.09%
											Wand 24.44%
											Wand 21.36%
											W/S 4.25%



**Figure D- 20: B2N3pAvgNav Novice Users Navigation Ratings**

**Table D- 20: B2N3pAvgNav Novice Users Navigation Ratings Statistics**

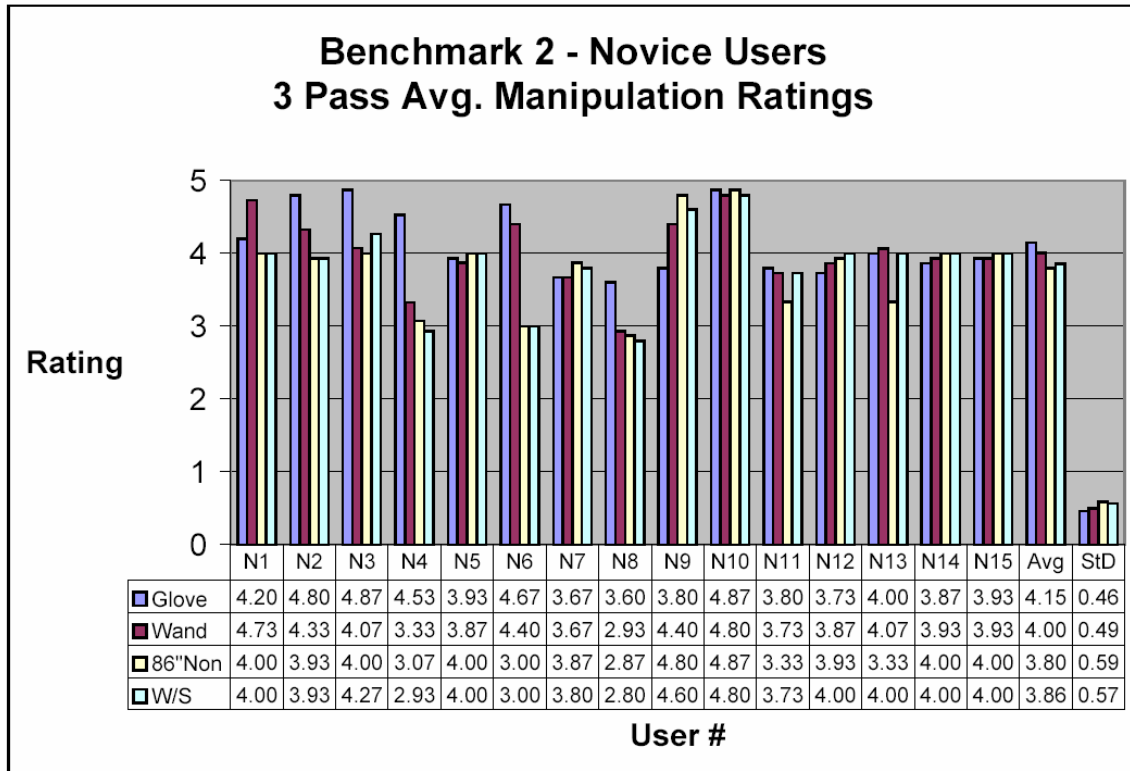
Benchmark 2, Novice Users, 3 Pass Avg., Navigation Ratings												
B2N3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.18	0.46	3.48	4.86	> 0.1000	Yes	11.08%				
Wand	15	4.11	0.56	3.00	5.00	> 0.1000	Yes	13.64%				
86*Non	15	3.85	0.58	2.76	4.95	> 0.1000	Yes	14.96%				
W/S	15	3.77	0.58	2.83	4.90	0.0737	Yes	15.40%				
B2N3pA		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	3.14	Pr > F		0.0651
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		0.4858	0.4858			Yes	14	0.17	0.6891	Yes	Neither	Neither
Glove vs 86*Non		0.6362	0.4251			Yes	14	2.63	0.1272	Yes	Neither	Neither
Glove vs W/S		0.6829	0.4086			Yes	14	3.83	0.0707	No	Means	Glove 10.90%
Wand vs 86*Non		0.0103	0.9190			Yes	14	7.04	0.0189	No	Means	Wand 8.97%
Wand vs W/S		0.0171	0.8959			Yes	14	10.53	0.0059	No	Means	Wand 8.97%
86*Non vs W/S		0.0008	0.9768			Yes	14	0.73	0.4070	Yes	Neither	Neither



**Figure D- 21: B2N3pAvgLoc Novice User Locating Ratings**

**Table D- 21: B2N3pAvgLoc Novice User Locating Ratings Statistics**

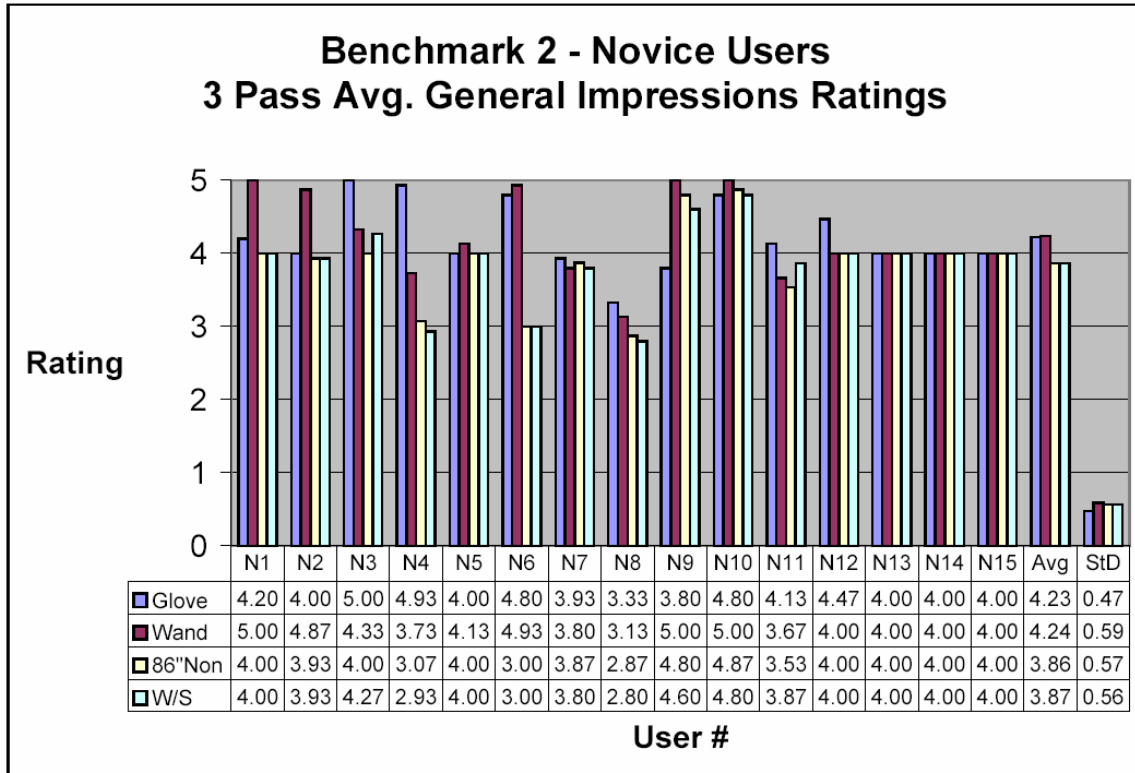
Benchmark 2, Novice Users, 3 Pass Avg., Locating Ratings										
B2N3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	15	4.20	0.38	3.67	4.80	> 0.1000	Yes	8.98%		
Wand	15	4.14	0.61	2.93	5.00	> 0.1000	Yes	14.86%		
86\"Non	15	3.87	0.67	2.60	5.00	> 0.1000	Yes	17.21%		
W/S	15	3.84	0.66	2.40	5.00	0.0929	Yes	17.17%		
Homogeneity of Variance				Roy's Greatest Root: F(3,12)				1.96	Pr > F	0.1733
B2N3pA		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					
Glove vs Wand		3.1042	0.0781			Yes	14	0.25	0.6249	Yes
Glove vs 86\"Non		1.4381	0.2305			Yes	14	3.80	0.0716	No
Glove vs W/S		1.3495	0.2454			Yes	14	5.28	0.0374	No
Wand vs 86\"Non		1.4594	0.2270			Yes	14	6.03	0.0278	No
Wand vs W/S		0.0015	0.9693			Yes	14	5.21	0.0386	No
86\"Non vs W/S		0.0001	0.9927			Yes	14	0.21	0.6572	Yes
										Significant?
										Statistically Better
										Neither
										Means
										Glove
										8.50%
										Means
										Glove
										9.38%
										Means
										Wand
										6.87%
										Means
										Wand
										7.73%
										Neither



**Figure D- 22: B2N3pAvgMov Novice User Manipulation Ratings**

**Table D- 22: B2N3pAvgMov Novice User Manipulation Ratings**

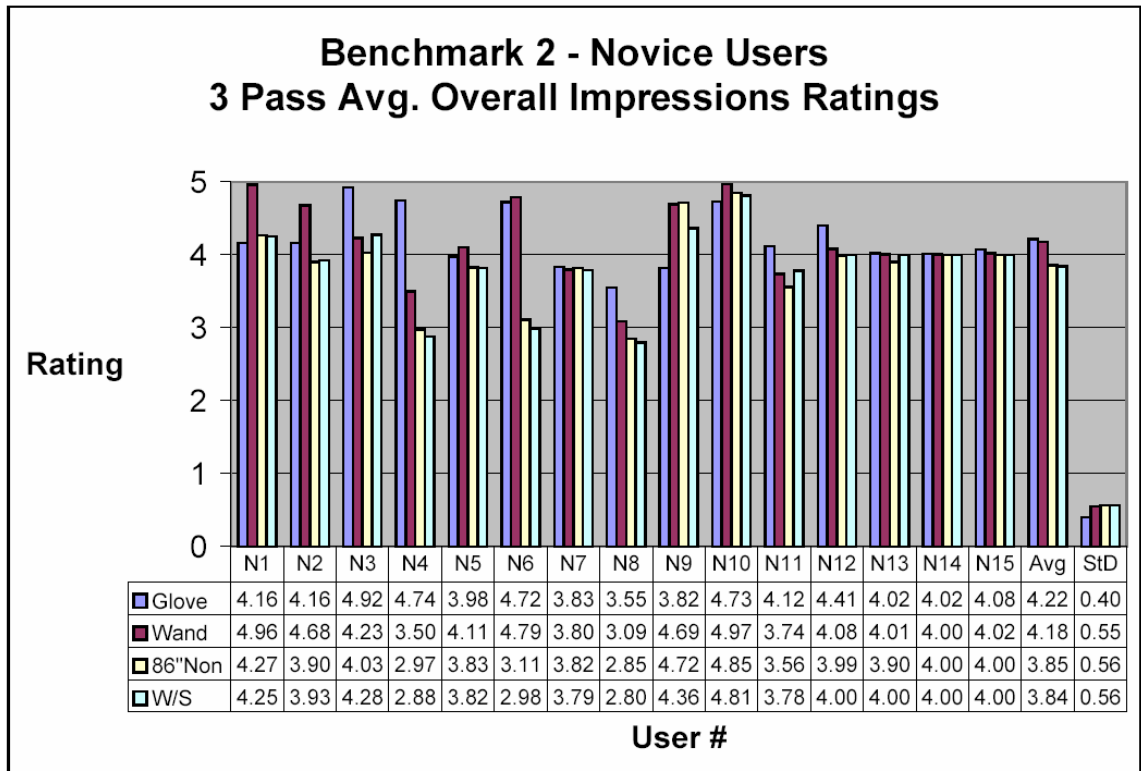
Benchmark 2, Novice Users, 3 Pass Avg., Manipulation Ratings												
B2N3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.15	0.47	3.60	4.87	0.0354	No	11.20%				
Wand	15	4.00	0.50	2.93	4.80	> 0.1000	Yes	12.36%				
86\"Non	15	3.80	0.59	2.87	4.87	0.0267	No	15.51%				
W/S	15	3.86	0.57	2.80	4.80	0.0538	Yes	14.65%				
Homogeneity of Variance					Roy's Greatest Root: F(3,12)			1.17	Pr > F	0.3616		
B2N3pA		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.0800	0.7815	Yes	14	1.43	0.2510	Yes	Neither	Neither
Glove vs 86\"Non				1.1700	0.2894	Yes	14	3.74	0.0736	No	Means	Glove 9.24%
Glove vs W/S				0.7100	0.4098	Yes	14	2.74	0.1199	Yes	Neither	Neither
Wand vs 86\"Non				0.5100	0.4832	Yes	14	2.87	0.1125	Yes	Neither	Neither
Wand vs W/S		0.2361	0.6270			Yes	14	1.70	0.2129	Yes	Neither	Neither
86\"Non vs W/S				0.0300	0.8609	Yes	14	0.97	0.3414	Yes	Neither	Neither



**Figure D- 23: B2N3pAvgGen Novice User General Impressions Ratings**

**Table D- 23: B2N3pAvgGen Novice User General Impressions Ratings Statistics**

Benchmark 2, Novice Users, 3 Pass Avg., General Impressions Ratings												
B2N3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.23	0.47	3.33	5.00	0.057	Yes	11.23%				
Wand	15	4.24	0.59	3.13	5.00	> 0.1000	Yes	13.91%				
86"Non	15	3.86	0.57	2.87	4.87	0.0041	No	14.68%				
W/S	15	3.87	0.56	2.80	4.80	0.0107	No	14.58%				
B2N3pA		Homogeneity of Variance				Roy's Greatest Root: F(3,12)			2.09	Pr > F		0.1544
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		0.6344	0.4257			Yes	14	0.01	0.9342	Yes	Neither	Neither
Glove vs 86"Non				0.5200	0.4759	Yes	14	3.66	0.0763	No	Means	Glove 9.40%
Glove vs W/S				0.4900	0.4906	Yes	14	3.82	0.0711	No	Means	Glove 9.28%
Wand vs 86"Non				0.0300	0.8588	Yes	14	7.11	0.0184	No	Means	Wand 9.73%
Wand vs W/S				0.0400	0.8415	Yes	14	6.43	0.0238	No	Means	Wand 9.62%
86"Non vs W/S				0.0000	0.9841	Yes	14	0.01	0.9115	Yes	Neither	Neither



**Figure D- 24: B2N3pAvgOvr Novice User Overall Impressions Ratings**

**Table D- 24: B2N3pAvgOvr Novice User Overall Impressions Ratings Statistics**

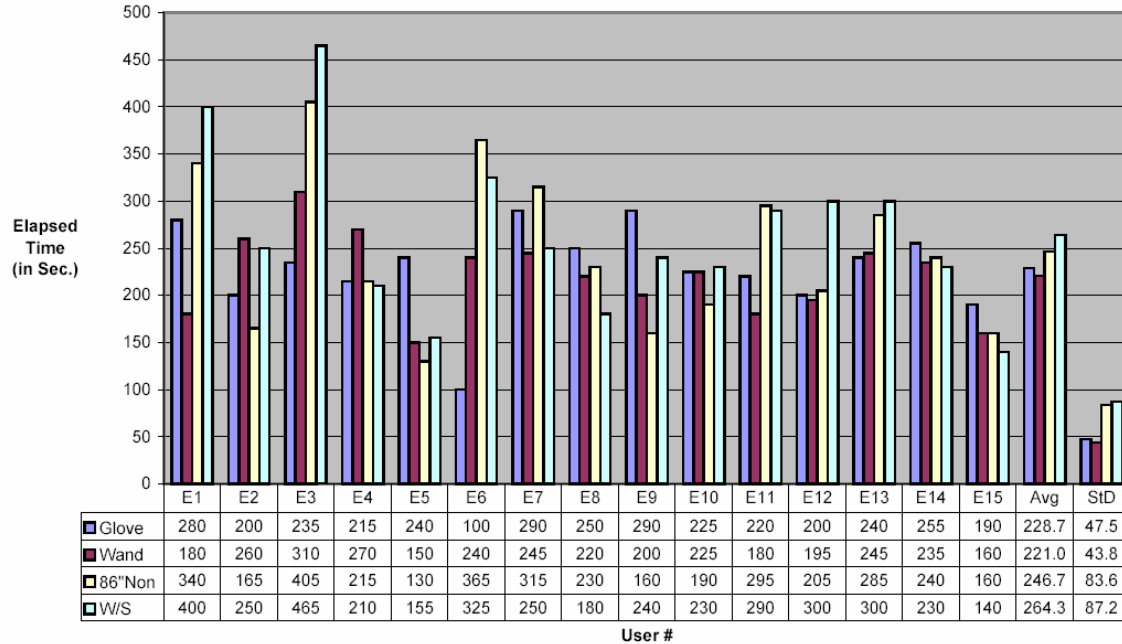
Benchmark 2, Novice Users, 3 Pass Avg., Overall Impressions Ratings												
B2N3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.22	0.40	3.55	4.92	0.0421	No	9.49%				
Wand	15	4.18	0.55	3.09	4.97	> 0.1000	Yes	13.12%				
86"Non	15	3.85	0.56	2.85	4.85	0.0747	Yes	14.62%				
W/S	15	3.85	0.56	2.80	4.81	0.0102	No	14.63%				
B2N3pA		Homogeneity of Variance				Roy's Greatest Root: F(3,12)			2.29	Pr > F 0.1301		
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				1.9000	0.1795	Yes	14	0.08	0.7852	Yes	Neither	Neither
Glove vs 86"Non				1.7600	0.1957	Yes	14	4.31	0.0568	No	Means	Glove 9.44%
Glove vs W/S				1.6700	0.2071	Yes	14	4.79	0.0460	No	Means	Glove 9.67%
Wand vs 86"Non		0.0100	0.9204			Yes	14	7.76	0.0146	No	Means	Wand 8.43%
Wand vs W/S				0.0100	0.9126	Yes	14	6.87	0.0201	No	Means	Wand 8.65%
86"Non vs W/S				0.0000	0.9963	Yes	14	0.05	0.8301	Yes	Neither	Neither

## **Appendix E**

### **Experienced User Benchmark 2 (Find/Repair) Detail**



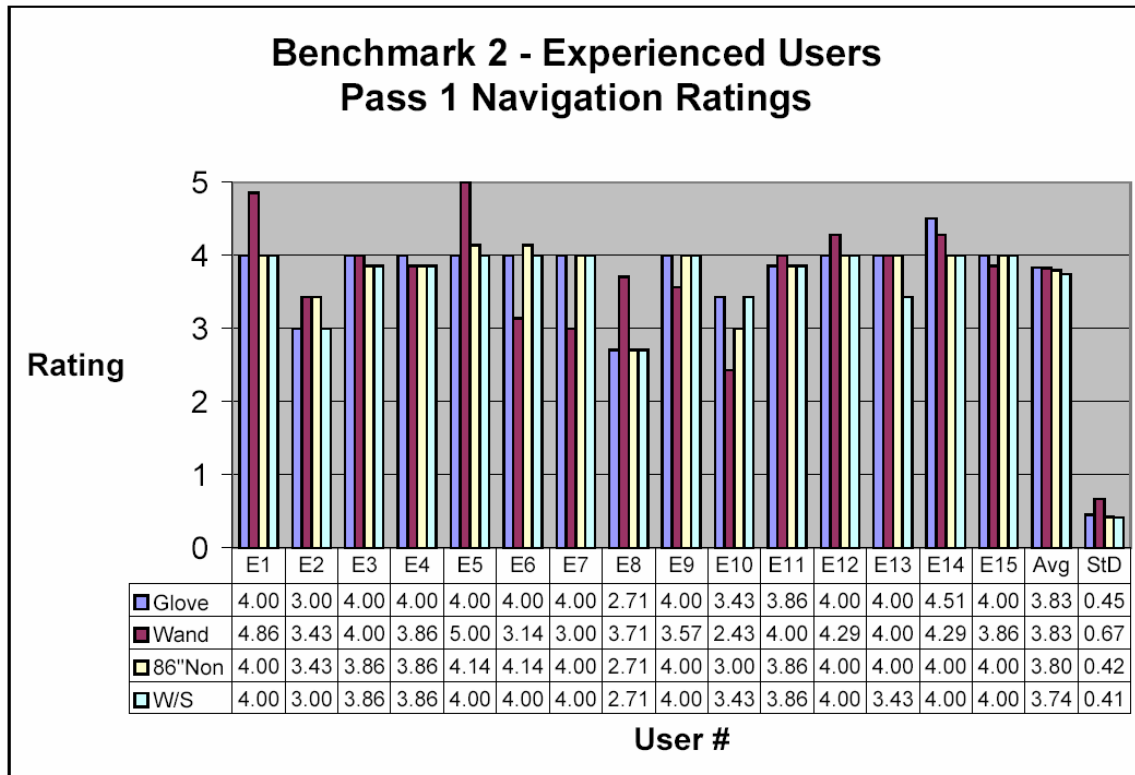
**Benchmark 2 - Experienced Users  
Pass 1 Elapsed Times**



**Figure E- 1: B2Ep1Tim Experienced User Elapsed Times**

**Table E- 1: B2Ep1Tim Experienced User Elapsed Times Statistics**

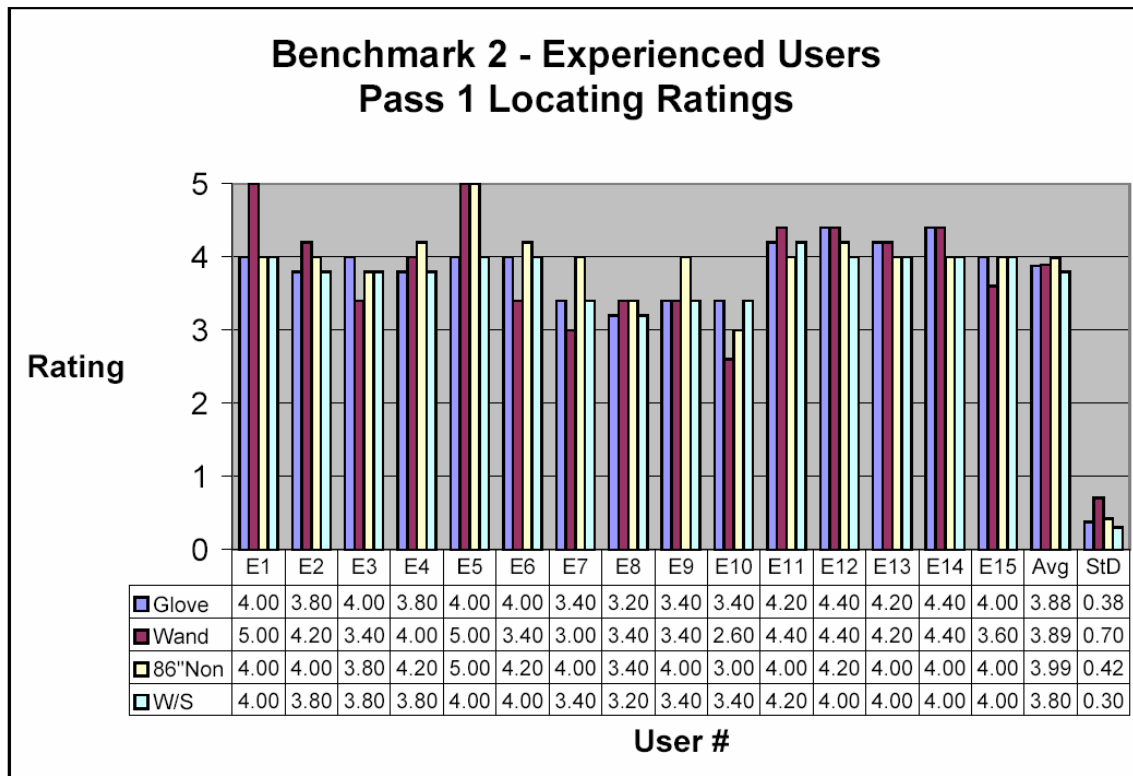
Benchmark 2, Experienced Users, Pass 1, Elapsed Times											
B2Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	228.7	47.49	100.0	290.0	> 0.1000	Yes	20.77%			
Wand	15	221.0	43.76	150.0	310.0	> 0.1000	Yes	19.80%			
86\"Non	15	246.7	83.59	130.0	405.0	>0.1000	Yes	33.88%			
W/S	15	264.3	87.18	140.0	465.0	> 0.1000	Yes	32.99%			
Homogeneity of Variance					Roy's Greatest Root: F(3,12)			1.42	Pr > F	0.286	
B2Ep1	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.0903	0.7638			Yes	29	0.20	0.6649	Yes	Neither	Neither
Glove vs 86\"Non	4.1099	0.0426			Yes	29	0.49	0.4958	Yes	Neither	Neither
Glove vs W/S	4.7077	0.0300			No	29	1.92	0.1877	Yes	Var	Glove 13.47%
Wand vs 86\"Non	5.3066	0.0212			No	29	1.83	0.1976	Yes	Var	Wand 10.42%
Wand vs W/S	5.9684	0.0146			No	29	4.64	0.0492	No	Var/Mean	Wand 16.38%
86\"Non vs W/S	0.0238	0.8775			Yes	29	1.80	0.2015	Yes	Neither	Neither



**Figure E- 2: B2Ep1Nav Experienced User Navigation Ratings**

**Table E- 2: B2Ep1Nav Experienced User Navigation Ratings Statistics**

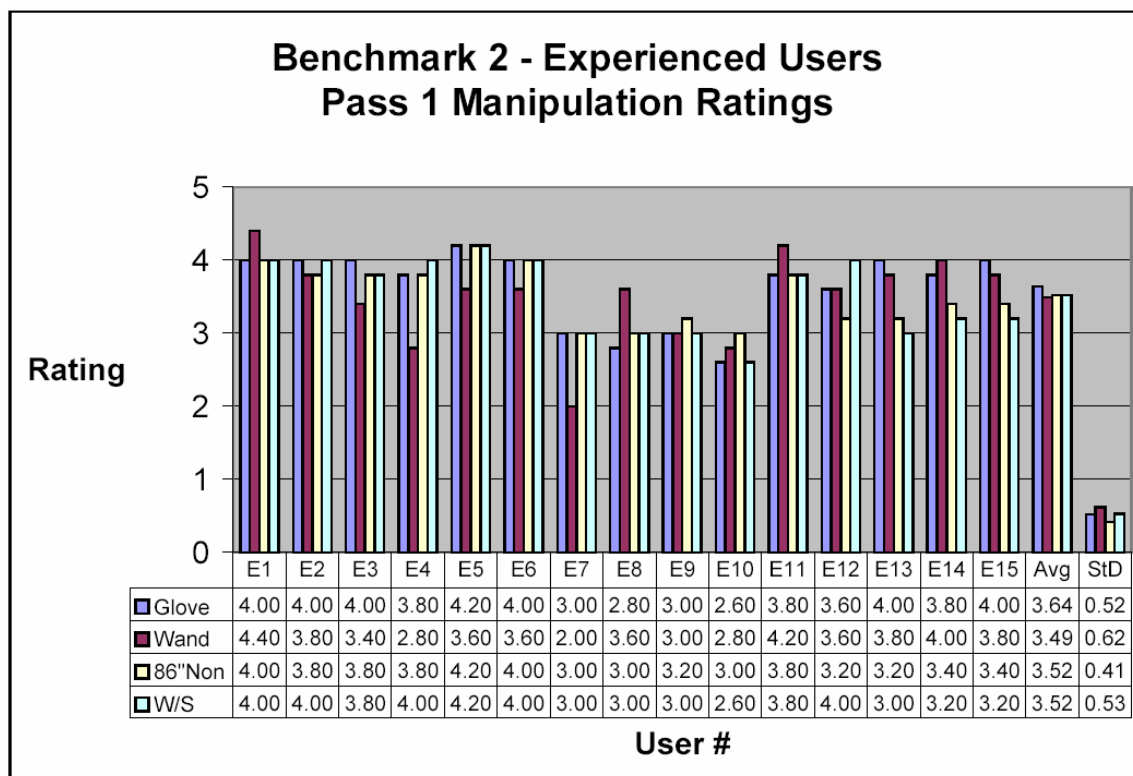
Benchmark 2, Experienced Users, Pass 1, Navigation Ratings											
B2Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	3.83	0.45	2.71	4.51	< 0.0001	No	11.78%			
Wand	15	3.83	0.67	2.43	5.00	> 0.1000	Yes	17.53%			
86"Non	15	3.80	0.42	2.71	4.14	< 0.0001	No	11.09%			
W/S	15	3.47	0.41	2.71	4.00	< 0.0001	No	11.88%			
B2Ep1	Homogeneity of Variance					Roy's Greatest Root: F(3,12)		1.00	Pr > F	0.4264	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			1.6400	0.2109	Yes	29	0.00	0.9784	Yes	Neither	Neither
Glove vs 86"Non			0.0400	0.8448	Yes	29	0.35	0.5628	Yes	Neither	Neither
Glove vs W/S			0.0700	0.7914	Yes	29	3.45	0.0846	No	Means	Glove 10.39%
Wand vs 86"Non			2.1100	0.1579	Yes	29	0.03	0.8548	Yes	Neither	Neither
Wand vs W/S			2.3400	0.1370	Yes	29	0.25	0.6270	Yes	Neither	Neither
86"Non vs W/S			0.0000	0.9514	Yes	29	0.98	0.3382	Yes	Neither	Neither



**Figure E- 3: B2Ep1Loc Experienced User Locating Ratings**

**Table E- 3: B2Ep1Loc Experienced User Locating Ratings Statistics**

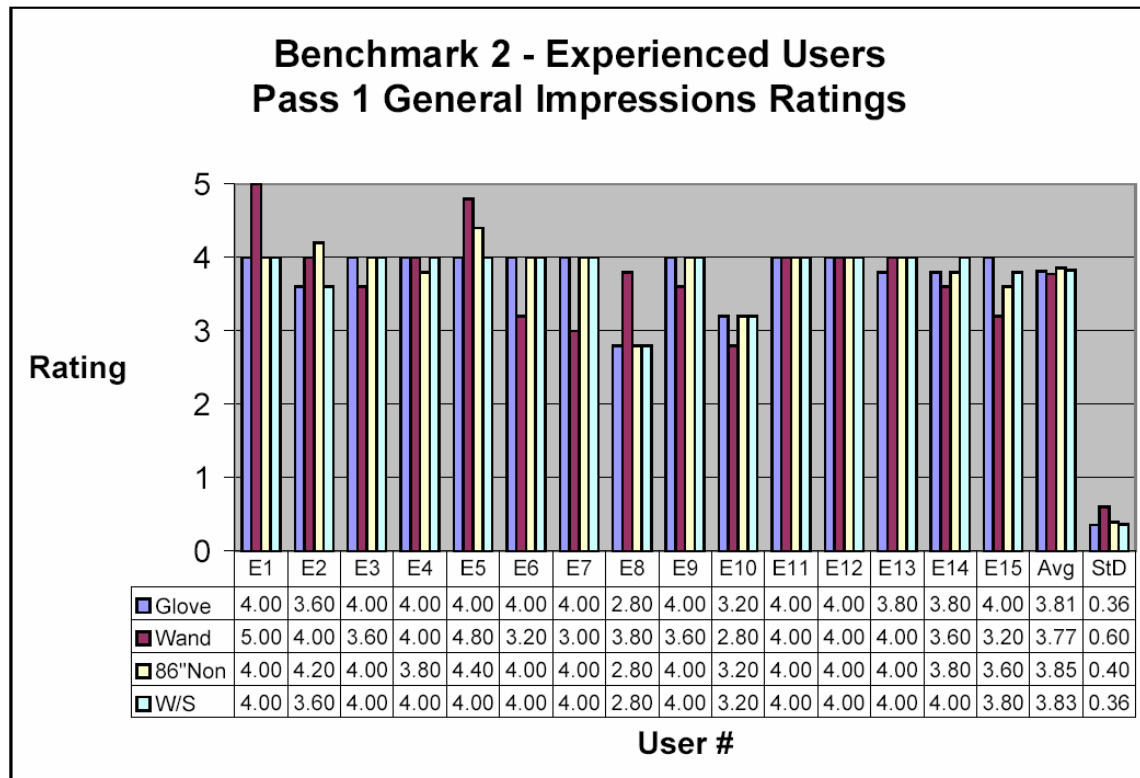
Benchmark 2, Experienced Users, Pass 1, Locating Ratings											
B2Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	3.88	0.38	3.20	4.40	0.0396	No	9.70%			
Wand	15	3.89	0.70	2.60	5.00	> 0.1000	Yes	18.10%			
86\"Non	15	3.99	0.42	3.00	5.00	0.0003	No	10.63%			
W/S	15	3.80	0.30	3.20	4.20	0.0026	No	7.96%			
B2Ep1	Homogeneity of Variance				Roy's Greatest Root: F(3,12)				3.25	Pr > F	0.0602
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			5.8200	0.0226	No	29	0.01	0.9240	Yes	Var	Wand 0.34%
Glove vs 86\"Non			0.1300	0.7173	Yes	29	1.03	0.3263	Yes	Neither	Neither
Glove vs W/S			1.1200	0.2989	Yes	29	4.42	0.0541	No	Means	Glove 2.11%
Wand vs 86\"Non			3.4100	0.0755	Yes	29	0.48	0.5008	Yes	Neither	Neither
Wand vs W/S			7.8800	0.0090	Yes	29	0.46	0.5093	Yes	Neither	Neither
86\"Non vs W/S			0.7800	0.3858	Yes	29	4.26	0.0580	No	Means	86\"Non 4.92%



**Figure E- 4: B2Ep1Mov Experienced User Manipulation Ratings**

**Table E- 4: B2Ep1Mov Experienced User Manipulation Ratings Statistics**

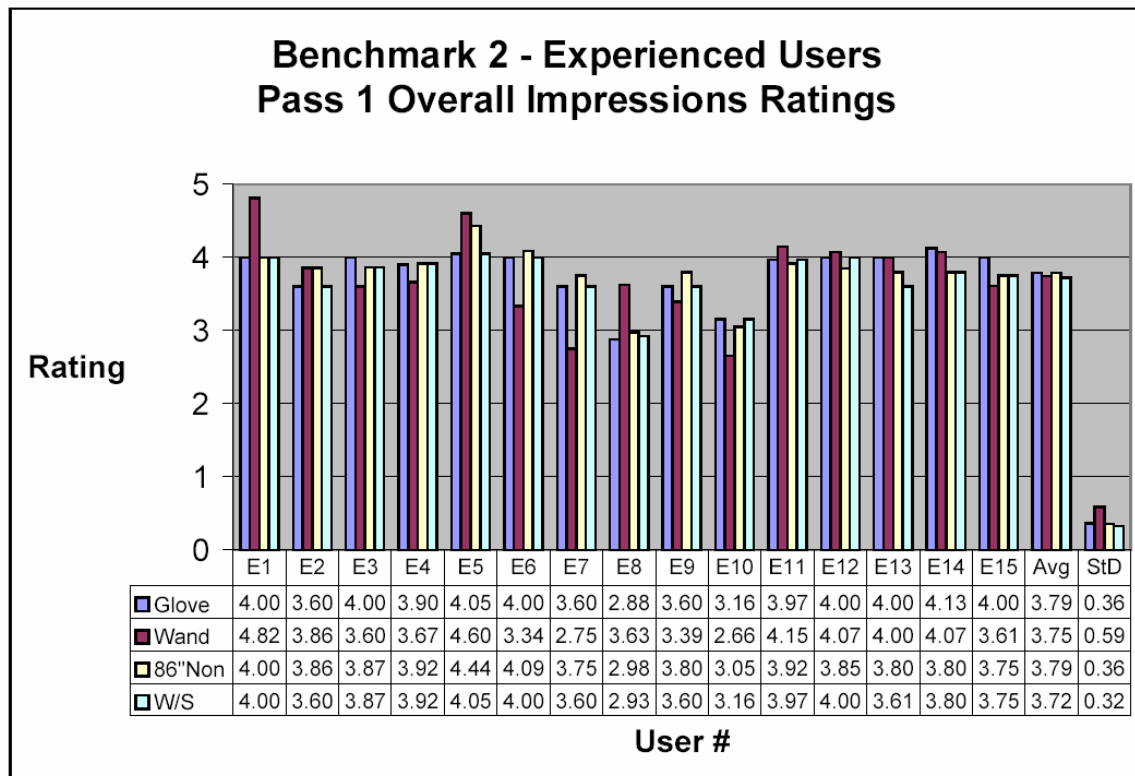
Benchmark 2, Experienced Users, Pass 1, Manipulation Ratings												
B2Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	3.64	0.52	2.60	4.20	0.0016	No	14.27%				
Wand	15	3.49	0.62	2.00	4.40	0.0251	No	17.70%				
86\"Non	15	3.52	0.41	3.00	4.20	0.0535	Yes	11.72%				
W/S	15	3.52	0.53	2.60	4.20	0.0249	No	15.00%				
B2Ep1	Homogeneity of Variance				Roy's Greatest Root: F(3,12)				0.09	Pr > F	0.4817	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.3900	0.5351	Yes	29	1.22	0.2889	Yes	Neither	Neither	
Glove vs 86\"Non			1.2300	0.2769	Yes	29	2.12	0.1671	Yes	Neither	Neither	
Glove vs W/S			0.0100	0.9264	Yes	29	1.47	0.2462	Yes	Neither	Neither	
Wand vs 86\"Non			1.7000	0.2026	Yes	29	0.03	0.8576	Yes	Neither	Neither	
Wand vs W/S			0.3800	0.5418	Yes	29	0.03	0.8734	Yes	Neither	Neither	
86\"Non vs W/S			2.9900	0.0950	Yes	29	0.00	1.0000	Yes	Neither	Neither	



**Figure E- 5: B2Ep1Gen Experienced User General Impressions Ratings**

**Table E- 5: B2Ep1Gen Experienced User General Impressions Ratings Statistics**

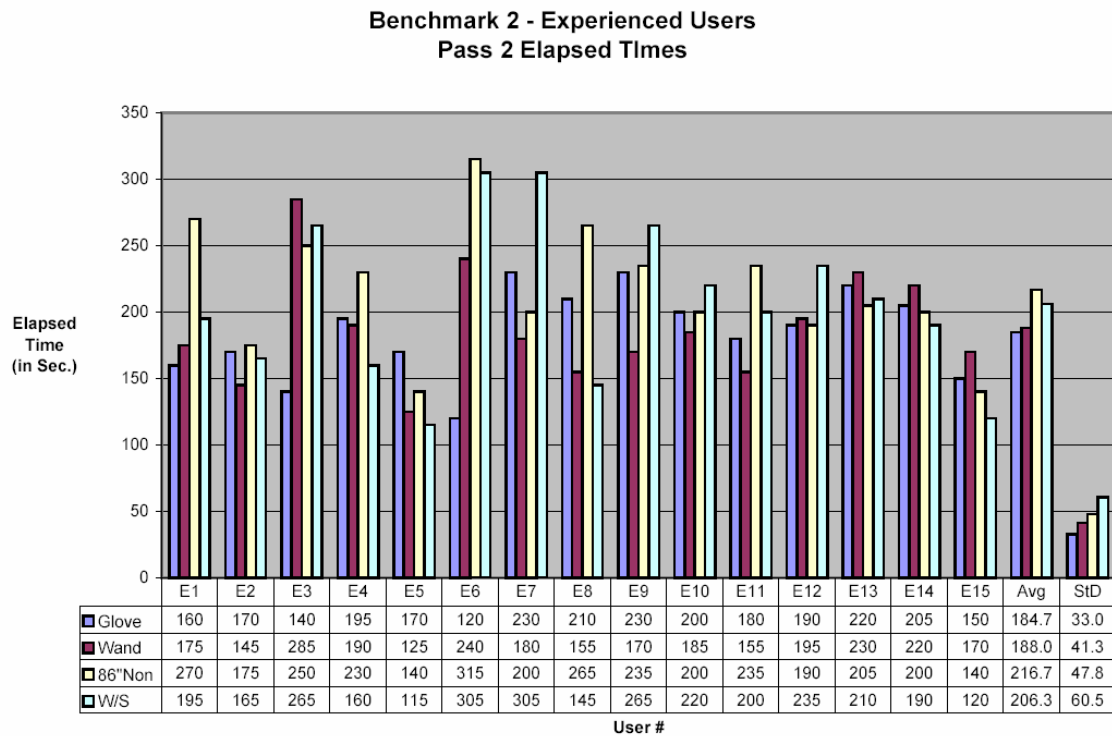
Benchmark 2, Experienced Users, Pass 1, General Impressions Ratings												
B2Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	3.81	0.36	2.80	4.00	< 0.0001	No	9.40%				
Wand	15	3.77	0.60	2.80	5.00	0.0482	No	16.01%				
86\"Non	15	3.85	0.40	2.80	4.40	0.0004	No	10.28%				
W/S	15	3.83	0.36	2.80	4.00	< 0.0001	No	9.45%				
B2Ep1	Homogeneity of Variance					Roy's Greatest Root: F(3,12)			0.28	Pr > F	0.8388	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			2.4900	0.1260	Yes	29	0.06	0.8103	Yes	Neither	Neither	
Glove vs 86\"Non			0.0700	0.7963	Yes	29	0.46	0.5103	Yes	Neither	Neither	
Glove vs W/S			0.0000	0.9829	Yes	29	0.32	0.5816	Yes	Neither	Neither	
Wand vs 86\"Non			1.8300	0.1870	Yes	29	0.30	0.5919	Yes	Neither	Neither	
Wand vs W/S			2.4000	0.1322	Yes	29	0.11	0.7443	Yes	Neither	Neither	
86\"Non vs W/S			0.0600	0.8147	Yes	29	0.24	0.6337	Yes	Neither	Neither	



**Figure E- 6: B2Ep1Ovr Experienced User Pass 1 Overall Impressions Ratings**

**Table E- 6: B2Ep1Ovr Experienced User Pass 1 Overall Impressions Ratings Statistics**

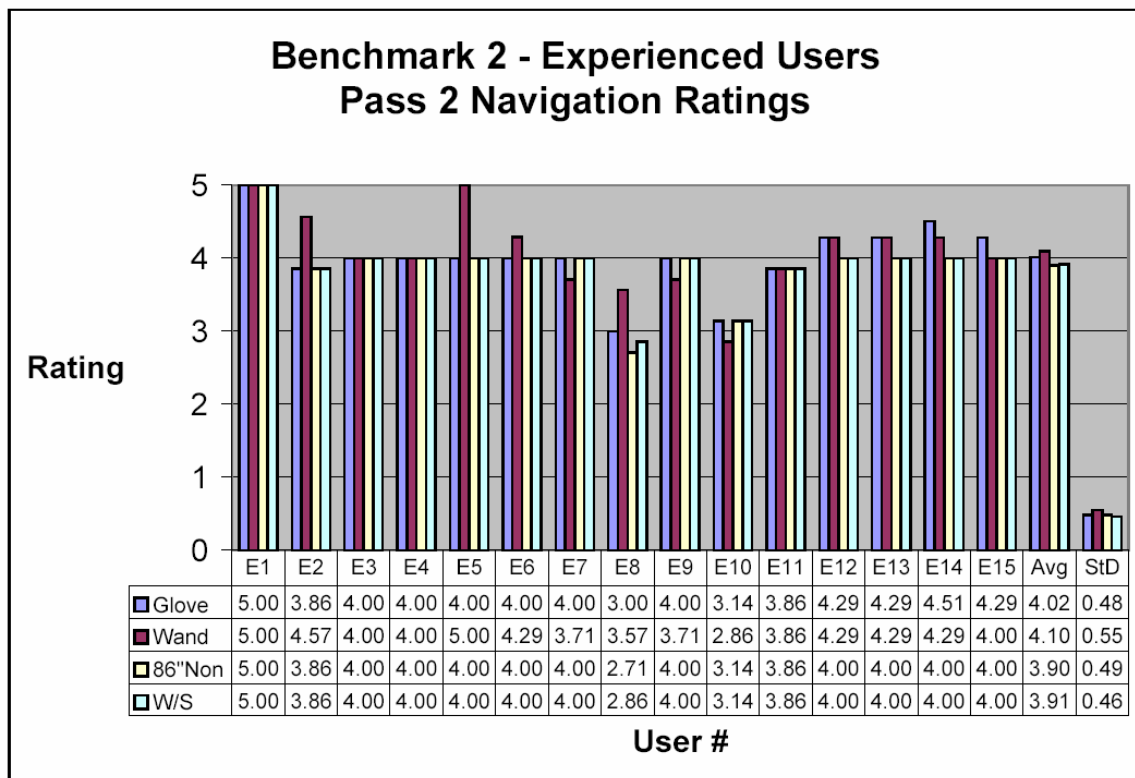
Benchmark 2, Experienced Users, Pass 1, Overall Impressions Ratings											
B2Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	3.79	0.36	2.88	4.13	0.0015	No	21.34%			
Wand	15	3.78	0.59	2.66	4.82	> 0.1000	Yes	52.51%			
86\"Non	15	3.79	0.36	2.98	4.44	0.0002	No	39.55%			
W/S	15	3.72	0.32	2.93	4.05	0.0543	Yes	26.33%			
B2Ep1	Homogeneity of Variance					Roy's Greatest Root:		F(3,12)	2.00	Pr > F	0.1679
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F value	Pr > F							
Glove vs Wand			2.5000	0.1253	Yes	29	0.12	0.7320	Yes	Neither	Neither
Glove vs 86\"Non			0.0000	0.9915	Yes	29	0.00	0.9899	Yes	Neither	Neither
Glove vs W/S			0.1100	0.7401	Yes	29	3.63	0.0774	No	Means	Glove 1.85%
Wand vs 86\"Non			2.4900	0.1258	Yes	29	0.12	0.7325	Yes	Neither	Neither
Wand vs W/S	4.5662	0.0326			No	29	0.04	0.8530	Yes	Var	Wand 1.50%
86\"Non vs W/S			0.1000	0.7537	Yes	29	3.27	0.0919	No	Means	86\"Non 1.83%



**Figure E- 7: B2Ep2Tim Experienced User Elapsed Times**

**Table E- 7: B2Ep2Tim Experienced User Elapsed Times Statistics**

Benchmark 2, Experienced Users, Pass 2, Elapsed Times											
B2Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	184.7	32.98	120.0	230.0	> 0.1000	Yes	17.86%			
Wand	15	188.0	41.27	125.0	285.0	> 0.1000	Yes	21.95%			
86\"Non	15	216.7	47.80	140.0	315.0	> 0.1000	Yes	22.06%			
W/S	15	206.3	60.49	115.0	305.0	> 0.1000	Yes	29.32%			
B2Ep2		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	1.98	Pr > F	0.1709
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand		0.6742	0.4116			Yes	29	0.05	0.8302	Yes	Neither
Glove vs 86\"Non		1.8120	0.1772			Yes	29	3.82	0.0710	No	Means
Glove vs W/S		4.6970	0.0302			No	29	1.55	0.2330	Yes	Var
Wand vs 86\"Non		0.2908	0.5897			Yes	29	5.47	0.0348	No	Means
Wand vs W/S		1.9304	0.1647			Yes	29	2.03	0.1762	Yes	Neither
86\"Non vs W/S		0.7428	0.3888			Yes	29	0.55	0.4690	Yes	Neither

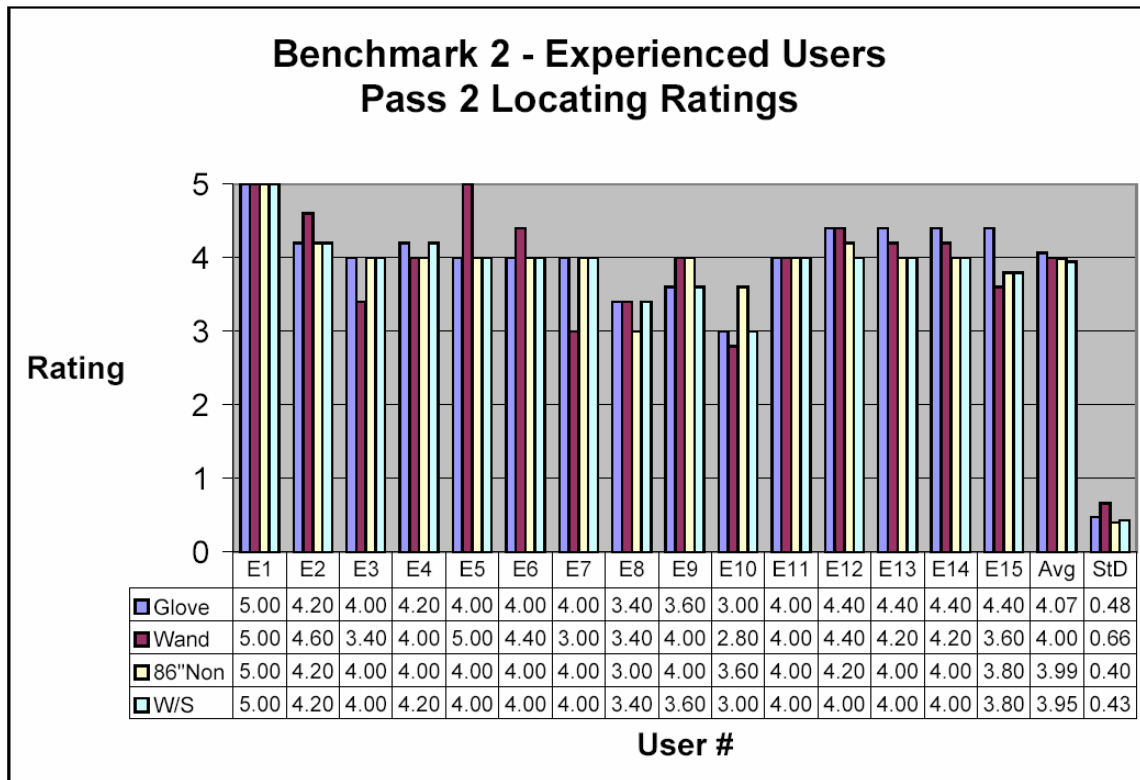


**Figure E- 8: B2Ep2Nav Experienced User Navigation Ratings**

**Table E- 8: B2Ep2Nav Experienced User Navigation Ratings Statistics**

Benchmark 2, Experienced Users, Pass 2, Navigation Ratings											
B2Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.02	0.48	3.00	5.00	0.0197	No	12.07%			
Wand	15	4.10	0.55	2.86	5.00	> 0.1000	Yes	13.36%			
86\"Non	15	3.91	0.49	2.71	5.00	< 0.0001	No	12.45%			
W/S	15	3.92	0.46	2.86	5.00	< 0.0001	No	11.76%			
Homogeneity of Variance					Roy's Greatest Root: F(3,12)			2.27	Pr > F		
B2Ep2	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.1700	0.6865	Yes	29	0.61	0.4471	Yes	Neither	Neither
Glove vs 86\"Non			0.0000	0.9935	Yes	29	6.34	0.0246	No	Means	Glove 2.84%
Glove vs W/S			0.0200	0.8838	Yes	29	5.71	0.0318	No	Means	Glove 2.58%
Wand vs 86\"Non			0.1300	0.7234	Yes	29	3.38	0.0875	No	Means	Wand 4.89%
Wand vs W/S			0.2800	0.6031	Yes	29	3.29	0.0911	No	Means	Wand 4.62%
86\"Non vs W/S			0.0200	0.8903	Yes	29	1.00	0.3343	Yes	Neither	Neither

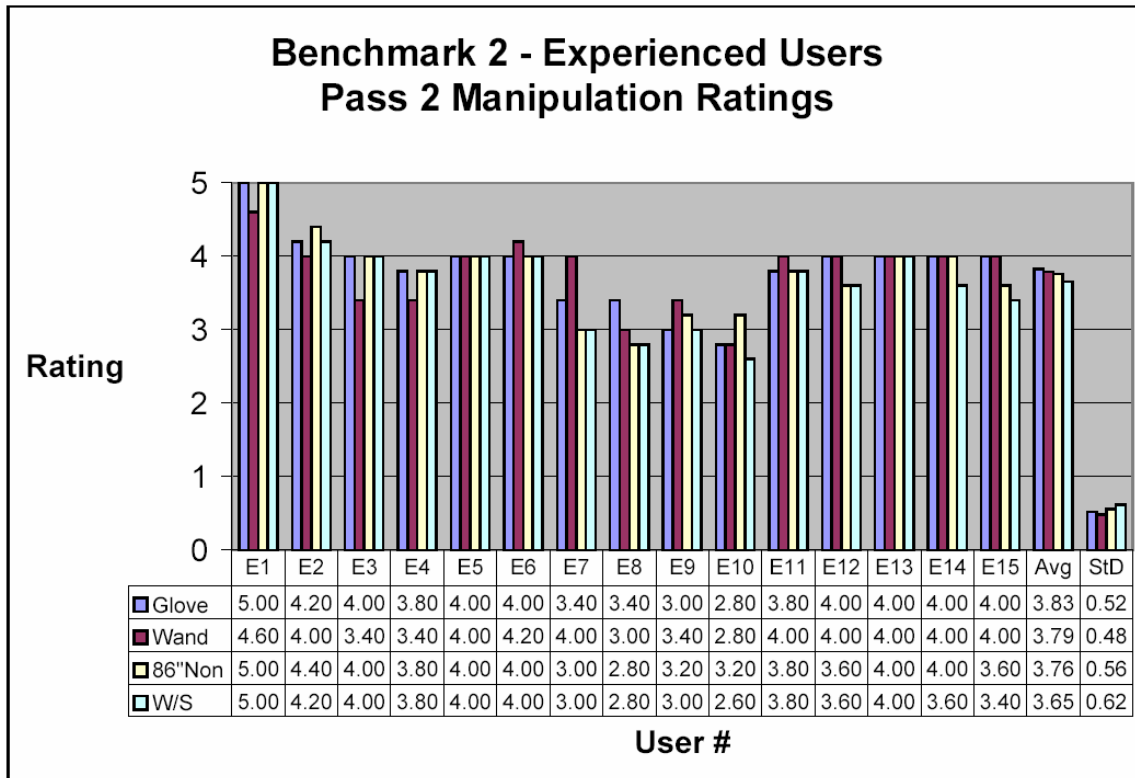




**Figure E- 9: B2Ep2Loc Experienced User Locating Ratings**

**Table E- 9: B2Ep2Loc Experienced User Locating Ratings Statistics**

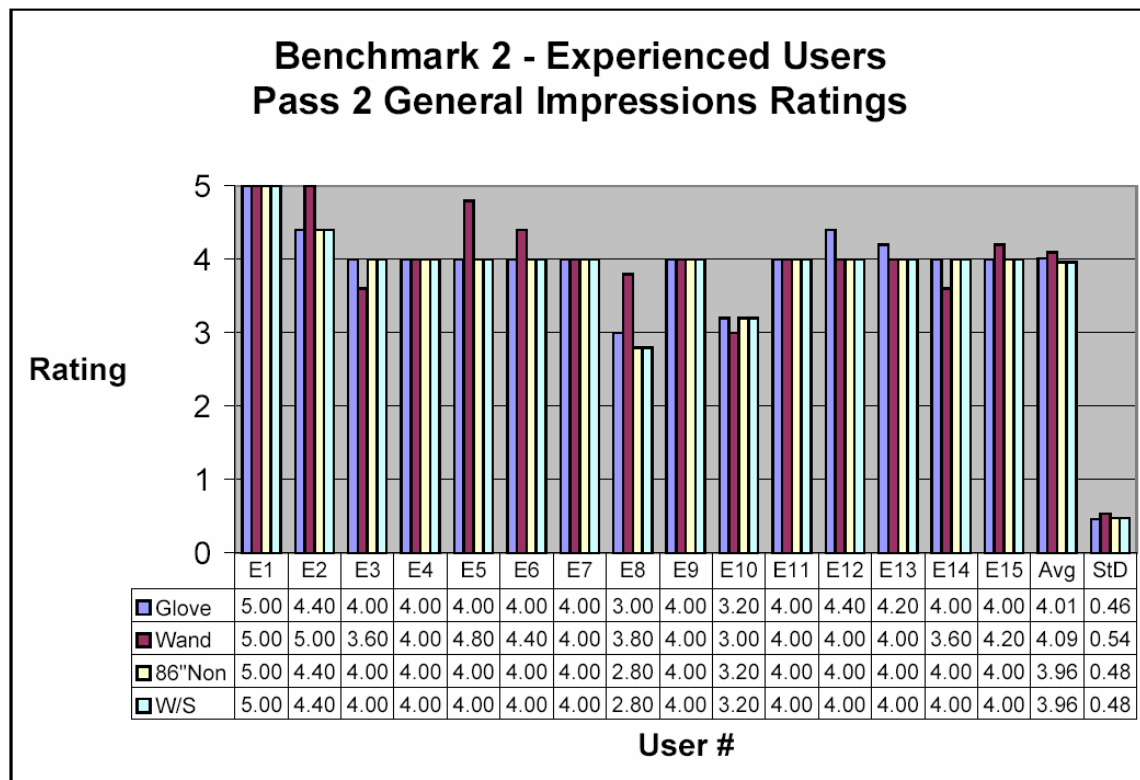
Benchmark 2, Experienced Users, Pass 2, Locating Ratings													
B2Ep2	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	CV	
Glove	15	4.07		0.47		3.00		5.00		0.0163	No	11.60%	
Wand	15	4.00		0.66		2.80		5.00		> 0.1000	Yes	16.48%	
86"Non	15	3.99		0.40		3.00		5.00		0.0003	No	10.12%	
W/S	15	3.95		0.43		3.00		5.00		0.0021	No	10.91%	
B2Ep2		Homogeneity of Variance					Roy's Greatest Root:		F(3,12)		1.52	Pr > F	0.2600
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand				1.6600	0.2084	Yes	29	0.26	0.6173	Yes	Neither	Glove	1.68%
Glove vs 86"Non				0.2200	0.6405	Yes	29	1.00	0.3343	Yes	Neither		Neither
Glove vs W/S				0.0900	0.7642	Yes	29	4.85	0.0450	No	Means	Glove	3.04%
Wand vs 86"Non				0.2200	0.6405	Yes	29	0.01	0.9207	Yes	Neither		Neither
Wand vs W/S				2.3700	0.1348	Yes	29	0.19	0.6702	Yes	Neither		Neither
86"Non vs W/S				0.0300	0.8677	Yes	29	0.46	0.5103	Yes	Neither		Neither



**Figure E- 10: B2Ep2Mov Experienced User Manipulation Ratings**

**Table E- 10: B2Ep2Mov Experienced User Manipulation Ratings Statistics**

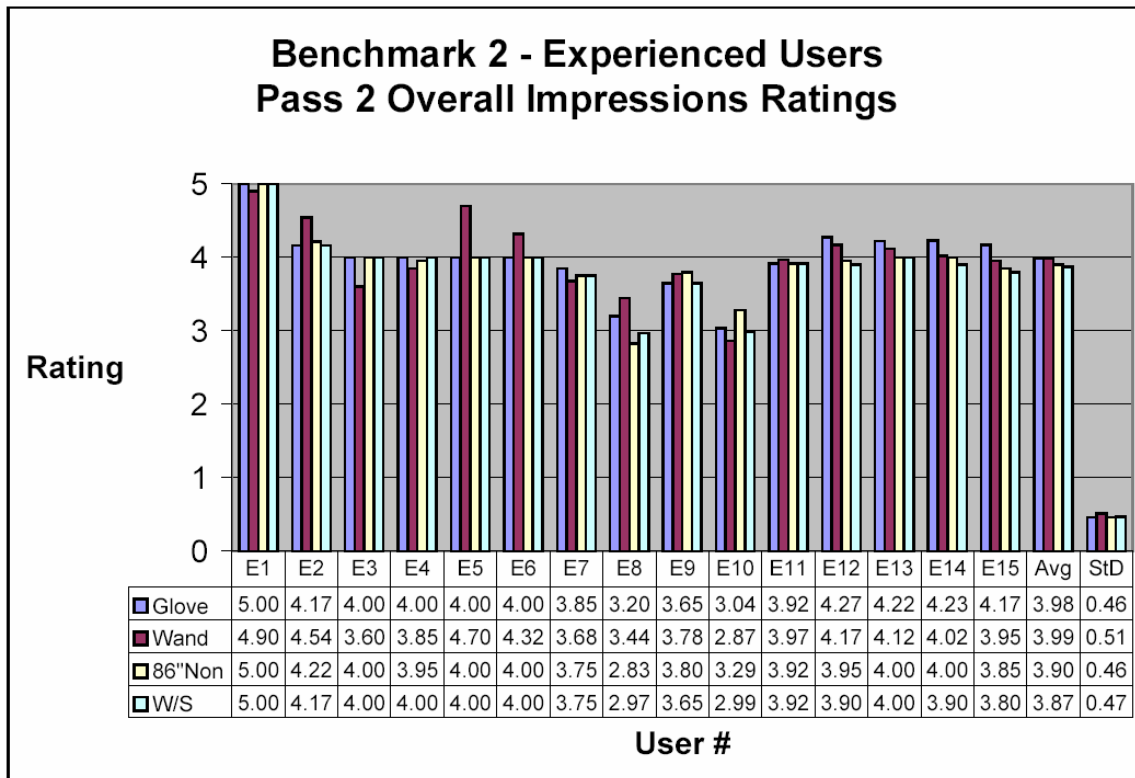
Benchmark 2, Experienced Users, Pass 2, Manipulation Ratings										
B2Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	15	3.83	0.52	2.80	5.00	0.0233	No	13.67%		
Wand	15	3.79	0.48	2.80	4.60	< 0.0001	No	12.70%		
86\"Non	15	3.76	0.56	2.80	5.00	> 0.1000	Yes	14.94%		
W/S	15	3.65	0.62	2.60	5.00	> 0.1000	Yes	17.00%		
B2Ep2		Homogeneity of Variance				Roy's Greatest Root: F(3,12)			Pr > F	0.0539
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					
Glove vs Wand				0.0900	0.7705	Yes	29	0.23	0.6384	Yes
Glove vs 86\"Non				0.3900	0.5351	Yes	29	0.92	0.3535	Yes
Glove vs W/S				1.2300	0.2769	Yes	29	7.99	0.0134	No
Wand vs 86\"Non				0.0100	0.9264	Yes	29	0.06	0.8093	Yes
Wand vs W/S				1.7000	0.2026	Yes	29	1.52	0.2377	Yes
86\"Non vs W/S		0.1359	0.7124			Yes	29	5.09	0.0406	No
										Significant?
										Statistically Better
										Means
										86\"Non
										4.76%
										2.93%



**Figure E- 11: B2Ep2Gen Experienced User General Impressions Ratings**

**Table E- 11: B2Ep2Gen Experienced User General Impressions Ratings Statistics**

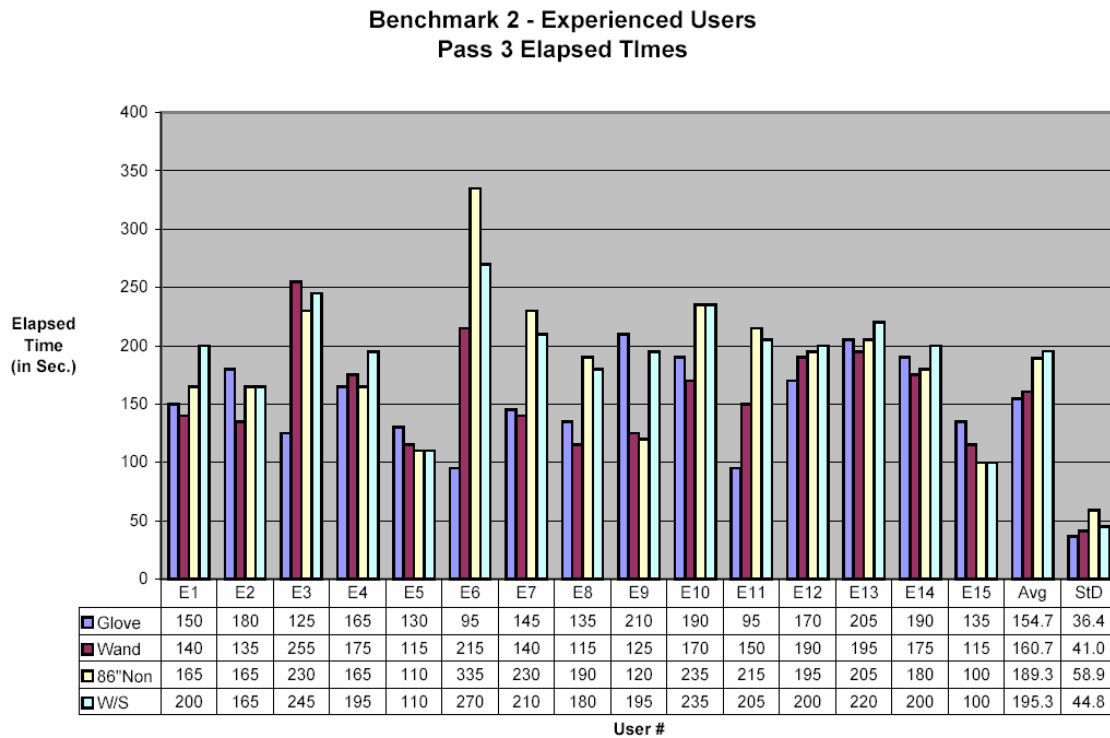
Benchmark 2, Experienced Users, Pass 2, General Impressions Ratings												
B2Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.01	0.46	3.00	5.00	< 0.0001	No	11.53%				
Wand	15	4.09	0.54	3.00	5.00	0.0249	No	13.17%				
86"Non	15	3.96	0.48	2.80	5.00	< 0.0001	No	12.10%				
W/S	15	3.96	0.48	2.80	5.00	< 0.0001	No	12.10%				
B2Ep2		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 1.19 Pr > F 0.353						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.2700	0.6063	Yes	29	0.58	0.4577	Yes	Neither	Neither
Glove vs 86"Non				0.0600	0.8088	Yes	29	3.03	0.1038	Yes	Neither	Neither
Glove vs W/S				0.0100	0.9218	Yes	29	1.67	0.2170	Yes	Neither	Neither
Wand vs 86"Non				0.0600	0.8184	Yes	29	1.63	0.2228	Yes	Neither	Neither
Wand vs W/S				0.1400	0.7073	Yes	29	1.69	0.2150	Yes	Neither	Neither
86"Non vs W/S				0.0200	0.8943	Yes	29	0.00	1.0000	Yes	Neither	Neither



**Figure E- 12: B2Ep2Ovr Experienced User Overall Impressions Ratings**

**Table E- 12: B2Ep2Ovr Experienced User Overall Impressions Ratings Statistics**

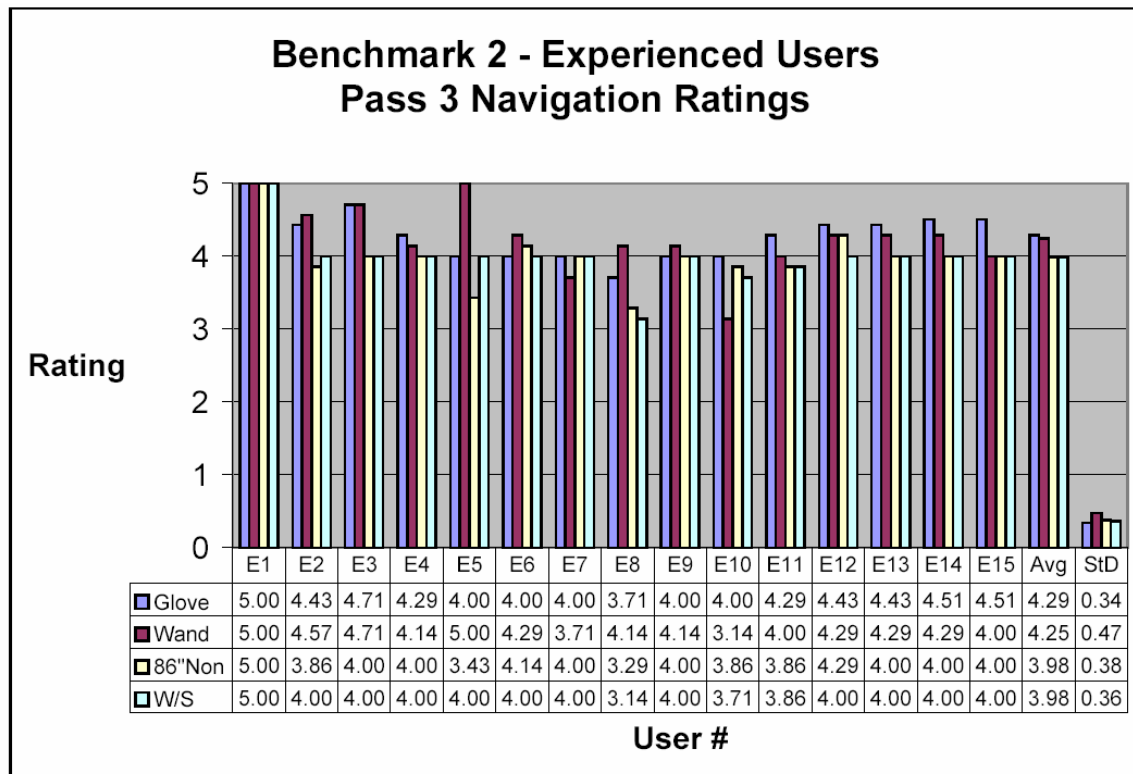
Benchmark 2, Experienced Users, Pass 2, Overall Impressions Ratings												
B2Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	3.98	0.46	3.04	5.00	> 0.1000	Yes	22.53%				
Wand	15	3.99	0.51	2.87	4.90	> 0.1000	Yes	24.73%				
86\"Non	15	3.90	0.46	2.83	5.00	0.0020	No	29.82%				
W/S	15	3.87	0.47	2.97	5.00	0.0082	No	27.39%				
B2Ep2	Homogeneity of Variance					Roy's Greatest Root:		F(3,12)	3.65	Pr > F	0.0443	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand	0.1637	0.6858			Yes	29	0.03	0.8675	Yes	Neither	Neither	
Glove vs 86\"Non			0.0000	0.9923	Yes	29	2.79	0.1182	Yes	Neither	Neither	
Glove vs W/S			0.0100	0.9354	Yes	29	3.58	0.0118	No	Means	Glove	2.87%
Wand vs 86\"Non			0.1200	0.7286	Yes	29	1.23	0.2861	Yes	Neither	Neither	
Wand vs W/S			0.0700	0.7928	Yes	29	3.00	0.1050	Yes	Neither	Neither	
86\"Non vs W/S			0.0100	0.9331	Yes	29	1.80	0.2009	Yes	Neither	Neither	



**Figure E- 13: B2Ep3Tim Experienced User Elapsed Times**

**Table E- 13: B2Ep3Tim Experienced User Elapsed Times Statistics**

Benchmark 2, Experienced Users, Pass 3, Elapsed Times											
B2Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	154.7	36.42	95.0	210.0	> 0.1000	Yes	23.54%			
Wand	15	160.7	41.01	115.0	255.0	> 0.1000	Yes	25.52%			
86\"Non	15	189.3	58.85	100.0	335.0	> 0.1000	Yes	31.09%			
W/S	15	195.3	44.82	100.0	270.0	0.0313	No	22.95%			
Homogeneity of Variance				Roy's Greatest Root: F(3,12) 7.60				Pr > F	0.0041		
B2Ep3	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.1896	0.6633			Yes	29	0.17	0.6894	Yes	Neither	Neither
Glove vs 86\"Non	2.9988	0.0833			No	29	2.86	0.1128	Yes	Var	Glove 18.28%
Glove vs W/S			0.5800	0.4531	Yes	29	7.34	0.0169	No	Means	Glove 20.79%
Wand vs 86\"Non	1.7273	0.1888			Yes	29	6.41	0.0239	No	Means	Wand 15.11%
Wand vs W/S			0.1000	0.7522	Yes	29	19.60	0.0006	No	Means	Wand 17.72%
86\"Non vs W/S			0.7000	0.4106	Yes	29	0.59	0.4568	Yes	Neither	Neither

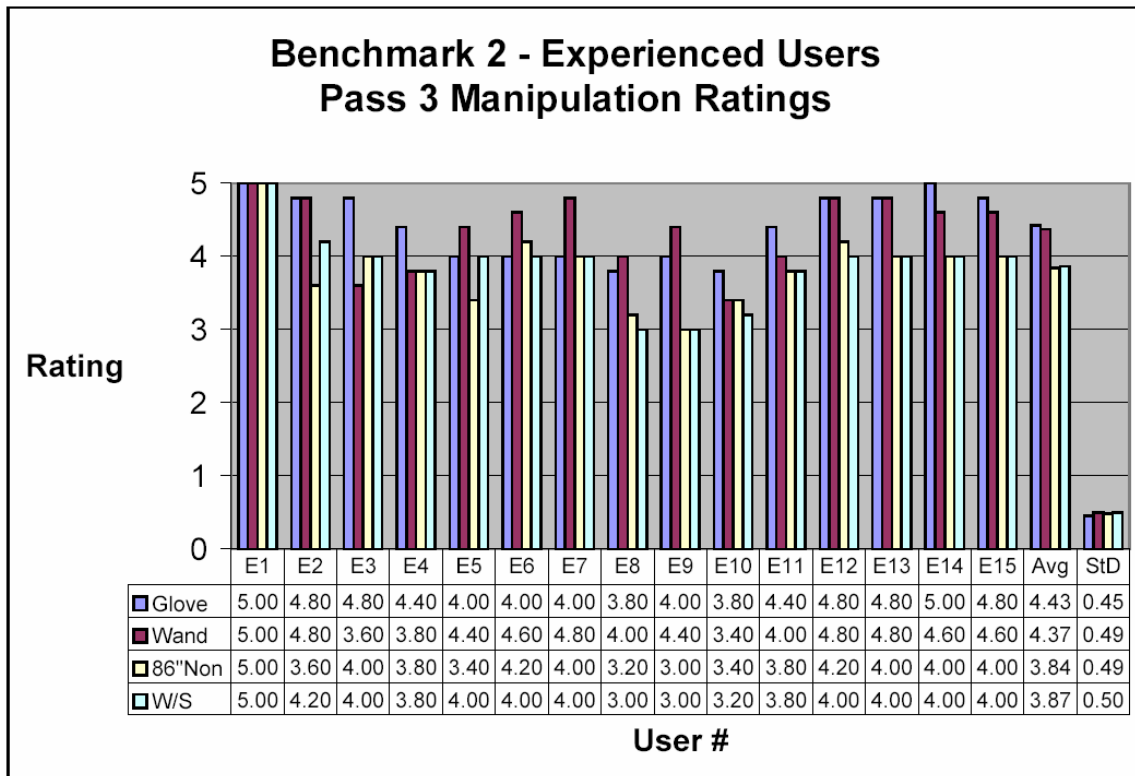


**Figure E- 14: B2Ep3Nav Experienced User Navigation Ratings**

**Table E- 14: B2Ep3Nav Experienced User Navigation Ratings Statistics**

Benchmark 2, Experienced Users, Pass 3, Navigation Ratings											
B2Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.29	0.34	3.71	5.00	0.0977	Yes	7.87%			
Wand	15	4.25	0.47	3.14	5.00	> 0.1000	Yes	11.13%			
86\"Non	15	3.98	0.38	3.29	5.00	0.0024	No	9.46%			
W/S	15	3.98	0.36	3.14	5.00	< 0.0001	No	9.10%			
B2Ep3	Homogeneity of Variance					Roy's Greatest Root: F(3,12) 7.29			Pr > F	0.0048	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	1.5139	0.2186			Yes	29	0.13	0.7227	Yes	Neither	Neither
Glove vs 86\"Non			0.1000	0.0267	No	29	20.29	0.0005	No	Var/Mean	Glove 7.66%
Glove vs W/S			0.0300	0.8544	Yes	29	23.34	0.0003	No	Means	Glove 7.69%
Wand vs 86\"Non			0.4400	0.5132	Yes	29	3.70	0.0749	No	Means	Wand 6.65%
Wand vs W/S			0.5200	0.4769	Yes	29	5.82	0.0302	No	Means	Wand 6.68%
86\"Non vs W/S			0.0100	0.9286	Yes	29	0.00	0.9784	Yes	Neither	Neither



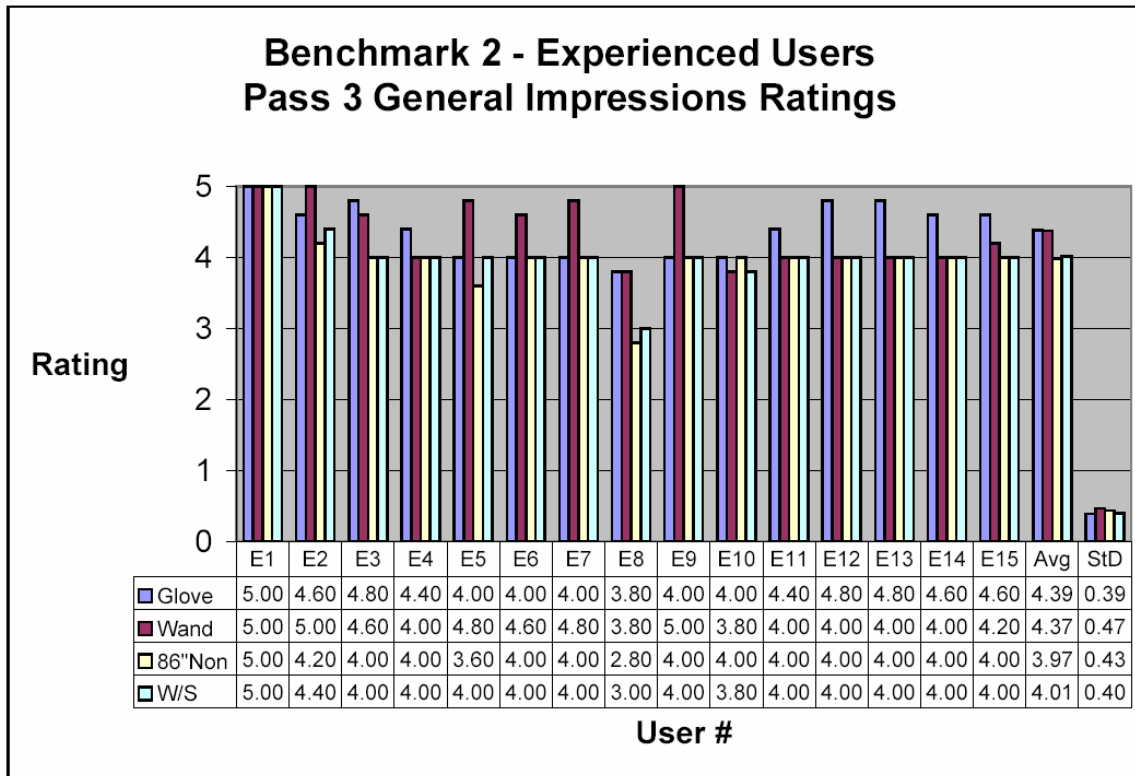


**Figure E- 16: B2Ep3Mov Experienced User Manipulation Ratings**

**Table E- 16: B2Ep3Mov Experienced User Manipulation Ratings Statistics**

Benchmark 2, Experienced Users, Pass 3, Manipulation Ratings												
B2Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.43	0.45	3.80	5.00	0.0067	No	10.23%				
Wand	15	4.37	0.49	3.40	5.00	0.0744	Yes	11.32%				
86"Non	15	3.84	0.49	3.00	5.00	> 0.1000	Yes	12.64%				
W/S	15	3.88	0.50	3.00	5.00	0.0039	No	12.89%				
B2Ep3		Homogeneity of Variance				Roy's Greatest Root: F(3,12)			13.29	Pr > F 0.0004		
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.2600	0.6129	Yes	29	0.16	0.6914	Yes	Neither	Neither
Glove vs 86"Non				0.0800	0.7767	Yes	29	32.89	<0.0001	Yes	Neither	Neither
Glove vs W/S				0.1500	0.6991	Yes	29	34.02	<0.0001	Yes	Neither	Glove 14.22%
Wand vs 86"Non		0.0053	0.9419			Yes	29	17.10	0.0010	No	Means	Wand 13.88%
Wand vs W/S				0.0000	0.9722	Yes	29	18.79	0.0007	No	Means	Wand 12.82%
86"Non vs W/S				0.0100	0.9252	Yes	29	0.17	0.6848	Yes	Neither	Neither

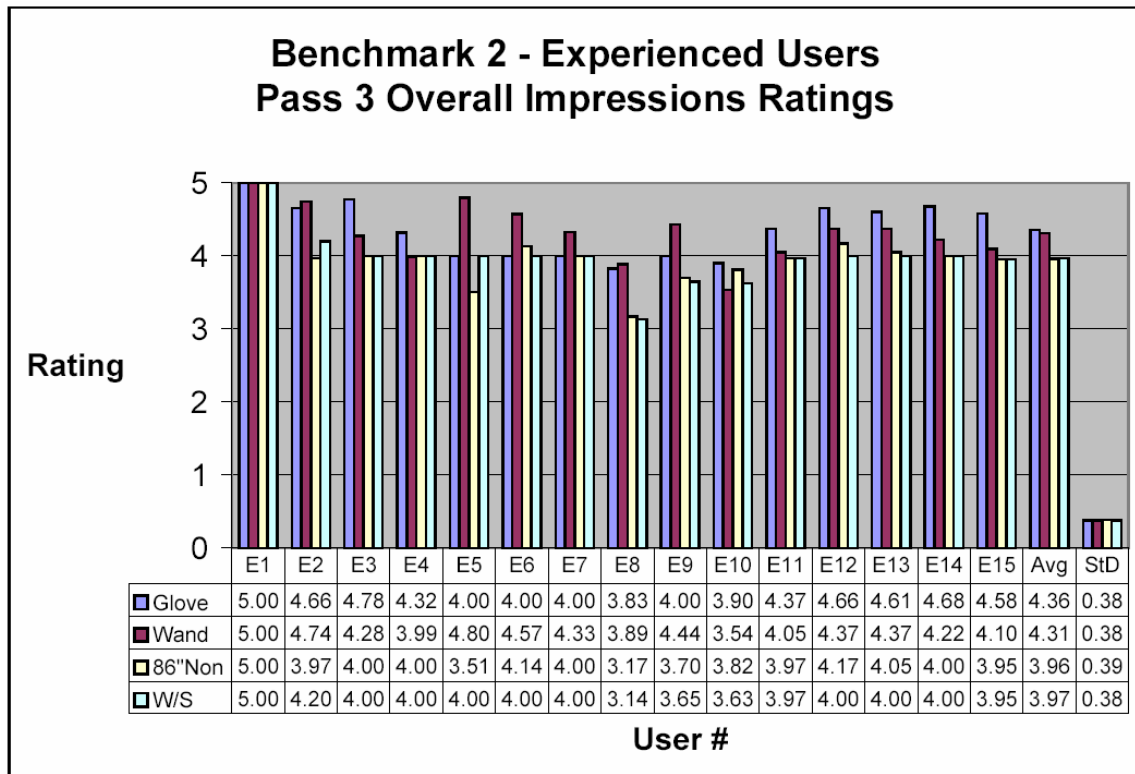




**Figure E- 17: B2Ep3Gen Experienced User General Impressions Ratings**

**Table E- 17: B2Ep3Gen Experienced User General Impressions Ratings Statistics**

Benchmark 2, Experienced Users, Pass 3, Overall Impressions Ratings											
B2Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.36	0.38	3.83	5.00	0.0324	No	14.43%			
Wand	15	4.31	0.38	3.54	5.00	> 0.1000	Yes	21.58%			
86\"Non	15	3.96	0.39	3.17	5.00	0.0324	No	24.75%			
W/S	15	3.97	0.38	3.14	5.00	0.0087	No	18.83%			
B2Ep3	Homogeneity of Variance				Roy's Greatest Root: F(3,12)		14.21	Pr > F	0.0003		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.0000	0.9938	Yes	29	0.19	0.6703	Yes	Neither	Neither
Glove vs 86\"Non			0.0100	0.9404	Yes	29	27.47	0.0001	No	Means	Glove 9.99%
Glove vs W/S			0.0000	0.9774	Yes	29	28.09	0.0001	No	Means	Glove 9.83%
Wand vs 86\"Non			0.0100	0.9412	Yes	29	11.85	0.0040	No	Means	Wand 8.83%
Wand vs W/S			0.0000	0.9757	Yes	29	20.26	0.0005	No	Means	Wand 8.67%
86\"Non vs W/S			0.0000	0.9727	Yes	29	0.02	0.8913	Yes	Neither	Neither

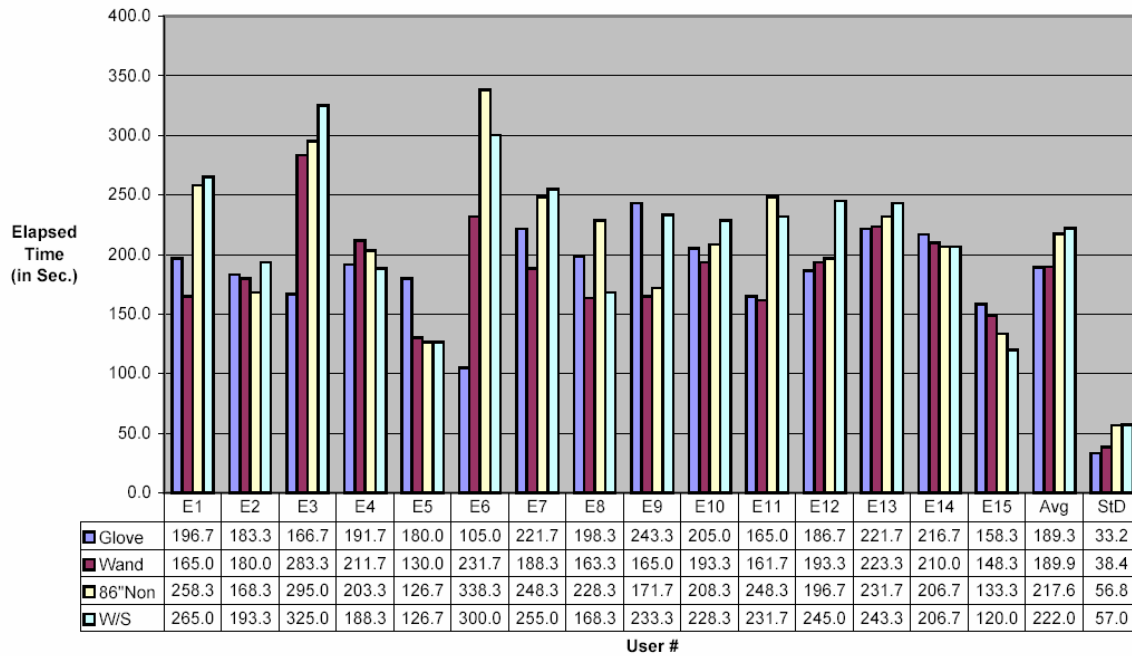


**Figure E- 18: B2Ep3Ovr Experienced User Overall Impressions Ratings**

**Table E- 18: B2Ep3Ovr Experienced User Overall Impressions Ratings Statistics**

Benchmark 2, Experienced Users, Pass 3, Overall Impressions Ratings												
B2Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.36	0.38	3.83	5.00	0.0324	No	14.43%				
Wand	15	4.31	0.38	3.54	5.00	> 0.1000	Yes	21.58%				
86\"Non	15	3.96	0.39	3.17	5.00	0.0324	No	24.75%				
W/S	15	3.97	0.38	3.14	5.00	0.0087	No	18.83%				
B2Ep3	Homogeneity of Variance					Roy's Greatest Root:			F(3,12)	14.21	Pr > F	0.0003
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.0000	0.9938	Yes	29	0.19	0.6703	Yes	Neither	Neither	
Glove vs 86\"Non			0.0100	0.9404	Yes	29	27.47	0.0001	No	Means	Glove	9.99%
Glove vs W/S			0.0000	0.9774	Yes	29	28.09	0.0001	No	Means	Glove	9.83%
Wand vs 86\"Non			0.0100	0.9412	Yes	29	11.85	0.0040	No	Means	Wand	8.83%
Wand vs W/S			0.0000	0.9757	Yes	29	20.26	0.0005	No	Means	Wand	8.67%
86\"Non vs W/S			0.0000	0.9727	Yes	29	0.02	0.8913	Yes	Neither	Neither	

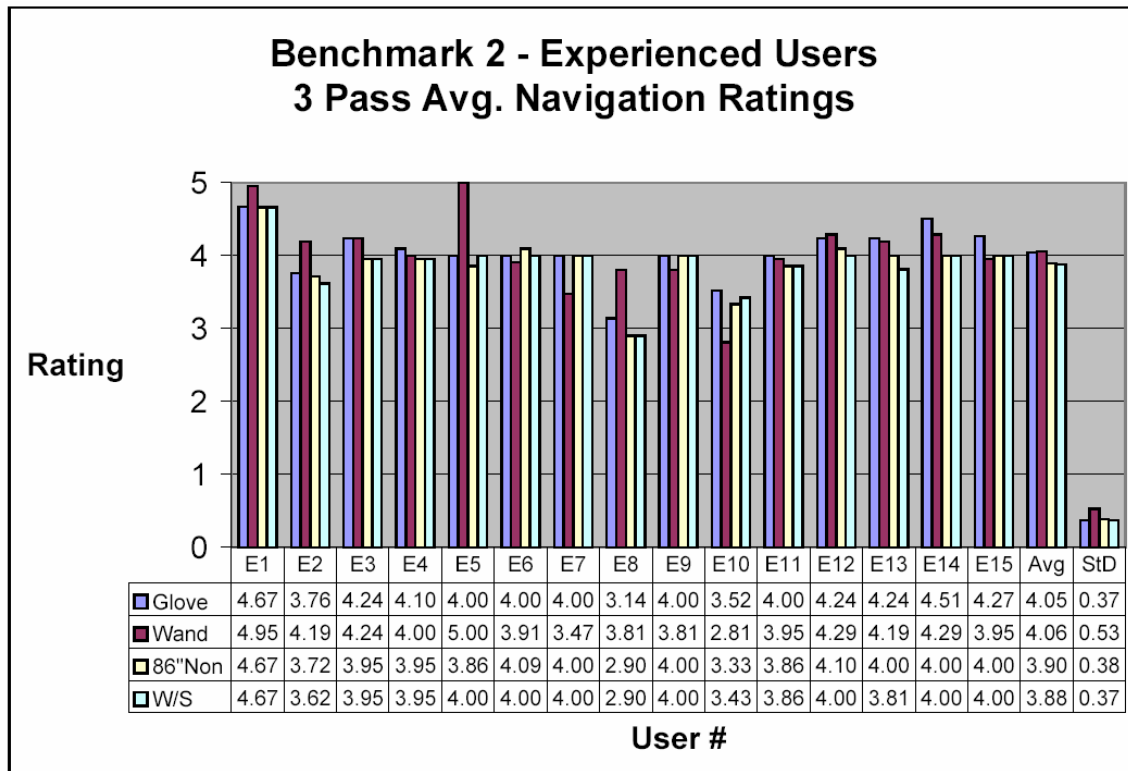
**Benchmark 2 - Experienced Users  
3 Pass Avg. Elapsed Times**



**Figure E- 19: B2E3pAvgTim Experienced Users Elapsed Times**

**Table E- 19: B2E3pAvgTim Experienced Users Elapsed Times Statistics**

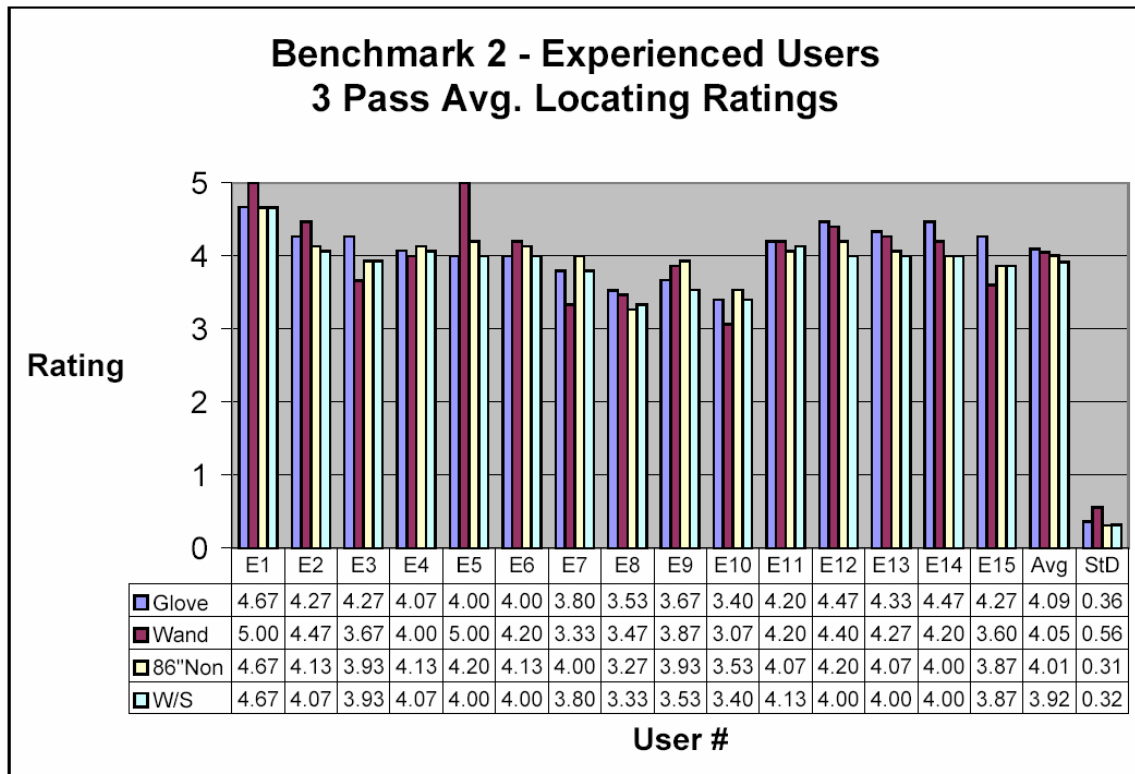
Benchmark 2, Experienced Users, 3 Pass Avg., Elapsed Times											
B2E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	189.3	33.22	105.0	243.3	> 0.1000	Yes	17.55%			
Wand	15	189.9	38.41	130.0	283.3	> 0.1000	Yes	20.23%			
86"Non	15	217.5	56.80	126.7	338.3	> 0.1000	Yes	26.11%			
W/S	15	222.0	56.97	120.0	325.0	> 0.1000	Yes	25.66%			
B2E3pA	Homogeneity of Variance				Roy's Greatest Root: F(3,12)			3.04	Pr > F	0.0707	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.2832	0.5946			Yes	29	0.00	0.9702	Yes	Neither	Neither
Glove vs 86"Non	3.7147	0.0539			Yes	29	2.07	0.1718	Yes	Neither	Neither
Glove vs W/S	3.7544	0.0527			Yes	29	3.33	0.0894	No	Means	Glove 14.73%
Wand vs 86"Non	2.0191	0.1553			Yes	29	6.47	0.0234	No	Means	Wand 12.69%
Wand vs W/S	2.0493	0.1523			Yes	29	10.40	0.0061	No	Means	Wand 14.46%
86"Non vs W/S	0.0001	0.9912			Yes	29	0.03	0.5910	Yes	Neither	Neither



**Figure E- 20: B2E3pAvgNav Experienced Users Navigation Ratings**

**Table E- 20: B2E3pAvgNav Experienced Users Navigation Ratings Statistics**

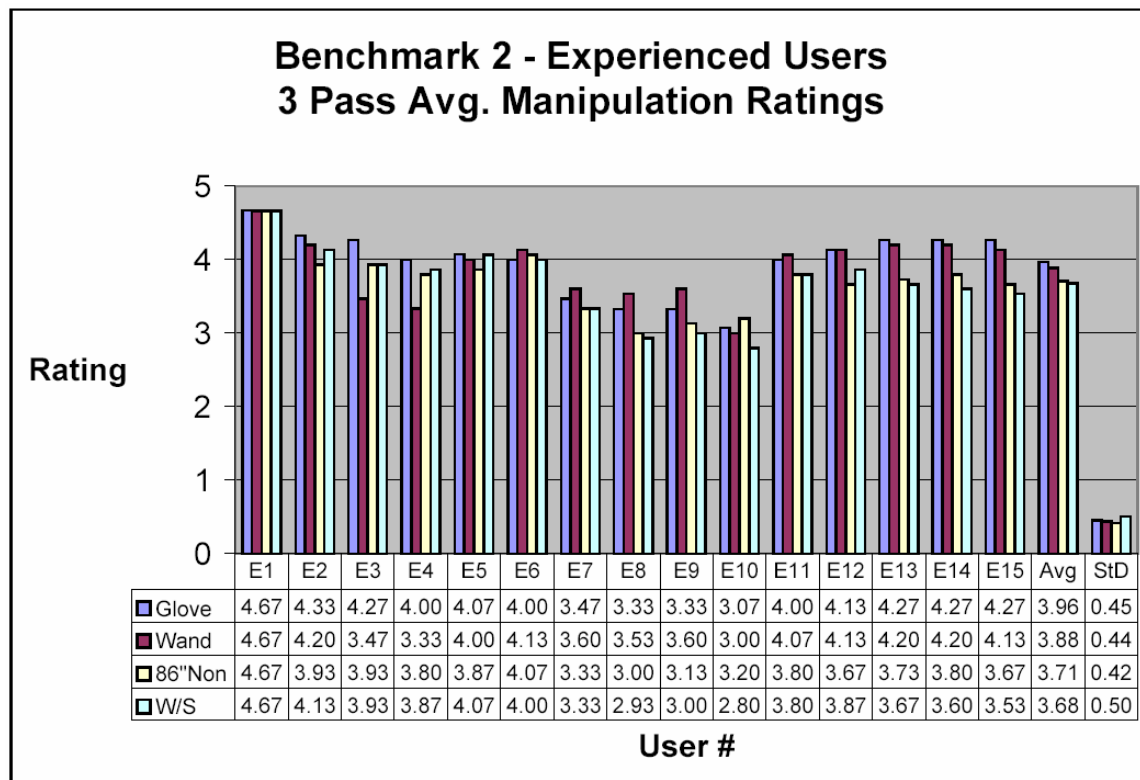
Benchmark 2, Experienced Users, 3 Pass Avg., Navigation Ratings											
B2E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.05	0.37	3.14	4.67	0.0117	No	9.25%			
Wand	15	4.06	0.53	2.81	5.00	> 0.1000	Yes	13.07%			
86"Non	15	3.90	0.39	2.90	4.67	0.0062	No	9.91%			
W/S	15	3.88	0.38	2.90	4.67	0.0005	No	9.67%			
B2E3pA	Homogeneity of Variance					Roy's Greatest Root: F(3,12)		4.76	Pr > F	0.0207	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.9900	0.3273	Yes	29	0.01	0.9219	Yes	Neither	Neither
Glove vs 86"Non			0.0100	0.9309	Yes	29	14.95	0.0017	No	Means	Glove 3.88%
Glove vs W/S			0.0000	0.9944	Yes	29	15.75	0.0014	No	Means	Glove 4.31%
Wand vs 86"Non			0.7800	0.3855	Yes	29	1.90	0.1901	Yes	Neither	Neither
Wand vs W/S			0.8900	0.3539	Yes	29	2.31	0.1505	Yes	Neither	Neither
86"Non vs W/S			0.0100	0.9423	Yes	29	0.60	0.4524	Yes	Neither	Neither



**Figure E- 21: B2E3pAvg Experienced User Locating Ratings**

**Table E- 21: B2E3pAvg Experienced User Locating Ratings Statistics**

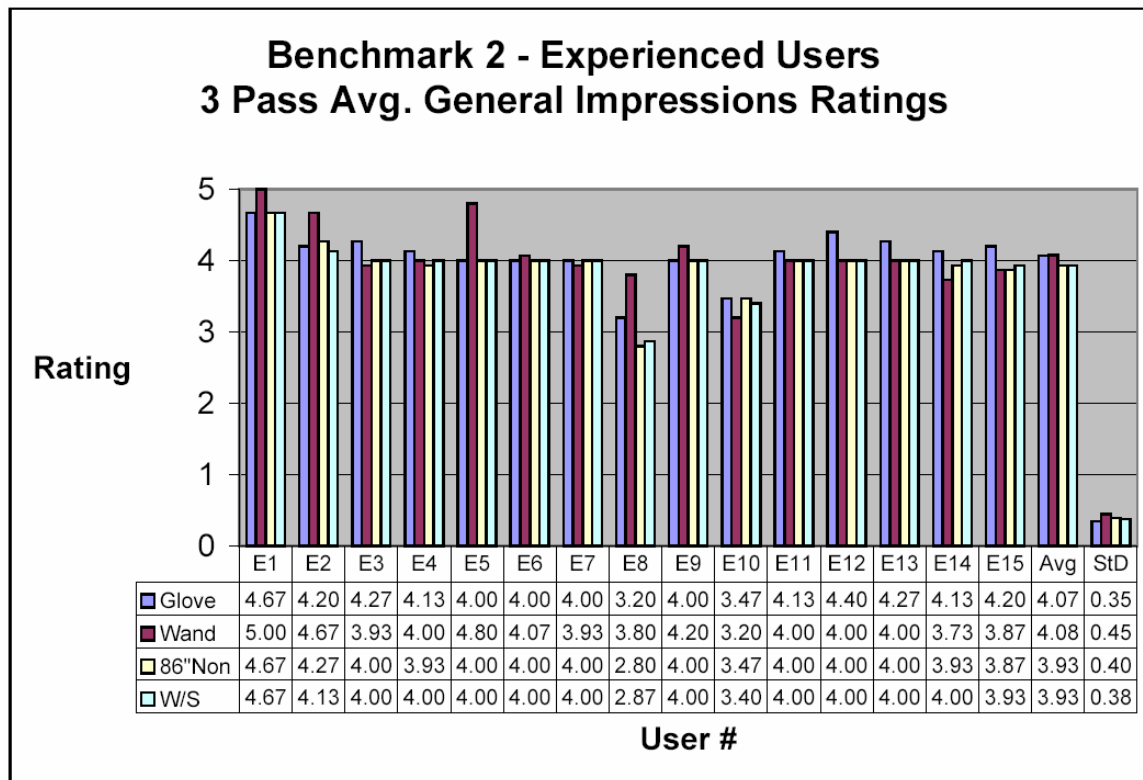
Benchmark 2, Experienced Users, 3 Pass Avg., Locating Ratings											
B2E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.10	0.36	3.40	4.67	> 0.1000	Yes	8.90%			
Wand	15	4.05	0.56	3.07	5.00	> 0.1000	Yes	13.87%			
86\"Non	15	4.01	0.31	3.27	4.67	0.0947	Yes	7.80%			
W/S	15	3.92	0.32	3.33	4.67	0.0929	Yes	8.27%			
B2E3pA		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)		
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand		2.4567	0.1170			Yes	29	0.17	0.6841	Yes	Neither
Glove vs 86\"Non		0.3150	0.5746			Yes	29	1.91	0.1891	Yes	Neither
Glove vs W/S		0.1842	0.6678			Yes	29	13.54	0.0025	No	Means
Wand vs 86\"Non		4.3946	0.0361			No	29	0.19	0.6657	Yes	Var
Wand vs W/S		3.8936	0.0485			Yes	29	1.83	0.1980	Yes	Neither
86\"Non vs W/S		0.0176	0.8945			Yes	29	7.67	0.0150	No	Means
											86\"Non
											2.27%



**Figure E- 22: B2E3pAvgMov Experienced User Manipulation Ratings**

**Table E- 22: B2E3pAvgMov Experienced User Manipulation Ratings Statistics**

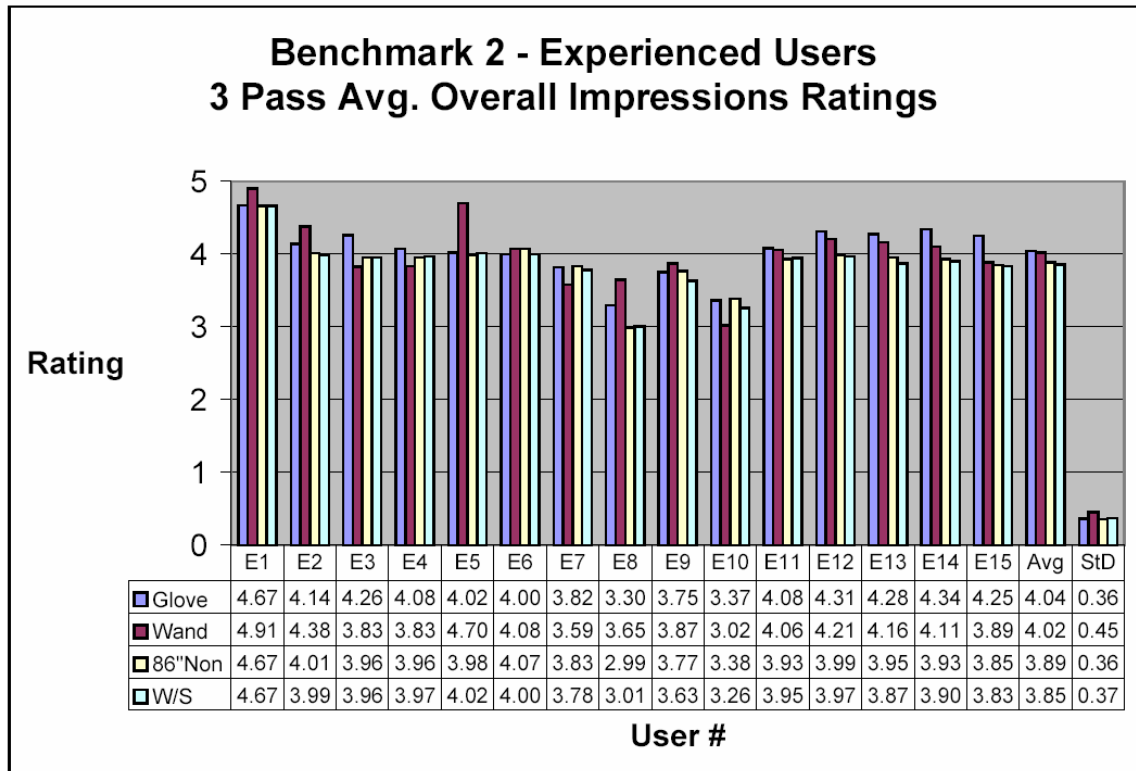
Benchmark 2, Experienced Users, 3 Pass Avg., Manipulation Ratings												
B2E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	3.97	0.45	3.07	4.67	0.0061	No	11.47%				
Wand	15	3.88	0.44	3.00	4.67	0.0924	Yes	11.29%				
86"Non	15	3.71	0.42	3.00	4.67	> 0.1000	Yes	11.28%				
W/S	15	3.68	0.50	2.80	4.67	> 0.1000	Yes	13.66%				
B2E3pA		Homogeneity of Variance				Roy's Greatest Root: F(3,12)			7.12	Pr > F		0.0053
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.0300	0.8725	Yes	29	1.16	0.2993	Yes	Neither	Neither
Glove vs 86"Non				0.1100	0.7411	Yes	29	21.09	0.0004	No	Means	Glove 6.96%
Glove vs W/S				0.1800	0.6744	Yes	29	22.01	0.0003	No	Means	Glove 7.74%
Wand vs 86"Non		0.0313	0.8597			Yes	29	4.27	0.0578	No	Means	Wand 4.77%
Wand vs W/S		0.2561	0.6170			Yes	29	4.65	0.0490	No	Means	Wand 5.54%
86"Non vs W/S		0.4568	0.4991			Yes	29	0.41	0.5323	Yes	Neither	Neither



**Figure E- 23: B2E3pAvgGen Experienced User General Impressions Ratings**

**Table E- 23: B2E3pAvgGen Experienced User General Impressions Ratings Statistics**

Benchmark 2, Experienced Users, 3 Pass Avg., General Impressions Ratings											
B2E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.07	0.35	3.20	4.67	0.0017	No	8.63%			
Wand	15	4.08	0.45	3.20	5.00	0.0181	No	11.00%			
86\"Non	15	3.93	0.40	2.80	4.67	0.0005	No	10.08%			
W/S	15	3.93	0.38	2.87	4.67	< 0.0001	No	9.71%			
B2E3pA	Homogeneity of Variance				Roy's Greatest Root: F(3,12)			5.57	Pr > F	0.0125	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.6000	0.4470	Yes	29	0.01	0.9322	Yes	Neither	Neither
Glove vs 86\"Non			0.0900	0.7703	Yes	29	11.32	0.0046	No	Means	Glove 3.61%
Glove vs W/S			0.0400	0.4645	Yes	29	15.24	0.0016	No	Means	Glove 3.61%
Wand vs 86\"Non			0.1300	0.7249	Yes	29	2.47	0.1202	Yes	Neither	Neither
Wand vs W/S			0.2200	0.6440	Yes	29	2.58	0.1308	Yes	Neither	Neither
86\"Non vs W/S			0.0100	0.9728	Yes	29	0.08	0.7828	Yes	Neither	Neither



**Figure E- 24: B2E3pAvgOvr Experienced User Overall Impressions Ratings**

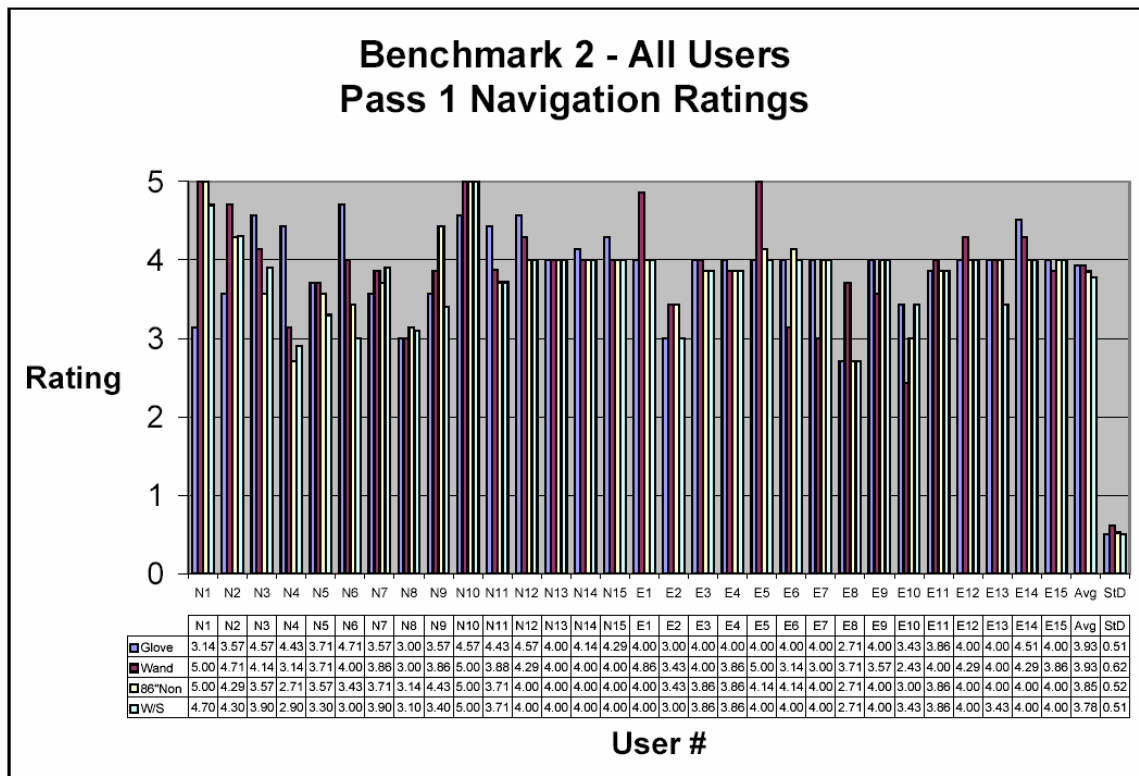
**Table E- 24: B2E3pAvgOvr Experienced User Overall Impressions Ratings Statistics**

Benchmark 2, Experienced Users, 3 Pass Avg., Overall Impressions Ratings													
B2E3pA	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	CV	
Glove	15	4.05		0.36		3.30		4.67		> 0.1000	Yes	16.03%	
Wand	15	4.02		0.45		3.02		4.91		> 0.1000	Yes	31.19%	
86"Non	15	3.89		0.36		2.99		4.67		0.0196	No	27.49%	
W/S	15	3.85		0.37		3.01		4.67		0.0076	No	20.24%	
Homogeneity of Variance						Roy's Greatest Root:		F(3,12)		10.18		Pr > F	0.0013
B2E3pA		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means ?	Significant?	Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand				0.5400	0.4668	Yes	29	0.10	0.0521	No	Means	Glove	0.65%
Glove vs 86"Non				0.0000	0.9569	Yes	29	13.36	0.0048	No	Means	Glove	4.12%
Glove vs W/S				0.0000	0.9662	Yes	29	21.08	0.0000	No	Means	Glove	4.96%
Wand vs 86"Non				0.5100	0.4804	Yes	29	3.04	0.4481	Yes	Neither	Wand	3.45%
Wand vs W/S				0.4100	0.5274	Yes	29	5.38	0.0173	No	Means	Wand	4.28%
86"Non vs W/S		0.0156	0.9005			Yes	29	5.16	0.0080	No	Means	86"Non	0.80%



## **Appendix F**

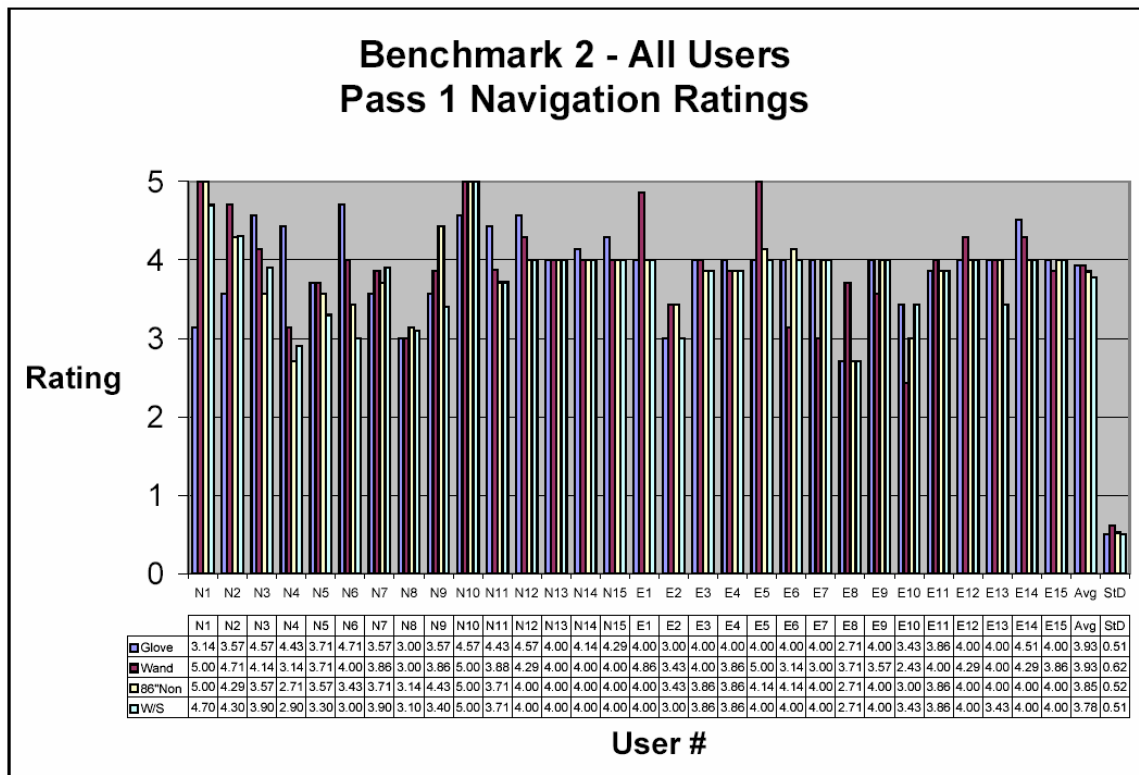
### All Users Benchmark 2 (Find/Repair) Detail



**Figure F- 1: B2Ap1Tim All User Elapsed Times**

**Table F- 1: B2Ap1Tim All User Elapsed Times Statistics**

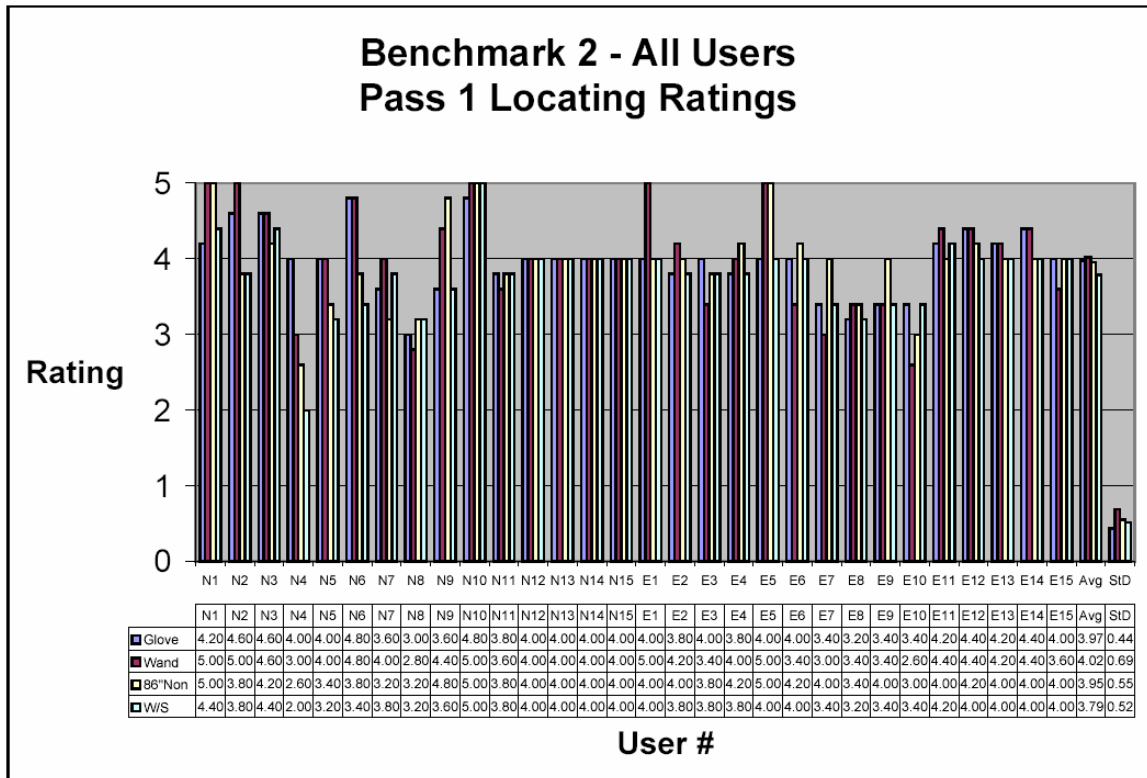
Benchmark 2, All Users, Pass 1, Elapsed Times										
B2Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	30	226.0	58.24	100.0	400.0	> 0.1000	Yes	25.77%		
Wand	30	231.8	51.46	150.0	365.0	> 0.1000	Yes	22.20%		
86*Non	30	258.9	70.99	130.0	405.0	> 0.1000	Yes	27.42%		
W/S	30	274.5	78.70	140.0	465.0	> 0.1000	Yes	28.67%		
B2Ap1	Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	3.25	Pr > F	0.0372
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						Statistically Better
Glove vs Wand	0.4358	0.5092			Yes	29	0.24	0.6251	Yes	Neither
Glove vs 86*Non	1.1102	0.2920			Yes	29	4.30	0.0472	No	Means
Glove vs W/S	2.5456	0.1106			Yes	29	7.23	0.0118	No	Means
Wand vs 86*Non	2.9022	0.8858			Yes	29	6.10	0.0197	No	Means
Wand vs W/S	4.9978	0.0254			No	29	10.22	0.0033	No	Var/Means
86*Non vs W/S	0.3023	0.5825			Yes	29	2.64	0.1149	Yes	Neither



**Figure F- 2: B2Ap1Nav All Users Navigation Ratings**

**Table F- 2: B2Ap1Nav All Users Navigation Ratings Statistics**

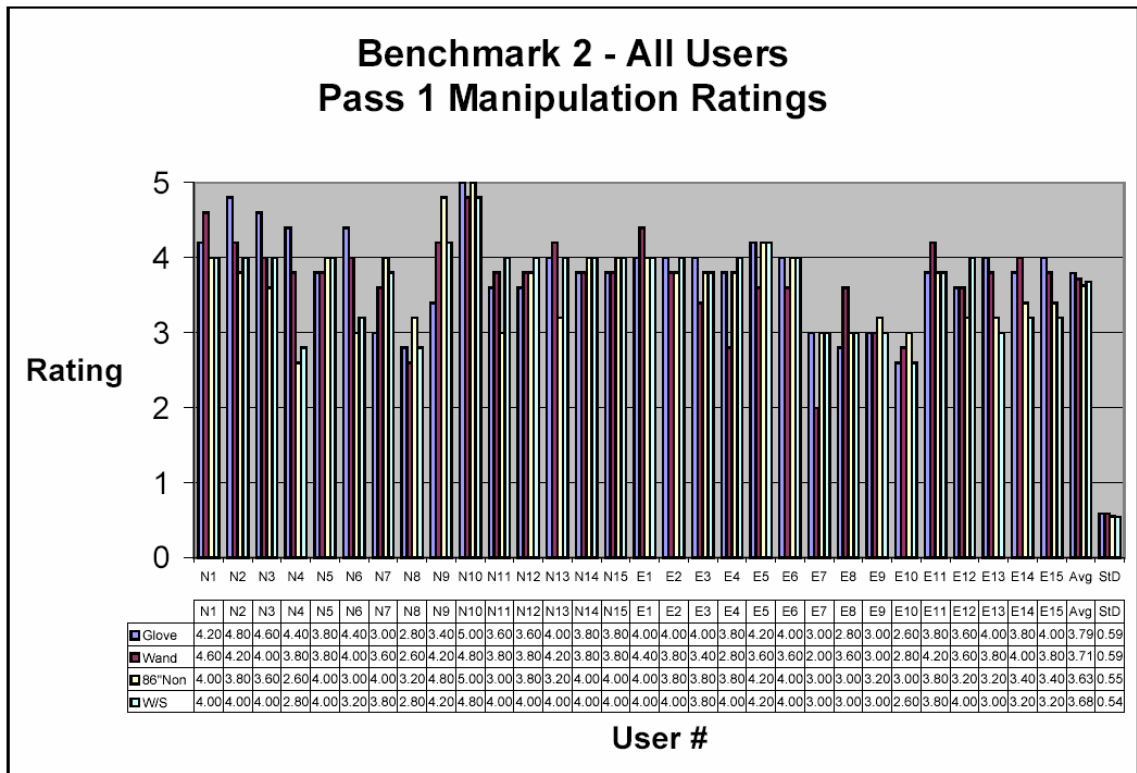
Benchmark 2, All Users, Pass 1, Navigation Ratings											
B2Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	3.93	0.51	2.71	4.71	0.0005	No	12.90%			
Wand	30	3.93	0.62	2.43	5.00	0.0549	Yes	15.73%			
86*Non	30	3.85	0.52	2.71	5.00	0.0079	No	13.61%			
W/S	30	3.78	0.51	2.71	5.00	0.0003	No	13.42%			
Homogeneity of Variance				Roy's Greatest Root:				F(3,27)	1.48	Pr > F	0.2411
B2Ap1		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						Statistically Better
Glove vs Wand				1.0900	0.3095	Yes	29	0.00	0.9481	Yes	Neither
Glove vs 86*Non				0.0300	0.8645	Yes	29	0.40	0.5344	Yes	Neither
Glove vs W/S				0.0000	0.9938	Yes	29	1.87	0.1820	Yes	Neither
Wand vs 86*Non				0.6700	0.4170	Yes	29	0.90	0.3496	Yes	Neither
Wand vs W/S				1.0200	0.3167	Yes	29	2.79	0.1055	Yes	Neither
86*Non vs W/S				0.0300	0.8746	Yes	29	2.16	0.1522	Yes	Neither



**Figure F- 3: B2Ap1Loc All Users Locating Ratings**

**Table F- 3: B2Ap1Loc All Users Locating Ratings Statistics**

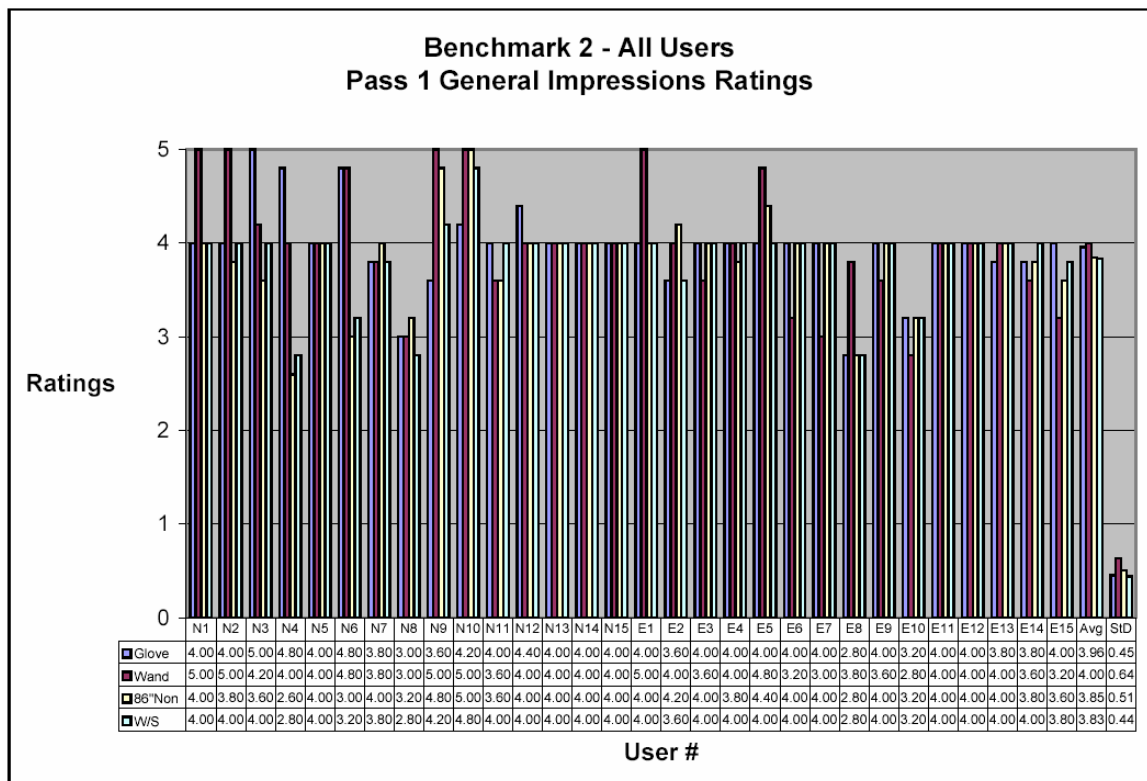
Benchmark 2, All Users, Pass 1, Locating Ratings											
B2Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	3.97	0.44	3.00	4.80	0.0068	No	11.04%			
Wand	30	4.02	0.69	2.60	5.00	0.0633	Yes	17.20%			
86\"Non	30	3.95	0.55	2.60	5.00	0.0034	No	13.98%			
W/S	30	3.79	0.52	2.00	5.00	0.0016	No	13.65%			
B2Ap1	Homogeneity of Variance				Roy's Greatest Root:				F(3,27) 2.79 Pr > F 0.0595		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			6.8000	0.0116	No	29	0.29	0.5966	Yes	Var	Wand 1.18%
Glove vs 86\"Non			1.2300	0.2719	Yes	29	0.04	0.8447	Yes	Neither	Neither
Glove vs W/S			0.3500	0.5558	Yes	29	4.46	0.0434	No	Means	Glove 4.91%
Wand vs 86\"Non			1.6900	0.1991	Yes	29	0.49	0.4910	Yes	Neither	Neither
Wand vs W/S			1.9000	0.1732	Yes	29	5.38	0.0276	No	Means	Wand 6.15%
86\"Non vs W/S			0.7000	0.7978	Yes	29	5.80	0.0226	No	Means	86\"Non 4.38%



**Figure F- 4: B2Ap1Mov All Users Manipulation Ratings**

**Table F- 4: B2Ap1Mov All Users Manipulation Ratings Statistics**

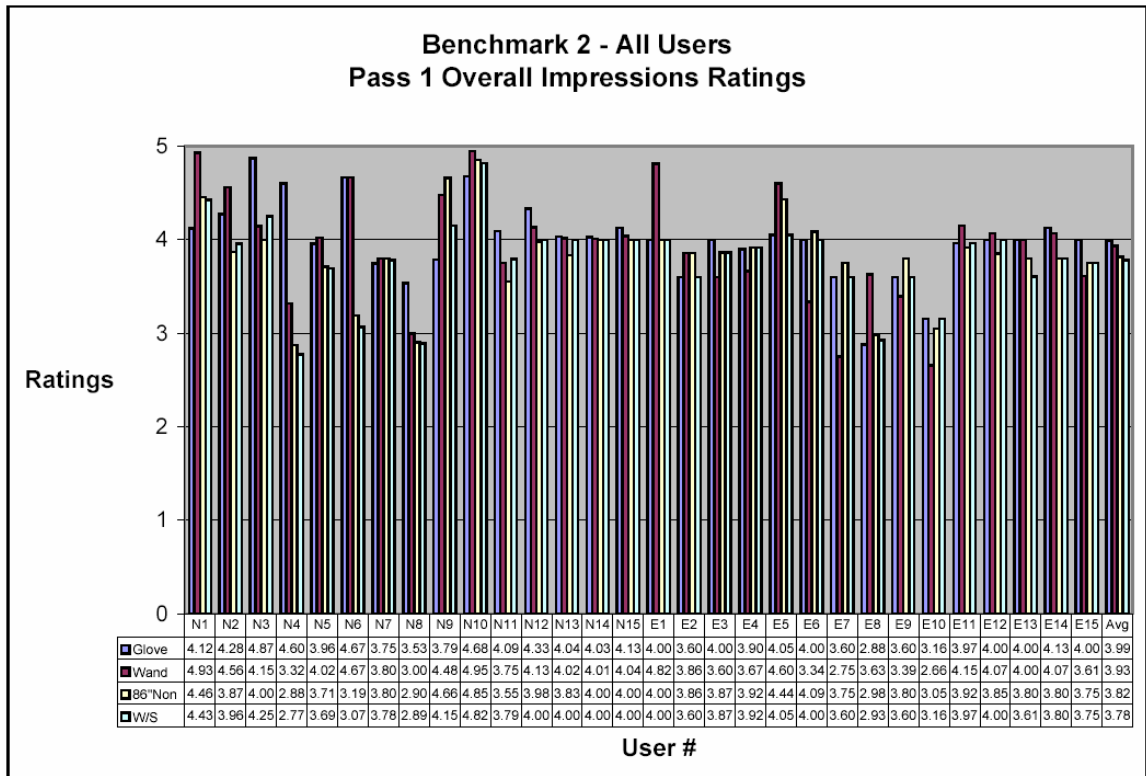
Benchmark 2, All Users, Pass 1, Manipulation Ratings											
B2Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	3.79	0.59	2.60	5.00	0.0250	No	15.51%			
Wand	30	3.71	0.59	2.00	4.80	0.0005	No	15.98%			
86*Non	30	3.63	0.55	2.60	5.00	0.0573	Yes	15.17%			
W/S	30	3.68	0.54	5.60	4.80	< 0.0001	No	14.74%			
B2Ap1		Homogeneity of Variance				Roy's Greatest Root:				F(3,27)	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				0.0000	0.9662	Yes	29	0.86	0.3604	Yes	Neither
Glove vs 86*Non				0.1400	0.7092	Yes	29	1.96	0.1721	Yes	Neither
Glove vs W/S				0.2500	0.6177	Yes	29	1.29	0.2662	Yes	Neither
Wand vs 86*Non				0.1200	0.7254	Yes	29	0.65	0.4267	Yes	Neither
Wand vs W/S				0.2000	0.6574	Yes	29	0.12	0.7308	Yes	Neither
86*Non vs W/S				0.0100	0.9317	Yes	29	0.70	0.4083	Yes	Neither



**Figure F- 5: B2Ap1Gen All Users General Impressions Ratings**

**Table F- 5: B2Ap1Gen All Users General Impressions Ratings Statistics**

Benchmark 2, All Users, Pass 1, General Impressions Ratings												
B2Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	3.96	0.45	2.80	5.00	< 0.0001	No	11.44%				
Wand	30	4.00	0.64	2.80	5.00	0.0002	No	15.92%				
86*Non	30	3.85	0.51	2.60	5.00	< 0.0001	No	13.14%				
W/S	30	3.83	0.44	2.80	4.80	< 0.0001	No	11.57%				
B2Ap1	Homogeneity of Variance					Roy's Greatest Root			F(3,27) 0.96 Pr > F 0.4237			
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			3.1800	0.0799	Yes	29	0.12	0.7336	Yes	Neither	Neither	
Glove vs 86*Non			0.2200	0.6412	Yes	29	0.86	0.3603	Yes	Neither	Neither	
Glove vs W/S			0.0100	0.9251	Yes	29	1.72	0.2001	Yes	Neither	Neither	
Wand vs 86*Non			1.5800	0.2143	Yes	29	1.71	0.2015	Yes	Neither	Neither	
Wand vs W/S			3.6400	0.0612	Yes	29	2.10	0.1584	Yes	Neither	Neither	
86*Non vs W/S			0.3200	0.5722	Yes	29	0.09	0.7634	Yes	Neither	Neither	

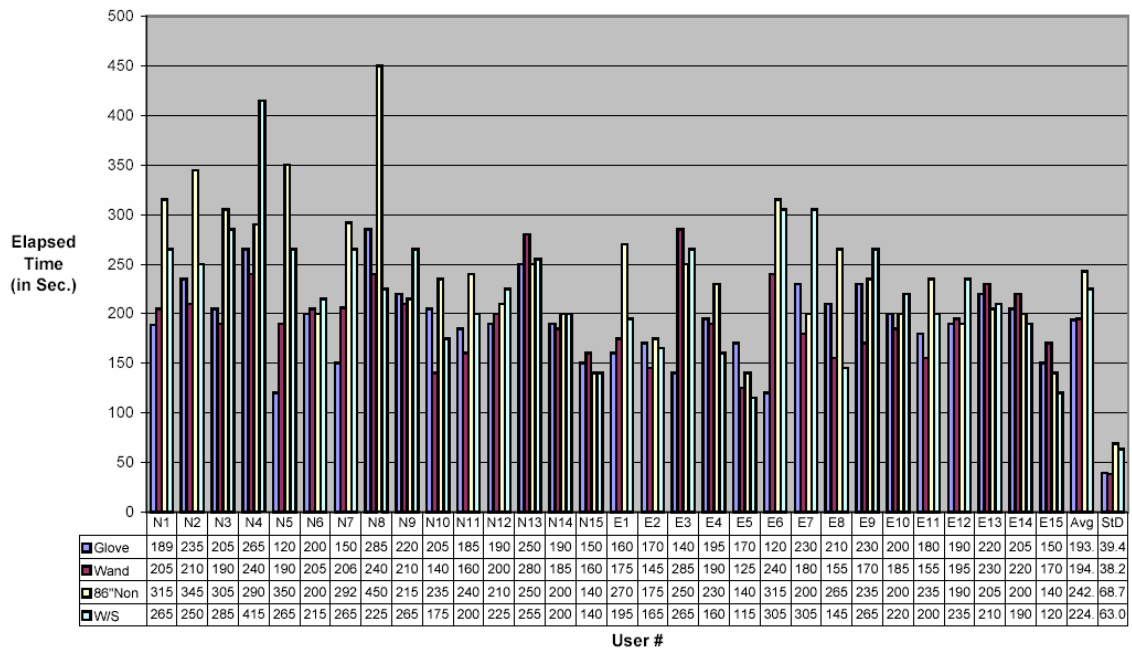


**Figure F- 6: B2Ap1Ovr All Users Pass 1 Overall Impressions Ratings**

**Table F- 6: B2Ap1Ovr All Users Pass 1 Overall Impressions Ratings Statistics**

Benchmark 2, All Users, Pass 1, Overall Impressions Ratings											
B2Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	3.99	0.42	2.88	4.87	0.0261	No	10.46%			
Wand	30	3.94	0.59	2.66	5.95	> 0.1000	Yes	14.96%			
86*Non	30	3.82	0.47	2.88	4.85	0.0019	No	12.24%			
W/S	30	3.78	0.45	2.77	4.82	0.0174	No	11.95%			
B2Ap1	Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	2.60	Pr > F	0.0725	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			2.9200	0.0930	Yes	29	0.39	0.3375	Yes	Neither	Neither
Glove vs 86*Non			1.4900	0.2267	Yes	29	3.45	0.0735	No	Means	Glove 4.53%
Glove vs W/S			0.1400	0.7115	Yes	29	6.03	0.0203	No	Means	Glove 5.55%
Wand vs 86*Non			1.4900	0.2267	Yes	29	1.86	0.1827	Yes	Neither	Neither
Wand vs W/S			1.8900	0.1743	Yes	29	3.16	0.0861	No	Means	Wand 4.05%
86*Non vs W/S			0.0300	0.8699	Yes	29	1.51	0.2284	Yes	Neither	Neither

**Benchmark 2 - All Users  
Pass 2 Elapsed Times**

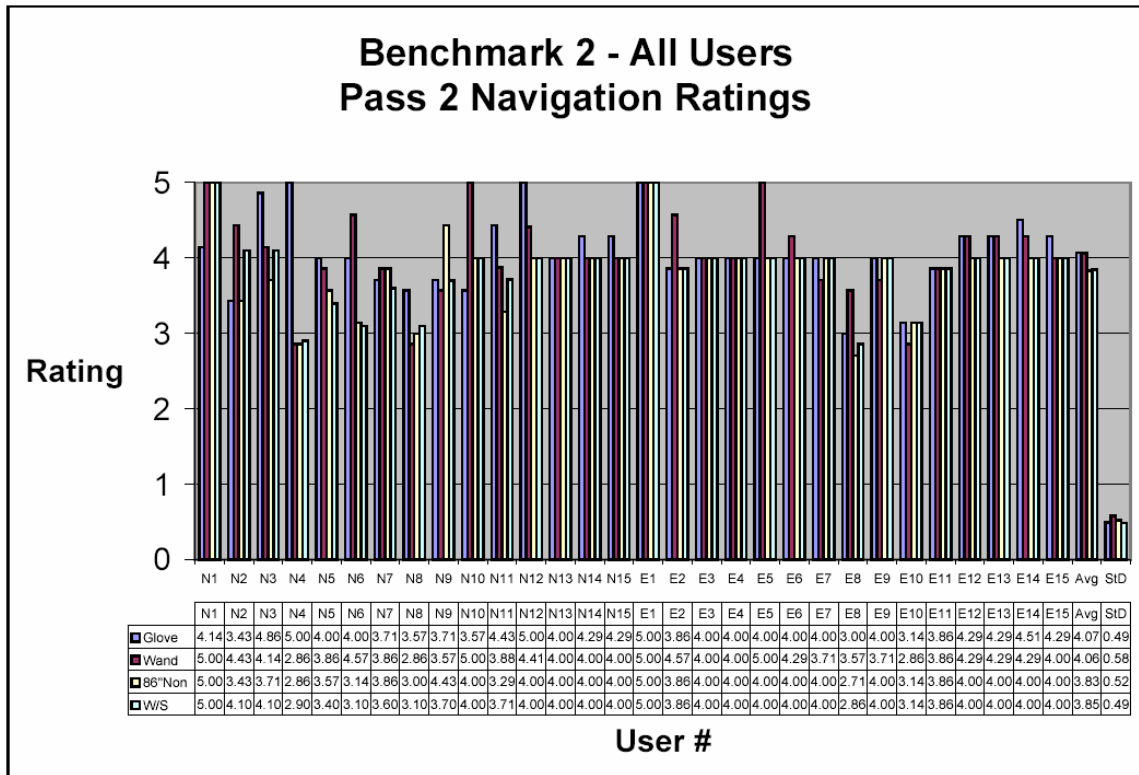


**Figure F- 7: B2Ap2Tim All Users Elapsed Times**

**Table F- 7: B2Ap2Tim All Users Elapsed Times Statistics**

Benchmark 2, All Users, Pass 2, Elapsed Times												
B2Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	193.6	39.44	120.0	285.0	> 0.1000	Yes	20.37%				
Wand	30	194.7	38.18	125.0	285.0	> 0.1000	Yes	19.61%				
86"Non	30	242.9	68.74	140.0	450.0	> 0.1000	Yes	28.30%				
W/S	30	224.7	63.01	115.0	415.0	> 0.1000	Yes	28.04%				
B2Ap2	Homogeneity of Variance				Roy's Greatest Root:			F(3,27)	7.12	Pr > F	0.0011	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand	0.0301	0.8623			Yes	29	0.01	0.9040	Yes	Neither	Neither	
Glove vs 86"Non	8.3814	0.0038			No	29	14.84	0.2384	Yes	Var	Glove 20.30%	
Glove vs W/S	6.0467	0.0140			No	29	7.13	0.0150	No	Var/Mean	Glove 13.84%	
Wand vs 86"Non	9.3382	0.0022			No	29	17.66	0.0002	No	Var/Mean	Wand 19.84%	
Wand vs W/S	6.8772	0.0087			No	29	10.43	0.0031	No	Var/Mean	Wand 13.35%	
86"Non vs W/S	0.2154	0.6425			Yes	29	2.29	0.1412	Yes	Neither	Neither	

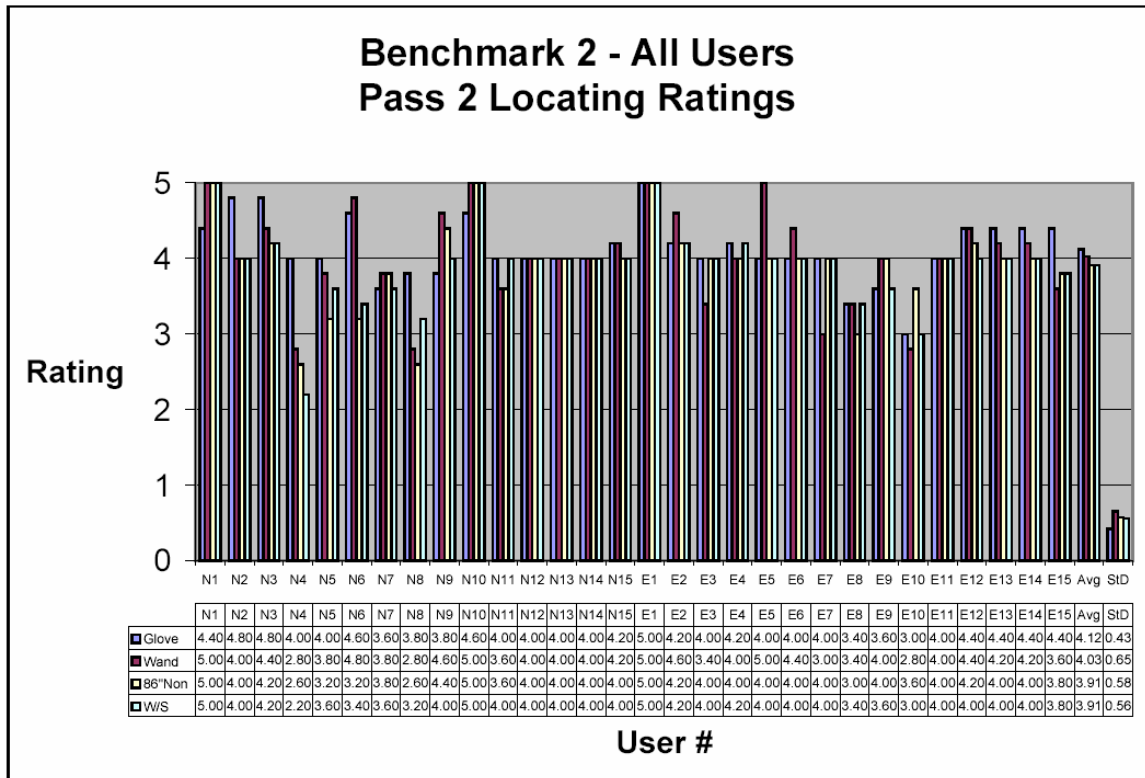




**Figure F- 8: B2Ap2Nav All Users Navigation Ratings**

**Table F- 8: B2Ap2Nav All Users Navigation Ratings Statistics**

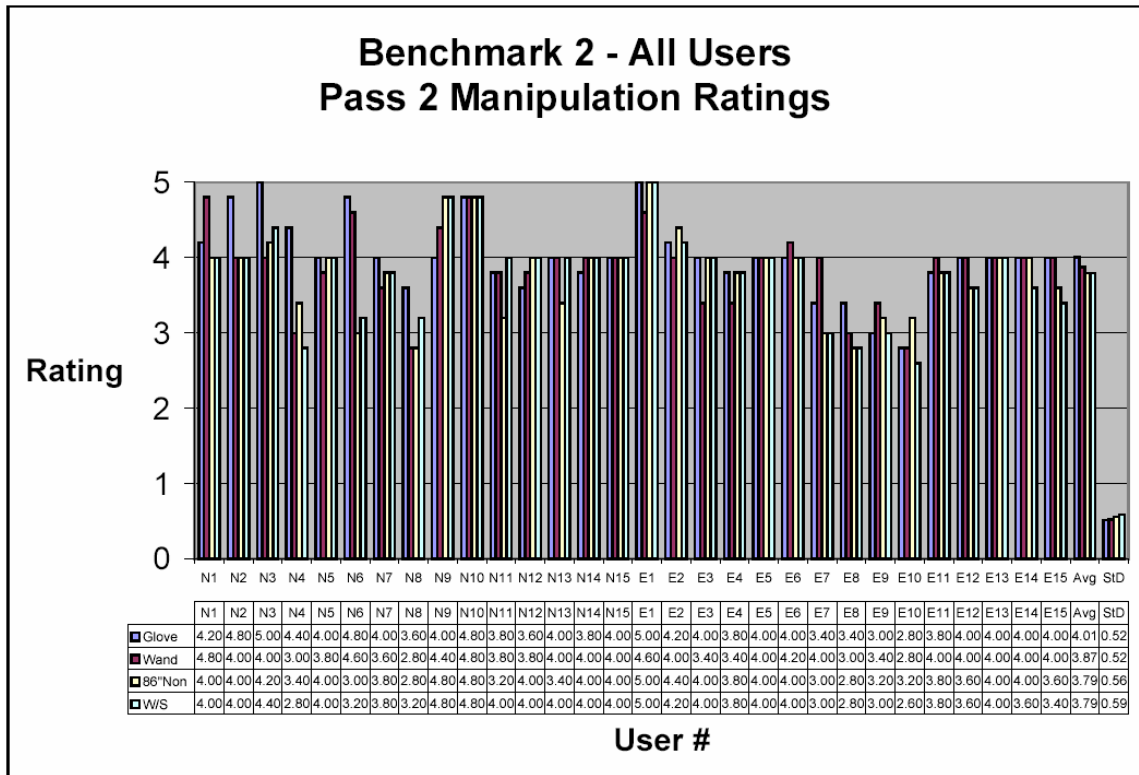
Benchmark 2, All Users, Pass 2, Navigation Ratings											
B2Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.08	0.49	3.00	5.00	0.0485	No	12.14%			
Wand	30	4.06	0.58	2.86	5.00	> 0.1000	Yes	14.27%			
86\"Non	30	3.83	0.52	2.71	5.00	< 0.0001	No	13.64%			
W/S	30	3.85	0.49	2.86	5.00	< 0.0001	No	12.70%			
B2Ap2		Homogeneity of Variance				Roy's Greatest Root:			F(3,27)		
		Bartlett's Test		Levene's Test		df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				0.6800	0.4113	Yes	29	0.01	0.9221	Yes	Neither
Glove vs 86\"Non				0.0700	0.7896	Yes	29	5.34	0.0282	No	Means
Glove vs W/S				0.0000	0.9524	Yes	29	5.31	0.0285	No	Means
Wand vs 86\"Non				0.2800	0.5996	Yes	29	7.03	0.0128	No	Means
Wand vs W/S				0.6900	0.4089	Yes	29	8.00	0.0084	No	Means
86\"Non vs W/S				0.0900	0.7614	Yes	29	0.22	0.6430	Yes	Neither



**Figure F- 9: B2Ap2Loc All Users Locating Ratings**

**Table F- 9: B2Ap2Loc All Users Locating Ratings Statistics**

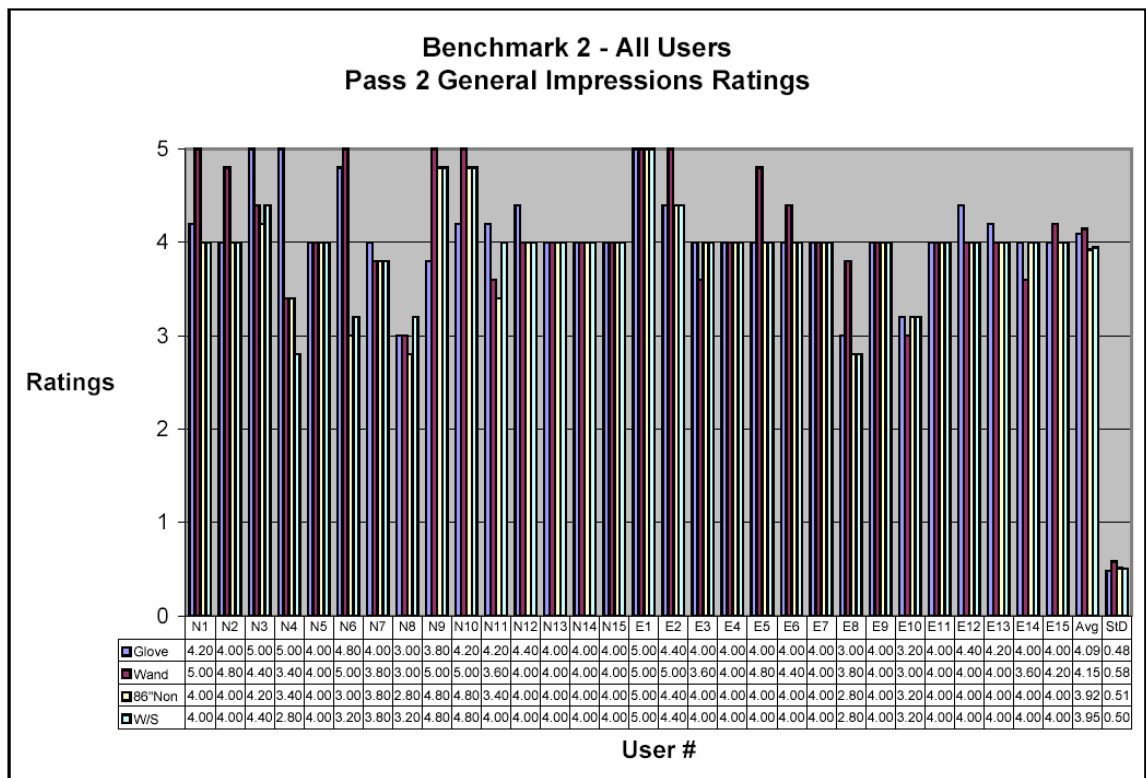
Benchmark 2, All Users, Pass 2, Locating Ratings											
B2Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.12	0.43	3.00	5.00	0.0078	No	10.33%			
Wand	30	4.03	0.65	2.80	5.00	0.0815	Yes	16.22%			
86\"Non	30	3.91	0.58	2.60	5.00	< 0.0001	No	14.75%			
W/S	30	3.91	0.56	2.20	5.00	< 0.0001	No	14.25%			
B2Ap2	Homogeneity of Variance				Roy's Greatest Root:		F(3,27)		1.42	Pr > F	0.1508
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			5.0900	0.0729	Yes	29	0.94	0.3402	Yes	Neither	Neither
Glove vs 86\"Non			1.7700	0.1884	Yes	29	4.49	0.0428	No	Means	Glove 5.29%
Glove vs W/S			1.0200	0.3176	Yes	29	5.84	0.0221	No	Means	Glove 5.29%
Wand vs 86\"Non			0.4600	0.5023	Yes	29	1.25	0.1958	Yes	Neither	Neither
Wand vs W/S			0.5900	0.4467	Yes	29	1.87	0.1819	Yes	Neither	Neither
86\"Non vs W/S			0.0200	0.8874	Yes	29	0.00	1.0000	Yes	Neither	Neither



**Figure F- 10: B2Ap2Mov All Users Manipulation Ratings**

**Table F- 10: B2Ap2Mov All Users Manipulation Ratings Statistics**

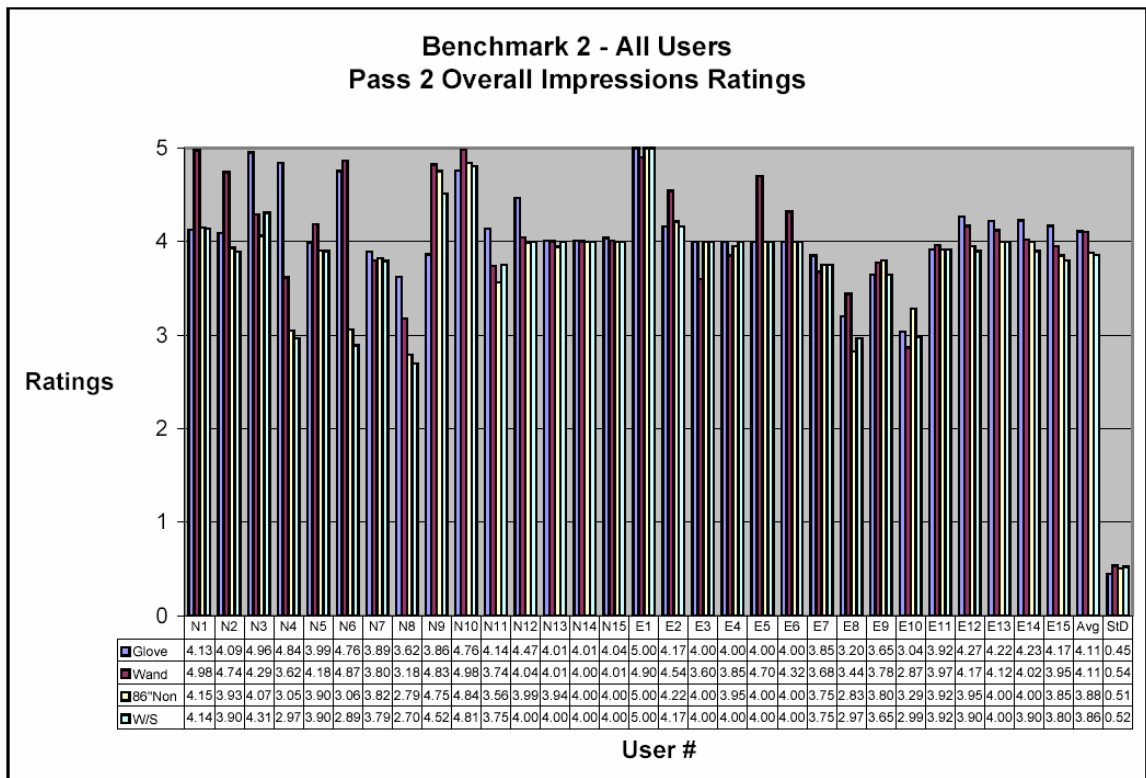
Benchmark 2, All Users, Pass 2, Manipulation Ratings											
B2Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.01	0.52	2.80	5.00	0.0001	No	12.94%			
Wand	30	3.87	0.52	2.80	4.80	0.0003	No	13.53%			
86*Non	30	3.79	0.56	2.80	5.00	0.0073	No	14.82%			
W/S	30	3.79	0.59	2.60	5.00	0.0027	No	15.57%			
B2Ap2	Homogeneity of Variance					Roy's Greatest Root:		F(3,27)	1.85	Pr > F	0.1618
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.0000	0.9574	Yes	29	2.57	0.1204	Yes	Neither	Neither
Glove vs 86*Non			0.1900	0.6633	Yes	29	5.27	0.0291	No	Means	Glove 5.64%
Glove vs W/S			0.4900	0.4864	Yes	29	5.57	0.0253	No	Means	Glove 5.64%
Wand vs 86*Non			0.1600	0.6908	Yes	29	0.83	0.3691	Yes	Neither	Neither
Wand vs W/S			0.4500	0.5040	Yes	29	0.96	0.3356	Yes	Neither	Neither
86*Non vs W/S			0.0800	0.7741	Yes	29	0.00	1.0000	Yes	Neither	Neither



**Figure F- 11: B2Ap2Gen All Users General Impressions Ratings**

**Table F- 11: B2Ap2Gen All Users General Impressions Ratings Statistics**

Benchmark 2, All Users, Pass 2, General Impressions Ratings												
B2Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	4.09	0.48	3.00	5.00	< 0.0001	No	11.74%				
Wand	30	4.15	0.58	3.00	5.00	0.0249	No	14.10%				
86*Non	30	3.92	0.51	2.80	5.00	< 0.0001	No	13.03%				
W/S	30	3.95	0.50	2.80	5.00	< 0.0001	No	12.76%				
B2Ap2		Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	2.32	Pr > F 0.0933		
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				1.1500	0.2870	Yes	29	0.27	0.6078	Yes	Neither	Neither
Glove vs 86*Non				0.0800	0.7803	Yes	29	3.24	0.0821	No	Means	Glove 4.41%
Glove vs W/S				0.0400	0.8330	Yes	29	2.07	0.1608	Yes	Neither	Neither
Wand vs 86*Non				0.5700	0.4533	Yes	29	6.52	0.0162	No	Means	Wand 5.79%
Wand vs W/S				0.6800	0.4128	Yes	29	5.06	0.0323	No	Means	Wand 5.07%
86*Non vs W/S				0.0000	0.9422	Yes	29	0.66	0.4235	Yes	Neither	Neither

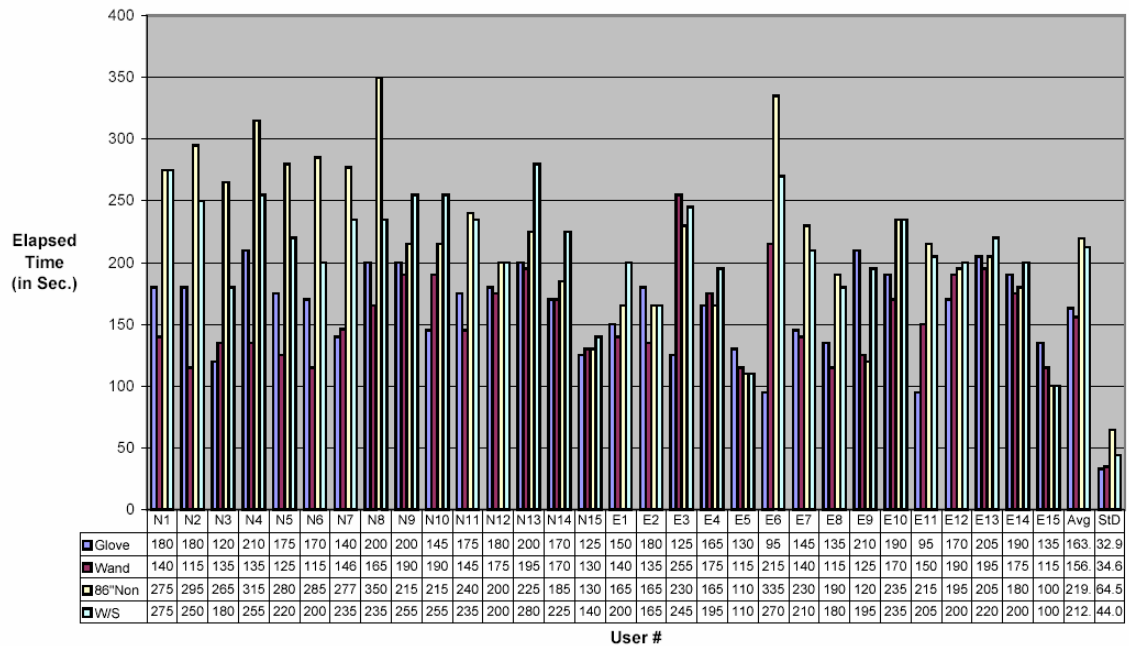


**Figure F- 12: B2Ap2Ovr All Users Overall Impressions Ratings**

**Table F- 12: B2Ap2Ovr All Users Overall Impressions Ratings Statistics**

Benchmark 2, All Users, Pass 2, Overall Impressions Ratings										
B2Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	30	4.11	0.45	3.04	5.00	0.0465	No	10.89%		
Wand	30	4.11	0.53	2.87	4.98	> 0.1000	Yes	13.02%		
86*Non	30	3.88	0.51	2.79	5.00	0.0021	No	13.04%		
W/S	30	3.86	0.52	2.70	5.00	0.0008	No	13.53%		
B2Ap2		Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	3.43	Pr > F 0.0311
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					
Glove vs Wand				0.9100	0.3430	Yes	29	0.00	0.9646	Yes
Glove vs 86*Non				0.3300	0.5701	Yes	29	5.75	0.0231	No
Glove vs W/S				0.5300	0.4686	Yes	29	7.16	0.0122	No
Wand vs 86*Non				0.0800	0.7719	Yes	29	8.41	0.0070	No
Wand vs W/S				0.0200	0.9007	Yes	29	9.50	0.0045	No
86*Non vs W/S				0.0200	0.8805	Yes	29	1.34	0.2562	Yes
										Significant?
										Statistically Better
Glove vs Wand										Neither
Glove vs 86*Non										Means
Glove vs W/S										Means
Wand vs 86*Non										Means
Wand vs W/S										Means
86*Non vs W/S										Neither

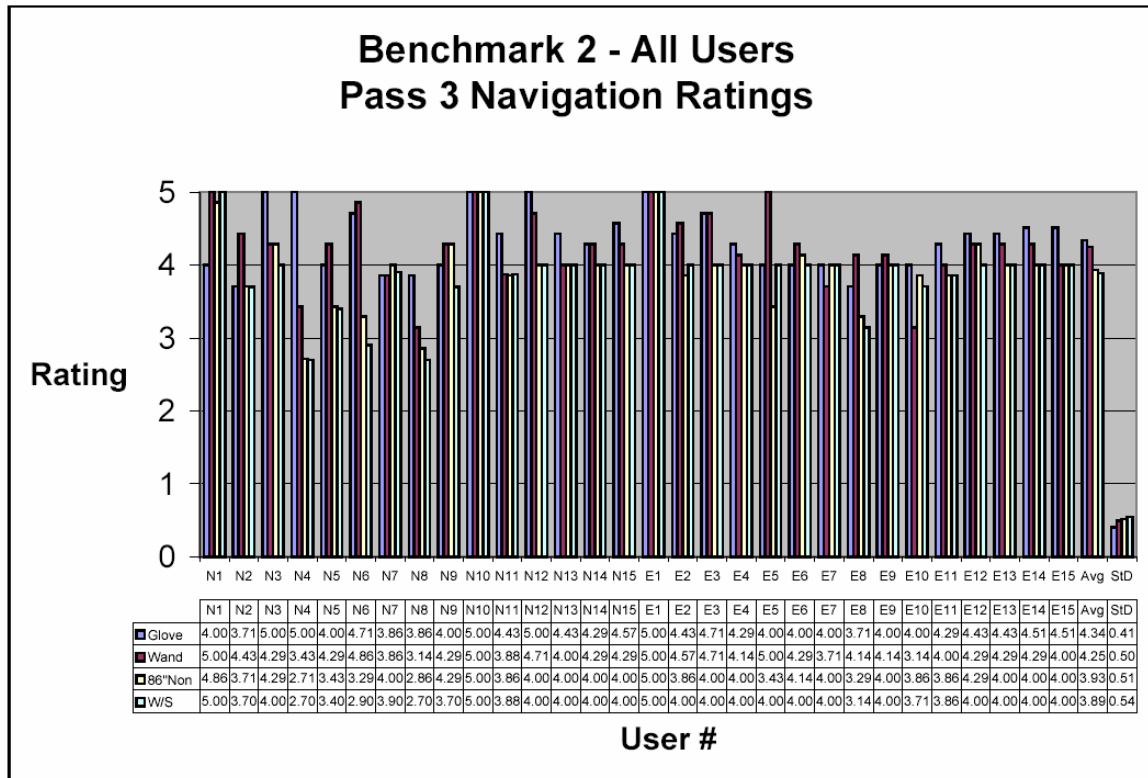
**Benchmark 2 - All Users  
Pass 3 Elapsed Times**



**Figure F- 13: B2Ap3Tim All Users Elapsed Times**

**Table F- 13: B2Ap3Tim All Users Elapsed Times Statistics**

Benchmark 2, All Users, Pass 3, Elapsed Times											
B2Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	163.0	32.87	95.0	210.0	0.079	Yes	20.17%			
Wand	30	156.0	34.59	115.0	255.0	0.0948	Yes	22.17%			
86"Non	30	219.7	64.45	100.0	350.0	> 0.1000	Yes	29.34%			
W/S	30	212.3	44.02	100.0	280.0	0.0975	Yes	20.73%			
B2Ap3	Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	22.93	Pr > F	<0.0001	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.0743	0.7852			Yes	29	0.06	0.4305	Yes	Neither	Neither
Glove vs 86"Non	12.0553	0.0005			No	29	19.99	0.0001	No	Var/Means	Glove 25.81%
Glove vs W/S	2.3978	0.1215			Yes	29	34.95	<0.0001	Yes	Neither	Neither
Wand vs 86"Non	10.3951	0.0013			No	29	26.39	<0.0001	Yes	Var	Wand 28.99%
Wand vs W/S	1.6398	0.2004			Yes	29	60.90	<0.0001	Yes	Neither	Neither
86"Non vs W/S	4.0496	0.0442			No	29	0.82	0.3739	Yes	Var	W/S 3.37%



**Figure F- 14: B2Ap3Nav All Users Navigation Ratings**

**Table F- 14: B2Ap3Nav All Users Navigation Ratings Statistics**

Benchmark 2, All Users, Pass 3, Navigation Ratings											
B2Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.34	0.41	3.71	5.00	0.0041	No	9.37%			
Wand	30	4.25	0.50	3.14	5.00	0.0321	No	11.67%			
86*Non	30	3.93	0.51	2.71	5.00	0.0010	No	13.07%			
W/S	30	3.89	0.54	2.70	5.00	< 0.0001	No	14.02%			
B2Ap3		Homogeneity of Variance				Roy's Greatest Root:			F(3,27)		
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				1.2700	0.2643	Yes	29	0.86	0.3613	Yes	Neither
Glove vs 86*Non				1.2900	0.2614	Yes	29	15.32	0.0005	No	Means
Glove vs W/S				1.7200	0.1943	Yes	29	17.19	0.0003	No	Means
Wand vs 86*Non				0.0300	0.8616	Yes	29	12.12	0.0016	No	Means
Wand vs W/S				0.1900	0.6615	Yes	29	17.11	0.0003	No	Means
86*Non vs W/S				0.0700	0.7969	Yes	29	1.88	0.1806	Yes	Neither

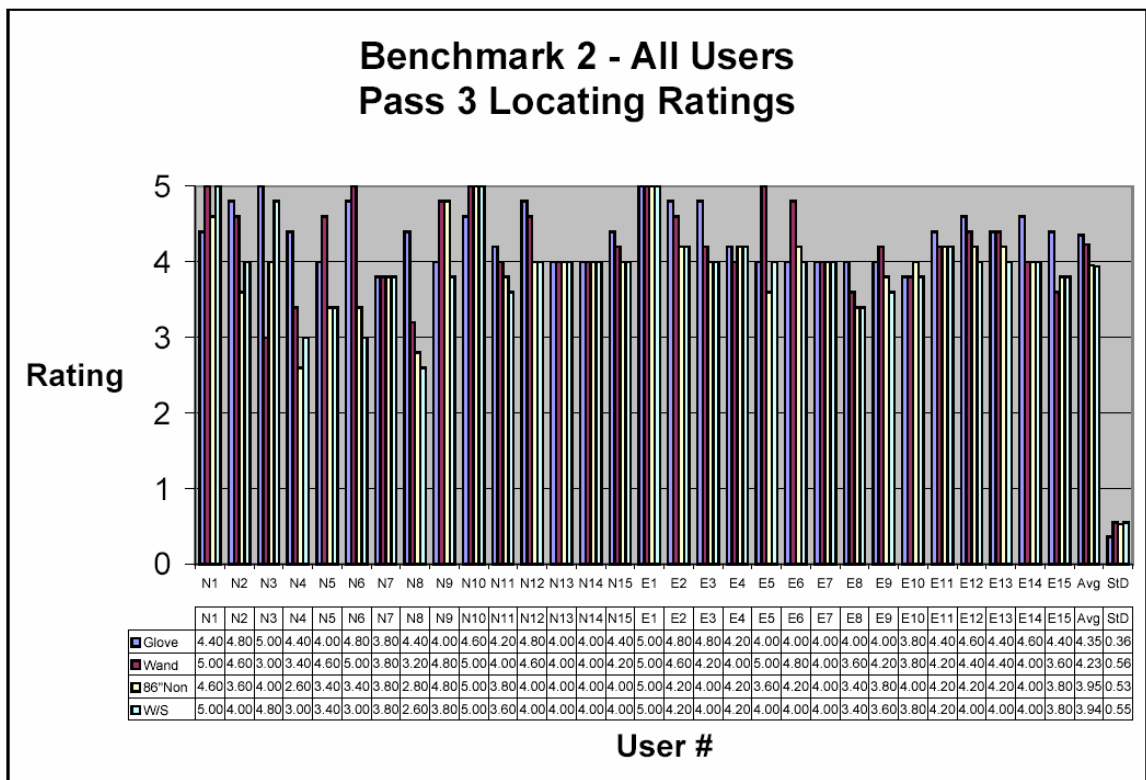
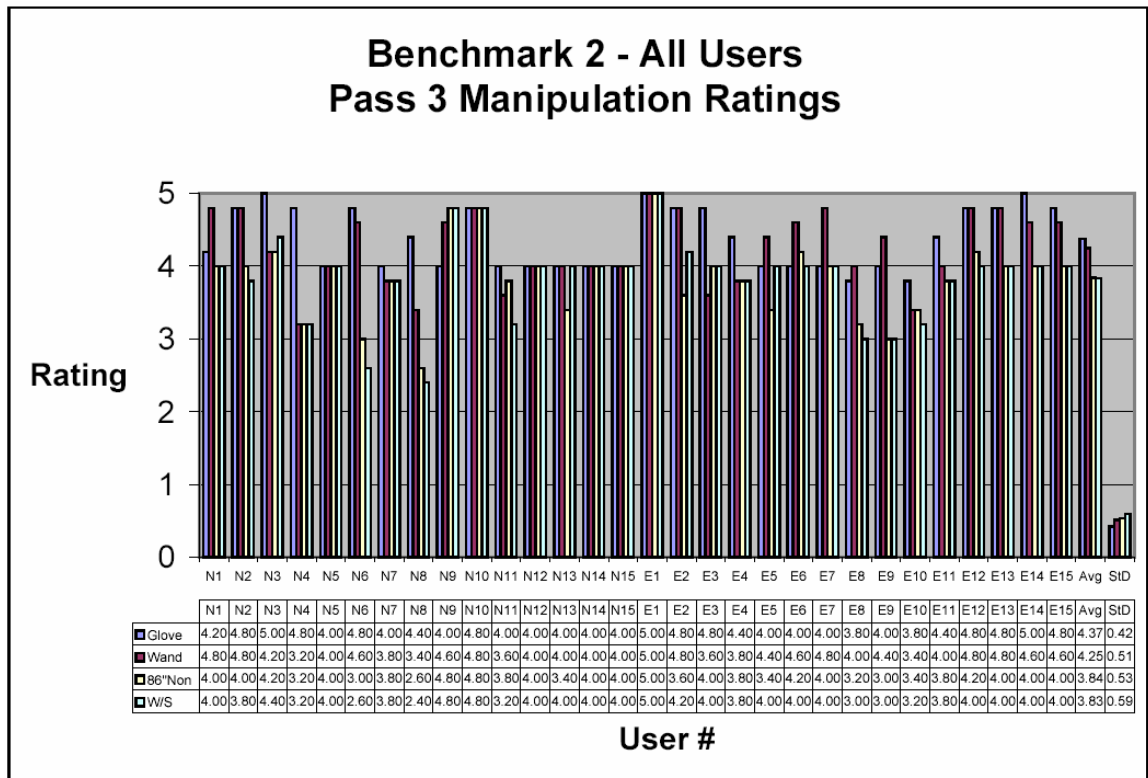


Figure F- 15: B2Ap3Loc All Users Locating Ratings

Table F- 15: B2Ap3Loc All Users Locating Ratings Statistics

Benchmark 2, All Users, Pass 3, Locating Ratings											
B2Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.35	0.36	3.80	5.00	0.0031	No	8.33%			
Wand	30	4.23	0.56	3.00	5.00	> 0.1000	Yes	13.14%			
86*Non	30	3.95	0.53	2.60	5.00	0.0087	No	13.40%			
W/S	30	3.94	0.55	2.60	5.00	0.0005	No	13.99%			
B2Ap3	Homogeneity of Variance					Roy's Greatest Root:		F(3,27)	5.70	Pr > F	0.0037
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			6.2900	0.0149	No	29	1.13	0.2966	Yes	Var	Glove 2.83%
Glove vs 86*Non			2.6700	0.1079	Yes	29	13.70	0.0609	No	Means	Glove 10.12%
Glove vs W/S			3.3800	0.0712	Yes	29	16.75	0.0003	No	Means	Glove 10.48%
Wand vs 86*Non			0.0700	0.7970	Yes	29	8.56	0.0066	No	Means	Wand 7.08%
Wand vs W/S			0.0000	0.9626	Yes	29	6.70	0.0149	No	Means	Wand 7.44%
86*Non vs W/S			0.0300	0.8546	Yes	29	0.06	0.8130	Yes	Neither	Neither

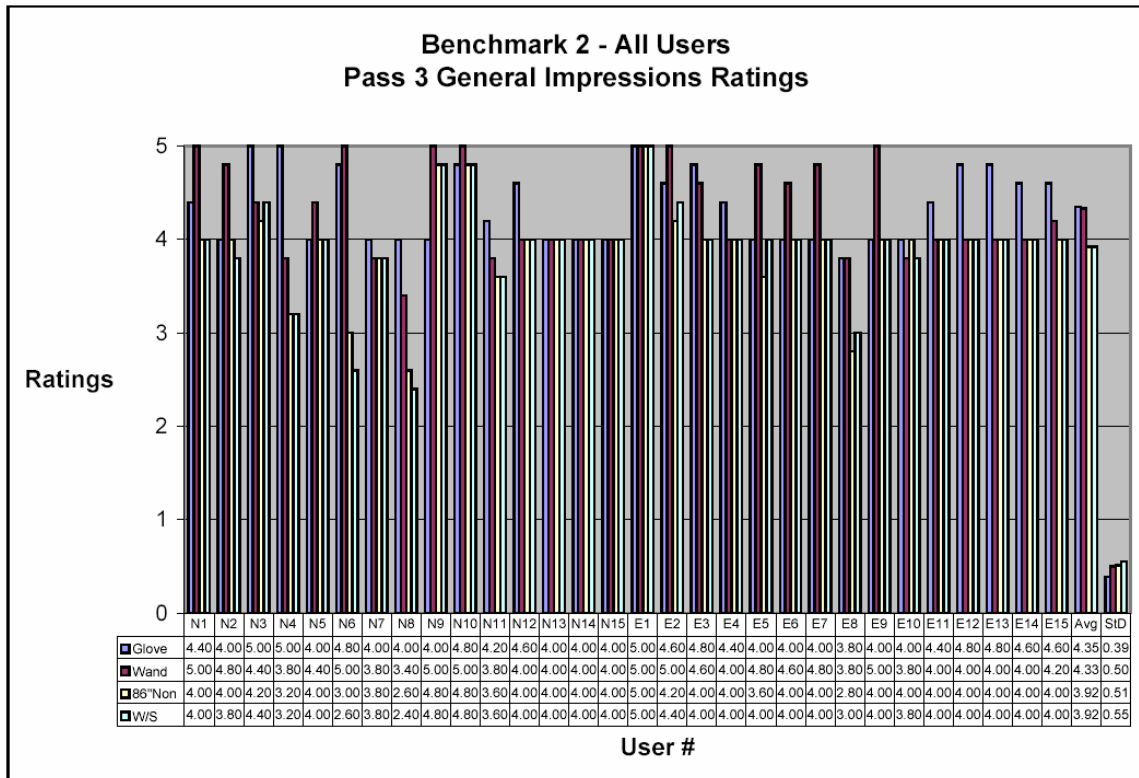




**Figure F- 16: B2Ap3Mov All Users Manipulation Ratings**

**Table F- 16: B2Ap3Mov All Users Manipulation Ratings Statistics**

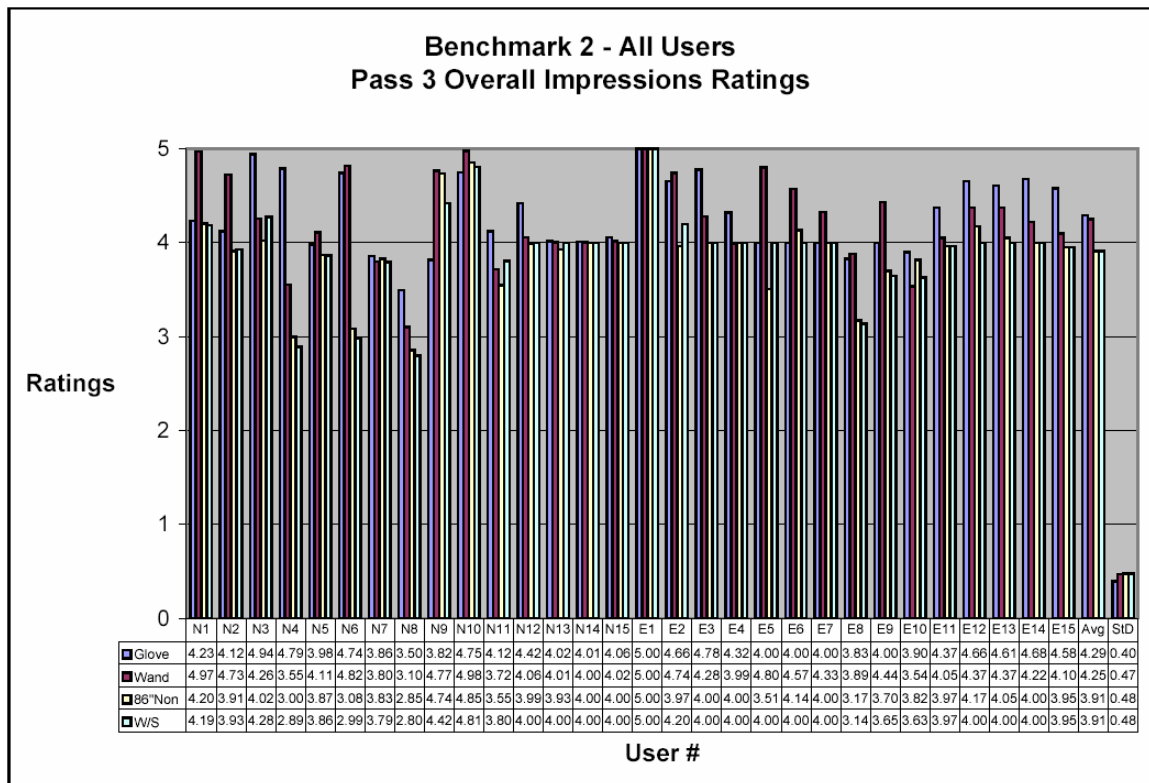
Benchmark 2, All Users, Pass 3, Manipulation Ratings											
B2Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.37	0.42	3.80	5.00	< 0.0001	No	9.66%			
Wand	30	4.25	0.51	3.20	5.00	0.008	No	12.03%			
86*Non	30	3.84	0.53	2.60	5.00	0.0106	No	13.91%			
W/S	30	3.83	0.59	2.40	5.00	< 0.0001	No	15.51%			
B2Ap3		Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	8.79	Pr > F	0.0003
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				2.5900	0.1132	Yes	29	1.67	0.2065	Yes	Neither
Glove vs 86*Non				1.7100	0.1964	Yes	29	24.10	<0.0001	Yes	Neither
Glove vs W/S				2.9100	0.0933	Yes	29	21.12	<0.0001	Yes	Neither
Wand vs 86*Non				0.0700	0.7954	Yes	29	18.64	0.0002	No	Means
Wand vs W/S				0.6900	0.4105	Yes	29	17.54	0.0002	No	Means
86*Non vs W/S				0.2800	0.5967	Yes	29	0.02	0.8869	Yes	Neither



**Figure F- 17: B2Ap3Gen All Users General Impressions Ratings**

**Table F- 17: B2Ap3Gen All Users General Impressions Ratings Statistics**

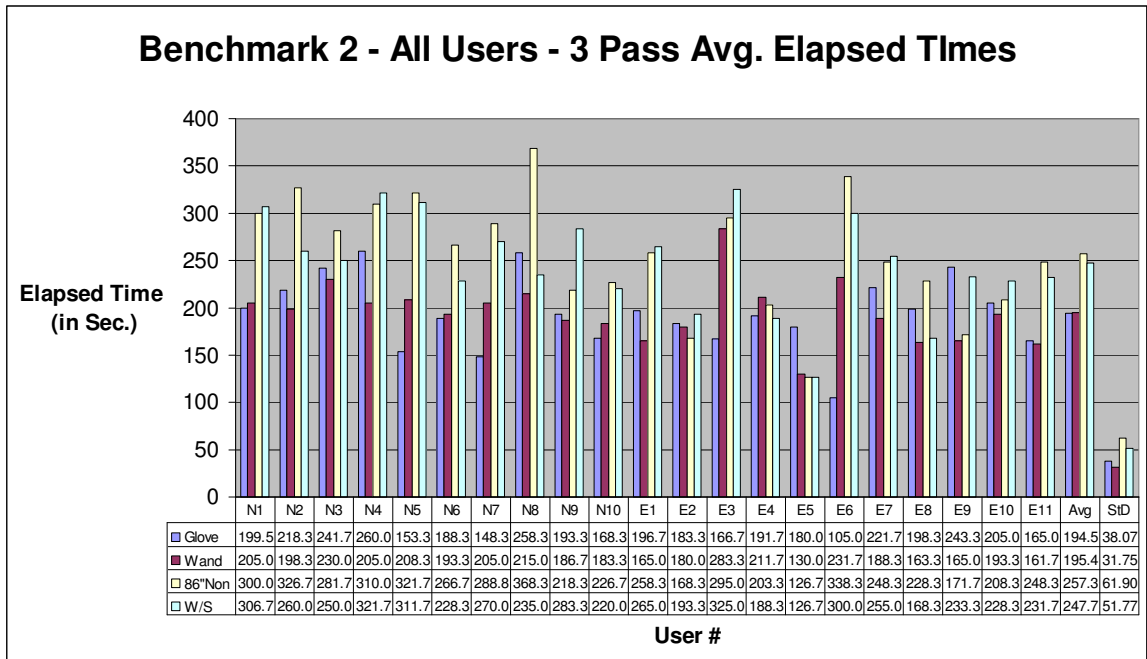
Benchmark 2, All Users, Pass 3, General Impressions Ratings												
B2Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	4.35	0.39	3.80	5.00	< 0.0001	No	9.00%				
Wand	30	4.33	0.50	3.40	5.00	< 0.0001	No	11.52%				
86\"Non	30	3.92	0.51	2.60	5.00	< 0.0001	No	13.03%				
W/S	30	3.92	0.55	2.40	5.00	< 0.0001	No	14.09%				
B2Ap3	Homogeneity of Variance					Roy's Greatest Root:			F(3,27)		Pr > F	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			5.1500	0.0270	No	29	0.03	0.8524	Yes	Var	Glove	0.46%
Glove vs 86\"Non			1.4100	0.2402	Yes	29	17.59	0.0002	No	Means	Glove	11.05%
Glove vs W/S			1.9000	0.1730	Yes	29	15.34	0.0005	No	Means	Glove	11.05%
Wand vs 86\"Non			0.0200	0.9019	Yes	29	20.21	0.0001	No	Means	Wand	10.54%
Wand vs W/S			0.2400	0.6246	Yes	29	17.88	0.0002	No	Means	Wand	10.54%
86\"Non vs W/S			0.1000	0.7520	Yes	29	0.00	1.0000	Yes	Neither	Neither	



**Figure F- 18: B2Ap3Ovr All Users Overall Impressions Ratings**

**Table F- 18: B2Ap3Ovr All Users Overall Impressions Ratings Statistics**

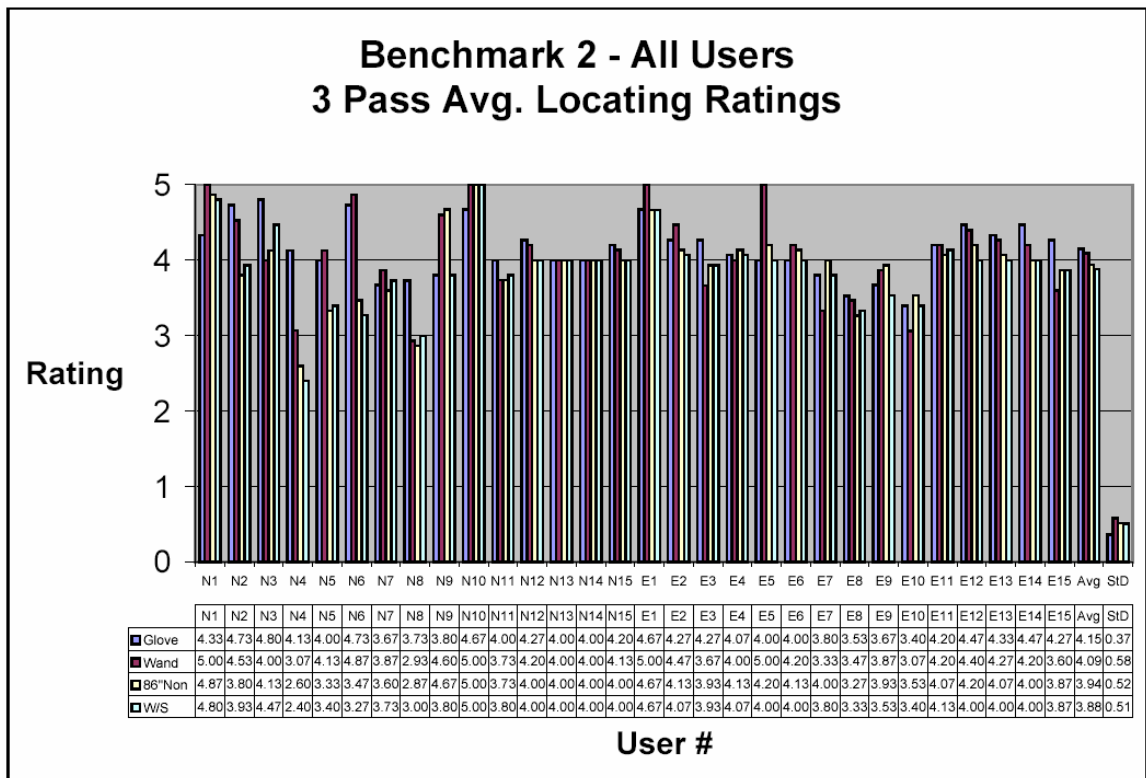
Benchmark 3, All Users, Pass 3, Overall Impressions Ratings											
B3Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.31	0.56	3.06	5.00	> 0.1000	Yes	12.96%			
Wand	30	4.11	0.52	3.36	5.00	< 0.0001	No	12.73%			
86\"Non	30	3.86	0.32	3.00	4.74	0.0080	No	8.32%			
W/S	30	3.72	0.41	2.87	5.00	0.0778	Yes	10.95%			
B3Ap3	Bartlett's Test		Levene's Test		Equal Var ?	Roy's Greatest Root:		F(3,27)	3.97	Pr > F	0.0183
	Chi-Sq	Pr > ChiSq	F Value	Pr > F		df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
Glove vs Wand			0.2900	0.6296	Yes	29	0.07	0.8033	Yes	Neither	Neither
Glove vs 86\"Non			9.1700	0.0037	No	29	6.13	0.0194	No	Var/Mean	Glove 10.37%
Glove vs W/S	2.7950	0.0946			Yes	29	9.46	0.0046	No	Means	Glove 13.64%
Wand vs 86\"Non			6.6900	0.0122	No	29	9.59	0.0043	No	Var/Mean	Wand 5.96%
Wand vs W/S			1.6900	0.1993	Yes	29	11.60	0.0019	No	Means	Wand 9.40%
86\"Non vs W/S			0.7500	0.3902	Yes	29	7.79	0.0092	No	Means	86\"Non 3.65%



**Figure F- 19: B2A3pAvgTim All Users Elapsed Times**

**Table F- 19: B2A3pAvgTim All Users Elapsed Times Statistics**

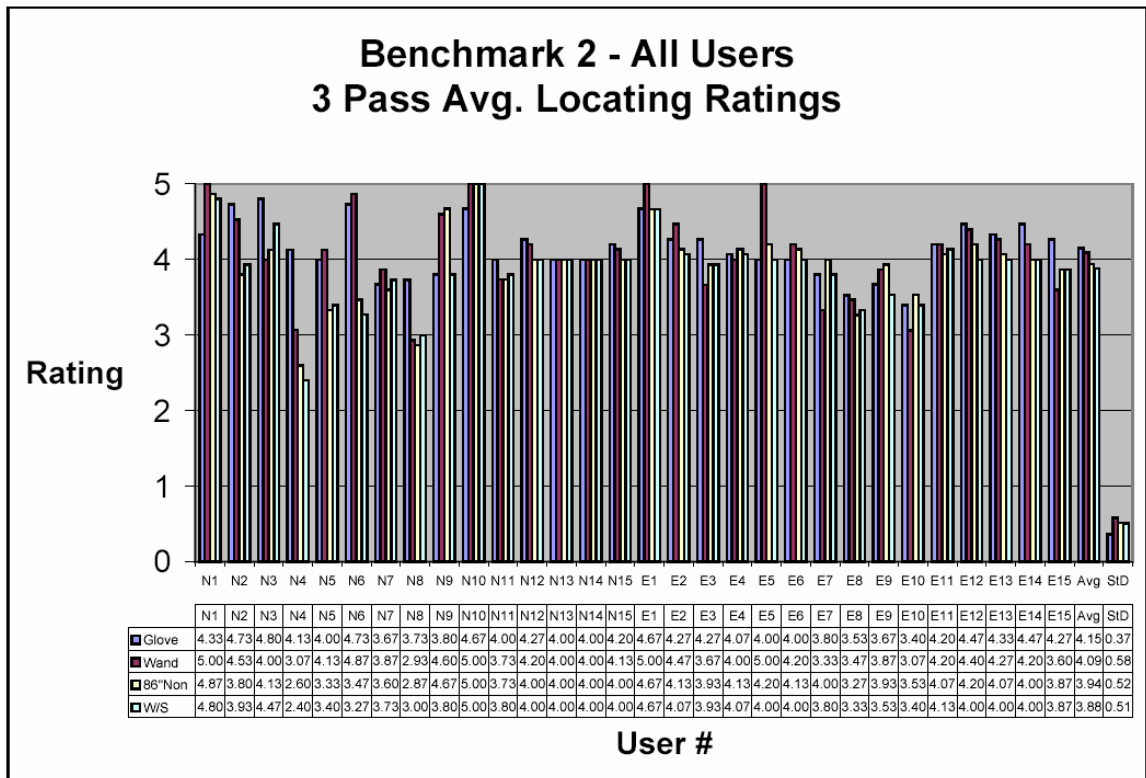
Benchmark 2, All Users, 3 Pass Avg., Elapsed Times												
B2A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	194.2	35.18	105.0	260.0	> 0.1000	Yes	18.12%				
Wand	30	194.2	32.38	130.0	283.3	> 0.1000	Yes	16.67%				
86*Non	30	240.5	61.62	126.7	368.3	> 0.1000	Yes	25.62%				
W/S	30	237.2	53.06	120.0	325.0	> 0.1000	Yes	22.37%				
B2A3pA		Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	11.72	Pr > F	<0.0001	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		0.1965	0.6576			Yes	29	0.00	0.9961	Yes	Neither	Neither
Glove vs 86*Non		8.5225	0.0035			No	29	14.65	0.0006	No	Var/Means	Glove 19.25%
Glove vs W/S		4.6844	0.0304			No	29	15.50	0.0005	No	Var/Means	Glove 18.13%
Wand vs 86*Non		11.0718	0.0009			No	29	26.57	<0.0001	Yes	Var	Wand 19.25%
Wand vs W/S		6.6912	0.0097			No	29	34.39	<0.0001	Yes	Var	Wand 18.13%
86*Non vs W/S		0.6352	0.4255			Yes	29	0.22	0.6407	Yes	Neither	Neither



**Figure F- 20: B2A3pAvgNav All Users Navigation Ratings**

**Table F- 20: B2A3pAvgNav All Users Navigation Ratings Statistics**

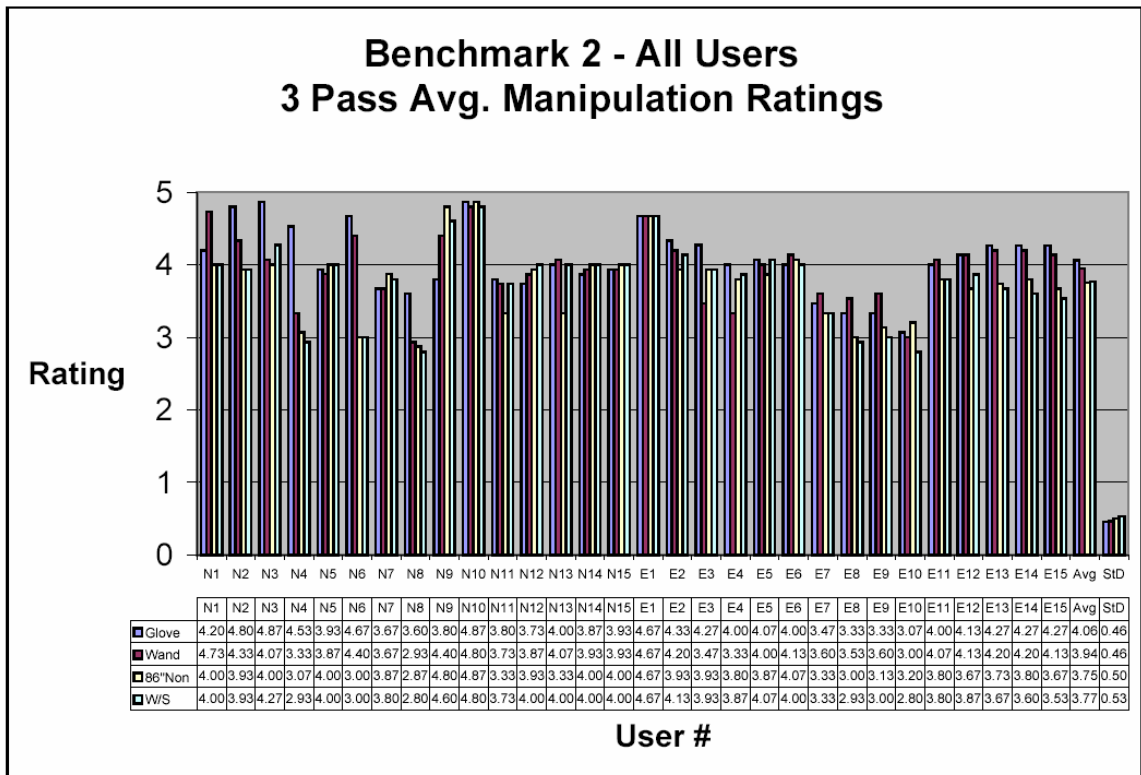
Benchmark 2, All Users, 3 Pass Avg., Navigation Ratings										
B2A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	30	4.11	0.42	3.14	4.86	> 0.1000	Yes	10.19%		
Wand	30	4.08	0.54	2.81	5.00	0.0228	No	13.14%		
86\"Non	30	3.87	0.48	2.76	4.95	0.005	No	12.45%		
W/S	30	3.84	0.48	2.83	4.90	< 0.0001	No	12.56%		
B2A3pA	Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	4.32	Pr > F	0.0130
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand			1.4900	0.2267	Yes	29	0.09	0.7692	Yes	Neither
Glove vs 86\"Non			0.4900	0.4877	Yes	29	5.41	0.0272	No	Means
Glove vs W/S			0.5100	0.4778	Yes	29	7.48	0.0105	No	Means
Wand vs 86\"Non			0.2700	0.6085	Yes	29	7.74	0.0094	No	Means
Wand vs W/S			0.2700	0.6023	Yes	29	10.53	0.0030	No	Means
86\"Non vs W/S			0.0000	0.9988	Yes	29	1.15	0.2929	Yes	Neither



**Figure F- 21: B2A3pAvg All Users Locating Ratings**

**Table F- 21: B2A3pAvg All Users Locating Ratings Statistics**

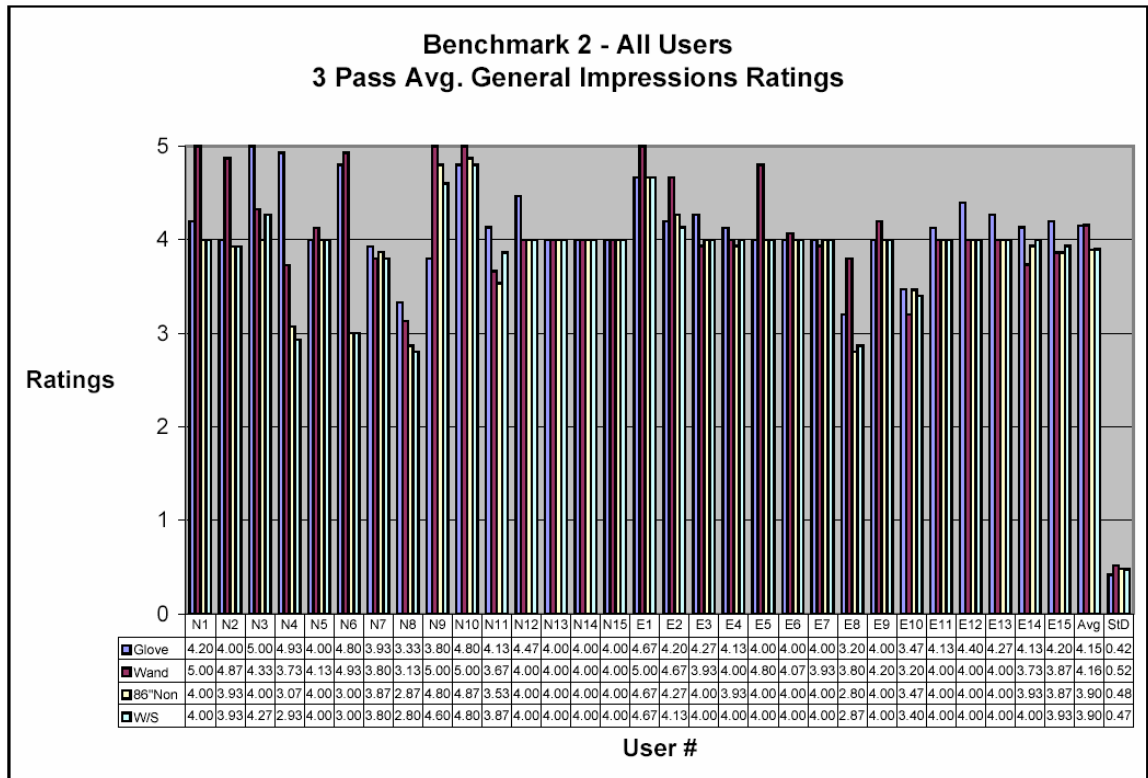
Benchmark 2, All Users, 3 Pass Avg., Locating Ratings											
B2A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.15	0.37	3.40	4.80	> 0.1000	Yes	8.88%			
Wand	30	4.09	0.58	2.93	5.00	> 0.1000	Yes	14.17%			
86\"Non	30	3.94	0.52	2.60	5.00	0.0212	No	13.10%			
W/S	30	3.88	0.51	2.40	5.00	0.008	No	13.20%			
B2A3pA	Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	4.20	Pr > F	0.0146	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	5.6792	0.0172	2.2100	0.1425	Yes	29	0.44	0.5137	Yes	Neither	Neither
Glove vs 86\"Non			1.7800	0.1879	Yes	29	5.16	0.0307	No	Means	Glove 5.30%
Glove vs W/S			0.3900	0.5334	Yes	29	10.51	0.0030	No	Means	Glove 6.93%
Wand vs 86\"Non			0.4000	0.5306	Yes	29	4.38	0.0451	No	Means	Wand 3.91%
Wand vs W/S			0.0000	0.9738	Yes	29	6.95	0.0133	No	Means	Wand 5.52%
86\"Non vs W/S						29	2.52	0.0123	No	Means	86\"Non 1.55%



**Figure F- 22: B2A3pAvgMov All Users Manipulation Ratings**

**Table F- 22: B2A3pAvgMov All Users Manipulation Ratings Statistics**

Benchmark 2, All Users, 3 Pass Avg., Manipulation Ratings											
B2A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.06	0.46	3.07	4.87	> 0.1000	Yes	11.38%			
Wand	30	3.94	0.46	2.93	4.50	> 0.1000	Yes	11.75%			
86*Non	30	3.75	0.50	2.87	4.87	0.0152	No	13.44%			
W/S	30	3.77	0.53	2.80	4.80	0.0579	Yes	14.14%			
B2A3pA		Homogeneity of Variance				Roy's Greatest Root:		F(3,27)	3.46	Pr > F	0.0300
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand		0.0004	0.9838			Yes	29	2.58	0.1191	Yes	Neither
Glove vs 86*Non				0.2500	0.6184	Yes	29	10.56	0.0029	No	Means
Glove vs W/S		0.5862	0.4439			Yes	29	9.37	0.0038	No	Means
Wand vs 86*Non				0.2200	0.6417	Yes	29	6.88	0.0138	No	Means
Wand vs W/S		0.5558	0.4560			Yes	29	5.87	0.0219	No	Means
86*Non vs W/S				0.1000	0.7498	Yes	29	0.18	0.6735	Yes	Neither

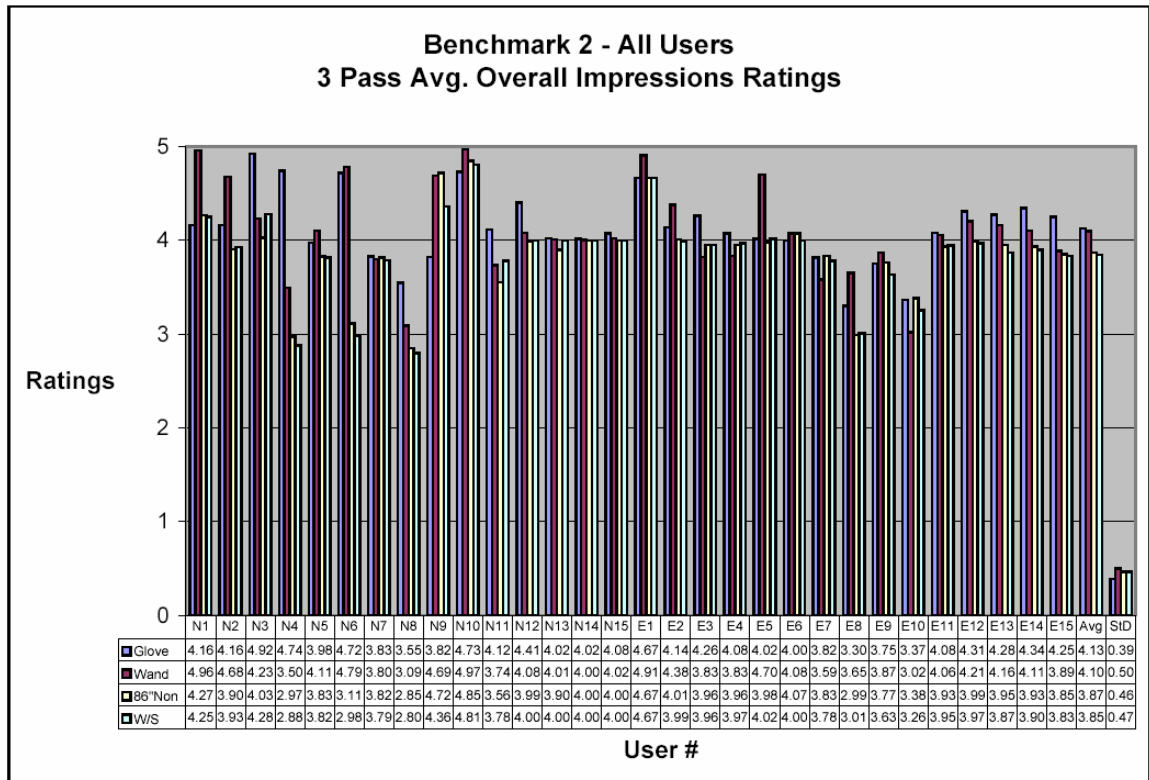


**Figure F- 23: B2A3pAvgGen All Users General Impressions Ratings**

**Table F- 23: B2A3pAvgGen All Users General Impressions Ratings Statistics**

Benchmark 2, All Users, 3 Pass Avg., General Impressions Ratings											
B2A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.15	0.42	3.20	5.00	0.0053	No	10.07%			
Wand	30	4.16	0.52	3.13	5.00	0.0007	No	12.53%			
86\"Non	30	3.90	0.48	2.80	4.87	< 0.0001	No	12.36%			
W/S	30	3.90	0.47	2.80	4.80	< 0.0001	No	12.16%			
B2A3pA	Homogeneity of Variance					Roy's Greatest Root:		F(3,27)	3.43	Pr > F	0.0312
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			1.5600	0.2165	Yes	29	0.01	0.9058	Yes	Neither	Neither
Glove vs 86\"Non			0.4400	0.5110	Yes	29	6.69	0.0150	No	Means	Glove 6.49%
Glove vs W/S			0.3500	0.5591	Yes	29	6.95	0.0133	No	Means	Glove 6.38%
Wand vs 86\"Non			0.1800	0.6727	Yes	29	9.58	0.0043	No	Means	Wand 6.78%
Wand vs W/S			0.2500	0.6175	Yes	29	8.79	0.0060	No	Means	Wand 6.67%
86\"Non vs W/S			0.0000	0.9457	Yes	29	0.05	0.8324	Yes	Neither	Neither





**Figure F- 24: B2A3pAvgOvr All Users Overall Impressions Ratings**

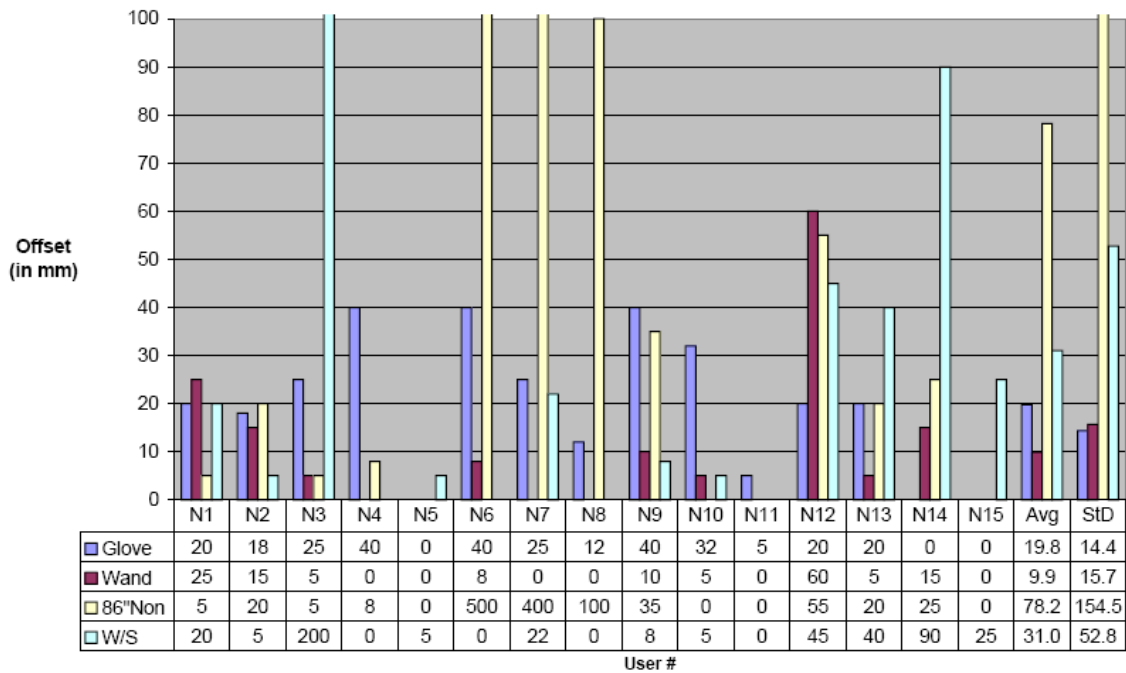
**Table F- 24: B2A3pAvgOvr All Users Overall Impressions Ratings**

Benchmark 2, All Users, 3 Pass Avg., Overall Impressions Ratings											
B2A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.13	0.39	3.30	4.92	> 0.1000	Yes	9.34%			
Wand	30	4.10	0.50	3.02	4.97	> 0.1000	Yes	12.22%			
86*Non	30	3.87	0.46	2.85	4.85	0.0005	No	11.98%			
W/S	30	3.85	0.47	2.80	4.81	0.0001	No	12.14%			
B2A3pA		Homogeneity of Variance				Roy's Greatest Root:			F(3,27)		
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand		1.9234	0.1655			Yes	29	0.16	0.6886	Yes	Neither
Glove vs 86*Non				0.7800	0.3822	Yes	29	8.33	0.0073	No	Means
Glove vs W/S				0.8500	0.3592	Yes	29	10.32	0.0032	No	Means
Wand vs 86*Non				0.1700	0.6833	Yes	29	10.49	0.0030	No	Means
Wand vs W/S				0.1300	0.7156	Yes	29	11.58	0.0020	No	Means
86*Non vs W/S				0.0100	0.9633	Yes	29	1.01	0.3241	Yes	Neither

## **Appendix G**

### **Novice User Benchmark 3 (Spatial Awareness) Detail**

**Benchmark 3 - Novice Users - Pass 1 - Part 1 Offsets**

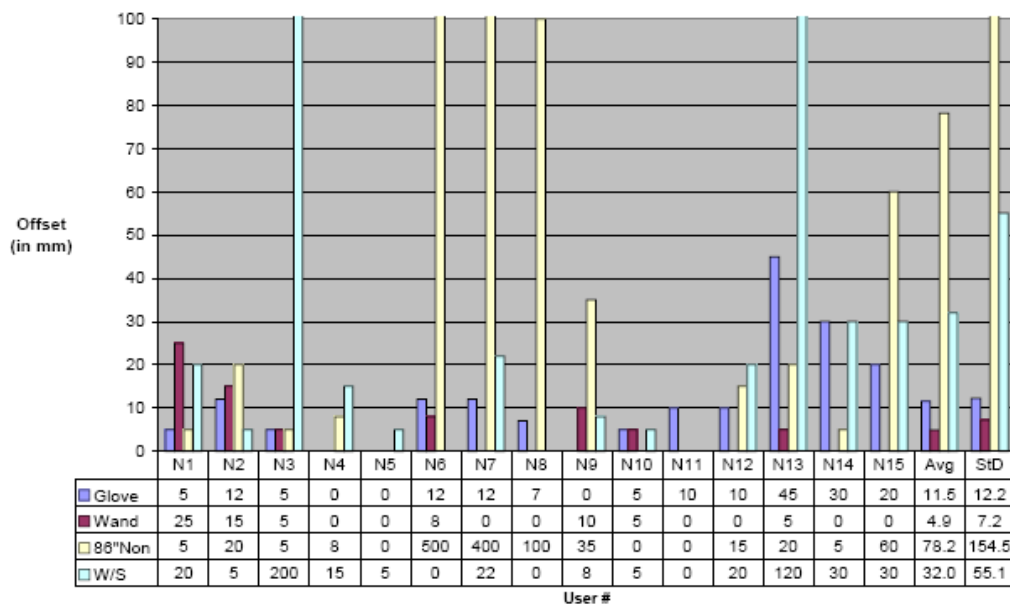


**Figure G- 1: B3Np1-1Off Novice Users Pass 1-Icon 1 Offsets**

**Table G- 1: B3Np1-1Off Novice Users Pass 1-Icon 1 Offsets Statistics**

Benchmark 2, Novice Users, Pass 1, Part 1 Offsets												
B3N1p1	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	CV
Glove	15	19.80		14.36		0.00		40.00		> 0.1000	Yes	72.53%
Wand	15	9.87		15.71		0.00		60.00		0.0057	No	159.22%
86"Non	15	78.20		154.20		0.00		500.00		< 0.1000	No	197.19%
W/S	15	31.00		52.50		0.00		200.00		0.0027	No	169.35%
B3N1p1	Homogeneity of Variance					Roy's Greatest Root: F(3,12)			2.04		Pr > F 0.1616	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.0500	0.8262	Yes	14	3.42	0.0855	No	Means	Wand	50.17%
Glove vs 86"Non			2.9100	0.0992	Yes	14	2.29	0.1528	Yes	Neither	Neither	
Glove vs W/S			1.6600	0.2084	Yes	14	0.60	0.4510	Yes	Neither	Neither	
Wand vs 86"Non			2.9000	0.0998	Yes	14	2.85	0.1132	Yes	Neither	Neither	
Wand vs W/S			1.5900	0.2171	Yes	14	2.36	0.1467	Yes	Neither	Neither	
86"Non vs W/S			2.2600	0.1438	Yes	14	1.13	0.3058	Yes	Neither	Neither	

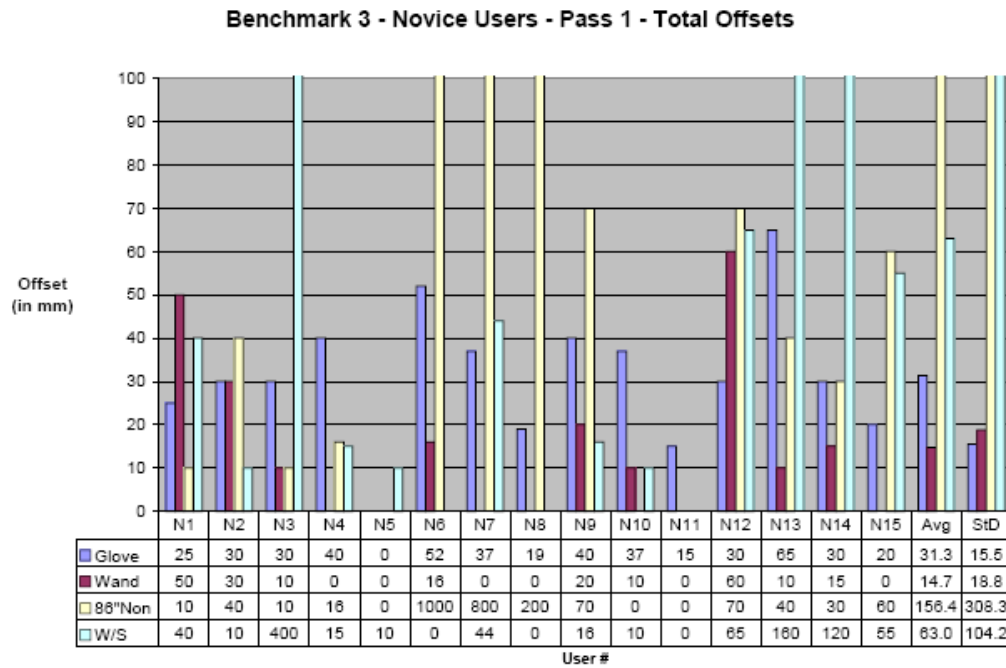
**Benchmark 3 - Novice Users - Pass 1 - Part 2 Offsets**



**Figure G- 2: B3Np1-2Off Novice Users Pass 1 –Icon 2 Offsets**

**Table G- 2: B3Np1-2Off Novice Users Pass2 –Icon 2 Offsets Statistics**

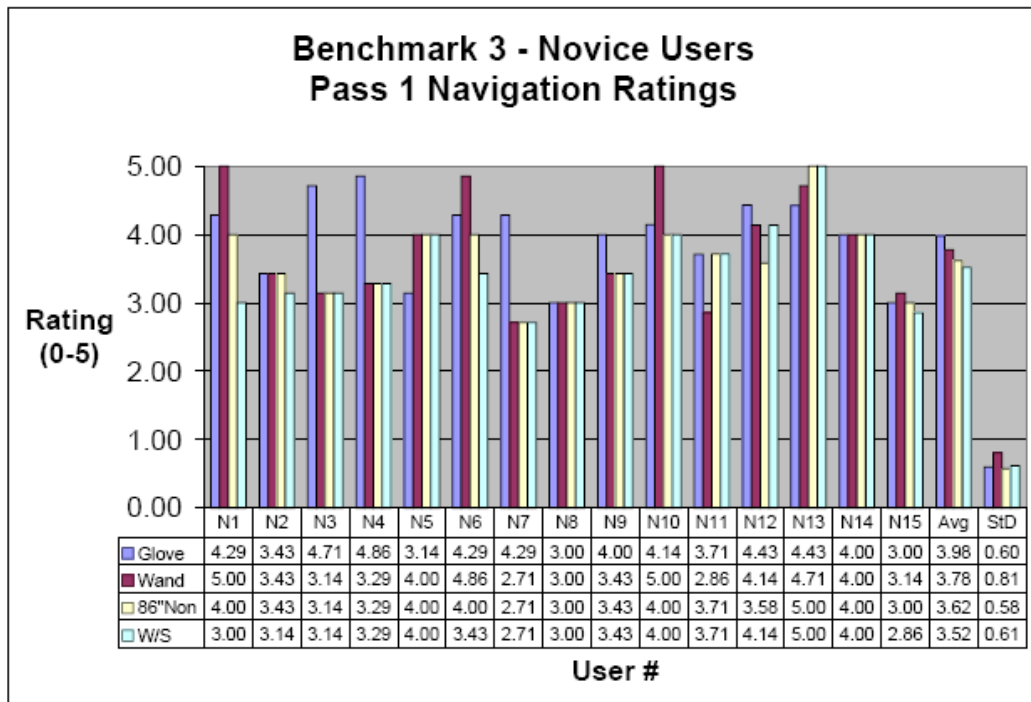
Benchmark 2, Novice Users, Pass 1, Part 2 Offsets												
B3N2p1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	11.53	12.21	0.00	45.00	0.0019	No	105.90%				
Wand	15	4.87	7.24	0.00	25.00	0.0021	No	148.74%				
86"Non	15	78.20	154.50	0.00	500.00	< 0.0001	No	197.57%				
W/S	15	32.00	55.09	0.00	200.00	< 0.0001	No	172.16%				
B3N2p1	Homogeneity of Variance					Roy's Greatest Root: F(3,12)			2.04	Pr > F	0.1616	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			1.3200	0.2597	Yes	14	3.42	0.0855	No	Means	Wand	57.79%
Glove vs 86"Non			2.9300	0.0982	Yes	14	2.29	0.1528	Yes	Neither	Neither	
Glove vs W/S			1.7300	0.1990	Yes	14	0.60	0.4510	Yes	Neither	Neither	
Wand vs 86"Non			2.9500	0.0969	Yes	14	2.85	0.1132	Yes	Neither	Neither	
Wand vs W/S			1.8600	0.1832	Yes	14	2.36	0.1467	Yes	Neither	Neither	
86"Non vs W/S			2.2700	0.1435	Yes	14	1.13	0.3058	Yes	Neither	Neither	



**Figure G- 3: B3Np1TotOff Novice Users Pass 1 Total Offsets**

**Table G- 3: B3Np1TotOff Novice Users Pass 1 Total Offsets Statistics**

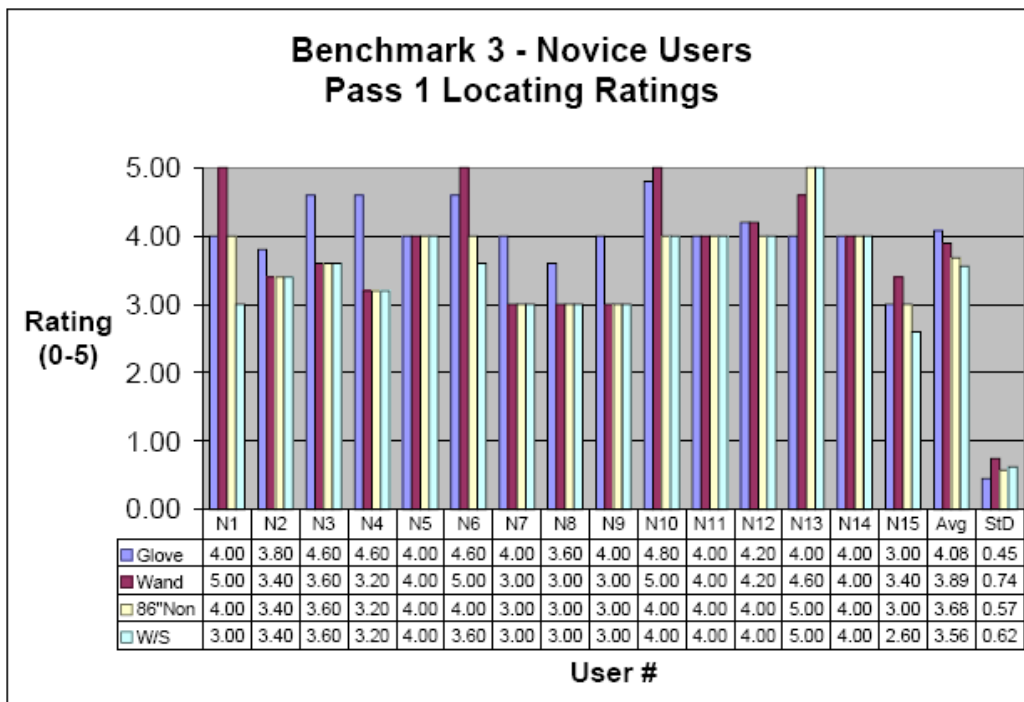
Benchmark 2, Novice Users, Pass 1, Total Offsets											
B3Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	31.33	15.47	0.00	65.00	> 0.1000	Yes	49.38%			
Wand	15	14.73	18.76	0.00	60.00	0.0587	Yes	127.36%			
86"Non	15	156.40	308.30	0.00	1000.00	< 0.0001	No	197.12%			
W/S	15	63.00	104.20	0.00	400.00	0.0012	No	165.40%			
B3Np1	Homogeneity of Variance				Roy's Greatest Root: F(3,12)				0.84	Pr > F 0.5404	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.4977	0.4805			Yes	14	1.11	0.3324	Yes	Neither	Neither
Glove vs 86"Non			2.9200	0.0987	Yes	14	2.24	0.1851	Yes	Neither	Neither
Glove vs W/S			1.7900	0.1919	Yes	14	0.34	0.5806	Yes	Neither	Neither
Wand vs 86"Non			2.9100	0.0991	Yes	14	0.08	0.7879	Yes	Neither	Neither
Wand vs W/S			1.7500	0.1966	Yes	14	0.98	0.3598	Yes	Neither	Neither
86"Non vs W/S			2.2500	0.1444	Yes	14	1.78	0.2306	Yes	Neither	Neither



**Figure G- 4: B3Np1Nav Novice Users Pass 1 Navigation Ratings**

**Table G- 4: B3Np1Nav Novice Users Pass 1 Navigation Ratings Statistics**

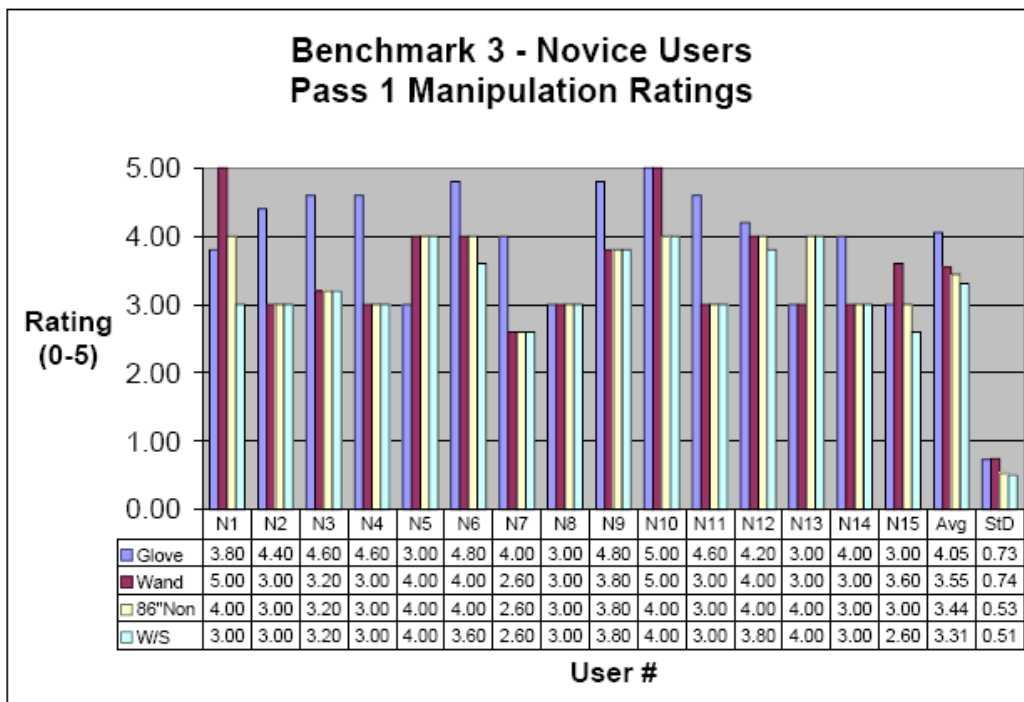
Benchmark 3, Novice Users, Pass 2, Navigation Ratings											
B3Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	3.98	0.5987	3.00	4.86	> 0.1000	Yes	15.04%			
Wand	15	3.78	0.8078	2.71	5.00	> 0.1000	Yes	21.36%			
86"Non	15	3.62	0.5756	2.71	5.00	> 0.1000	Yes	15.90%			
W/S	15	3.52	0.6146	2.71	5.00	> 0.1000	Yes	17.44%			
B3Np1	Homogeneity of Variance				Roy's Greatest Root: F(3,12)				1.57	Pr > F 0.2207	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	1.1882	0.2757			Yes	14	0.00	0.9912	Yes	Neither	Neither
Glove vs 86"Non	0.0213	0.8838			Yes	14	3.45	0.0734	No	Means	Glove 9.05%
Glove vs W/S	0.0088	0.9252			Yes	14	1.03	0.3176	Yes	Neither	Neither
Wand vs 86"Non	1.5193	0.2127			Yes	14	3.20	0.0842	No	Means	Wand 4.26%
Wand vs W/S	0.9960	0.3183			Yes	14	1.33	0.2579	Yes	Neither	Neither
86"Non vs W/S	0.0576	0.8104			Yes	14	1.49	0.2325	Yes	Neither	Neither



**Figure G- 5: B3Np1Loc Novice Users Pass 1 Locating Ratings**

**Table G- 5: B3Np1Loc Novice Users Pass 1 Locating Ratings Statistics**

Benchmark 3, Novice Users, Pass 2, Locating Ratings												
B3Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.08	0.4523	3.00	4.80	0.0232	No	11.09%				
Wand	15	3.89	0.7440	3.00	5.00	> 0.1000	Yes	19.11%				
86"Non	15	3.68	0.5747	3.00	5.00	0.0162	No	15.62%				
W/S	15	3.56	0.6197	2.60	5.00	> 0.1000	Yes	17.41%				
B3Np1	Homogeneity of Variance					Roy's Greatest Root: F(3,12)		1.57	Pr > F		0.2207	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand	1.1882	0.2757			Yes	14	0.00	0.9912	Yes	Neither	Neither	
Glove vs 86"Non	0.0213	0.8838			Yes	14	3.45	0.0734	No	Means	Glove 9.80%	
Glove vs W/S	0.0088	0.9252			Yes	14	1.03	0.3176	Yes	Neither	Neither	
Wand vs 86"Non	1.5193	0.2127			Yes	14	3.20	0.0842	No	Means	Wand 5.47%	
Wand vs W/S	0.9960	0.3183			Yes	14	1.33	0.2579	Yes	Neither	Neither	
86"Non vs W/S	0.0576	0.8104			Yes	14	1.49	0.2325	Yes	Neither	Neither	

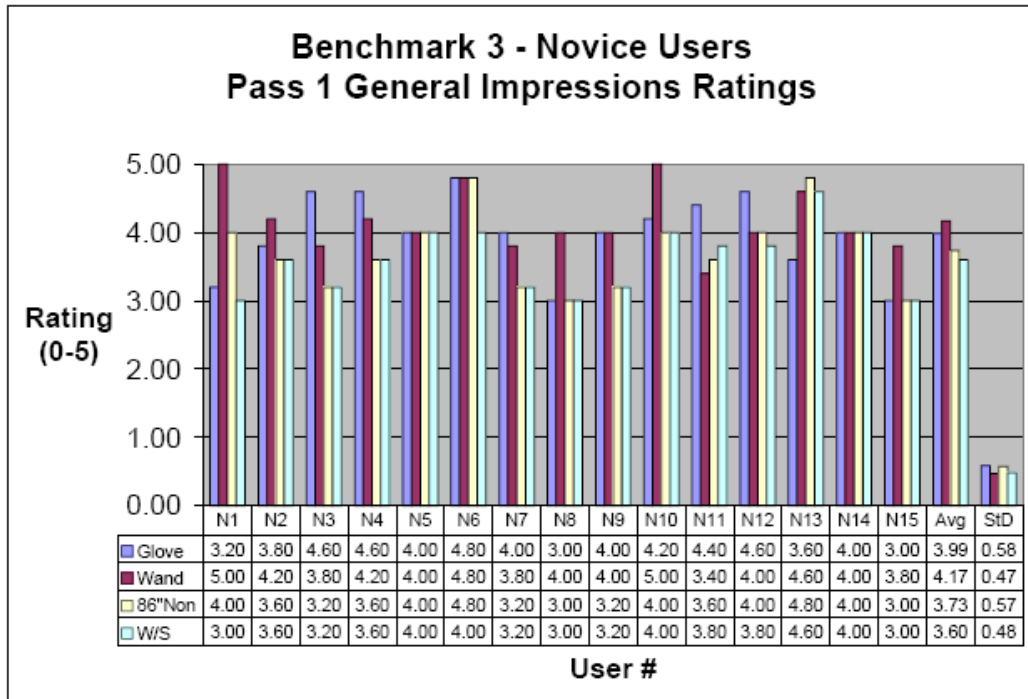


**Figure G- 6: B3Np1Mov Novice Users Manipulation Ratings**

**Table G- 6: B3Np1Mov Novice Users Manipulation Ratings Statistics**

Benchmark 3, Novice Users, Pass 2, Manipulation Ratings											
B3Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.05	0.7347	3.00	5.00	> 0.1000	Yes	18.13%			
Wand	15	3.55	0.7425	2.60	5.00	0.0243	No	20.93%			
86"Non	15	3.44	0.5302	2.60	4.00	0.0062	No	15.41%			
W/S	15	3.04	0.5063	2.60	4.00	0.0071	No	16.67%			
B3Np1	Homogeneity of Variance				Roy's Greatest Root: F(3,12) 5.02 Pr > F 0.0175						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.0000	0.9599	Yes	14	4.27	0.9912	Yes	Neither	Neither
Glove vs 86"Non			3.9100	0.0578	Yes	14	7.13	0.0734	No	Means	Glove 15.12%
Glove vs W/S			4.4000	0.0451	No	14	11.66	0.3176	Yes	Var	Glove 25.07%
Wand vs 86"Non			1.9800	0.1703	Yes	14	0.75	0.0842	No	Means	Wand 3.02%
Wand vs W/S			2.2900	0.1415	Yes	14	1.91	0.2579	Yes	Neither	Neither
86"Non vs W/S			0.1400	0.7130	Yes	14	3.41	0.2325	Yes	Neither	Neither

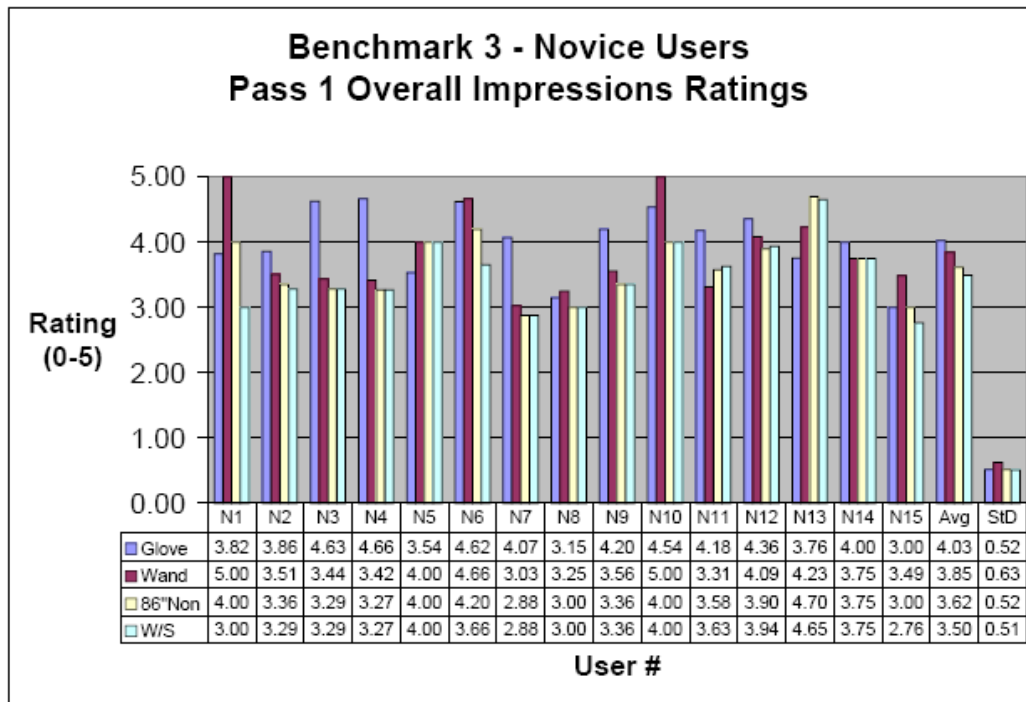




**Figure G- 7: B3Np1Gen Novice Users Pass 1 General Impressions Ratings**

**Table G- 7: B3Np1Gen Novice Users Pass 1 General Impressions Ratings Statistics**

Benchmark 3, Novice Users, Pass 2, General Impressions Ratings											
B3Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	3.99	0.5829	3.00	4.80	> 0.1000	Yes	14.62%			
Wand	15	4.17	0.4713	3.40	5.00	0.0170	No	11.29%			
86"Non	15	3.73	0.5740	3.00	4.80	> 0.1000	Yes	15.38%			
W/S	15	3.60	0.4540	3.00	4.60	> 0.1000	Yes	12.61%			
B3Np1	Homogeneity of Variance					Roy's Greatest Root: F(3,12)		5.74	Pr > F 0.0113		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.6066	0.4361			Yes	14	0.88	0.3653	Yes	Neither	Neither
Glove vs 86"Non	0.0032	0.9549			Yes	14	2.09	0.1699	Yes	Neither	Neither
Glove vs W/S	0.4646	0.4955			Yes	14	6.60	0.0223	No	Means	Glove 9.71%
Wand vs 86"Non			0.7200	0.4020	Yes	14	14.04	0.0022	No	Means	Wand 10.54%
Wand vs W/S			0.0200	0.9017	Yes	14	14.91	0.0017	No	Means	Wand 13.73%
86"Non vs W/S	0.3913	0.5316			Yes	14	2.50	0.1362	Yes	Neither	Neither

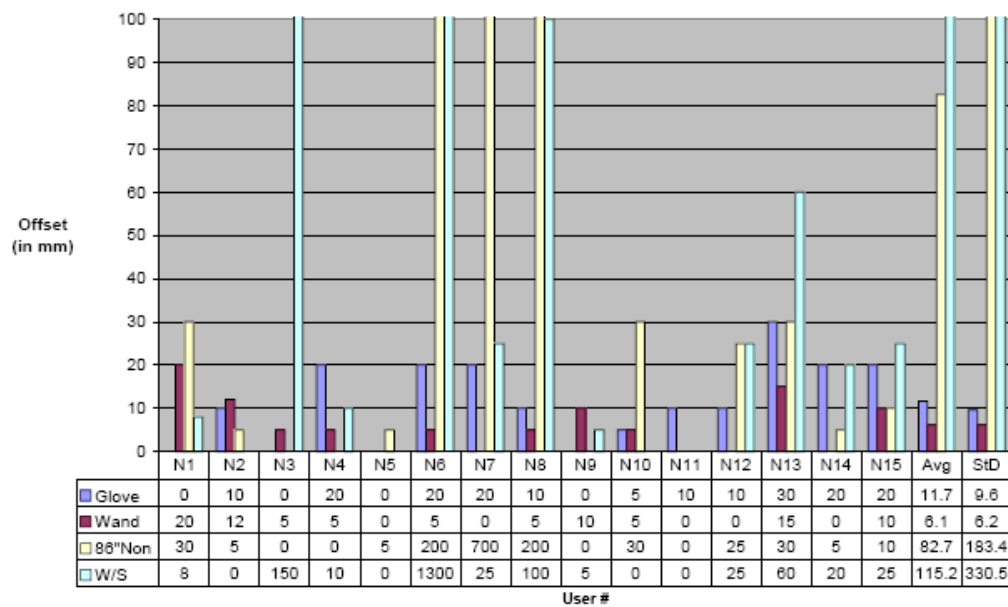


**Figure G- 8: B3Np1Ovr Novice Users Pass 1 Overall Impressions Ratings**

**Table G- 8: B3Np1Ovr Novice Users Pass 1 Overall Impressions Ratings Statistics**

Benchmark 3, Novice Users, Pass 2, Overall Impressions Ratings										
B3Np1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	15	4.03	0.5178	3.00	4.66	> 0.1000	Yes	12.86%		
Wand	15	3.85	0.6297	3.03	5.00	0.0696	Yes	16.36%		
86\"Non	15	3.62	0.5155	2.88	4.70	> 0.1000	Yes	14.25%		
W/S	15	3.50	0.5146	2.76	4.65	> 0.1000	Yes	14.71%		
B3Np1	Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	3.46	Pr > F	0.0511
	Bartlett's Test		Levene's Test		df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand	0.5130	0.4738			Yes	14	0.90	0.3592	Yes	Neither
Glove vs 86\"Non	0.0003	0.9870			Yes	14	5.39	0.0293	No	Means Glove 10.11%
Glove vs W/S	0.0003	0.9862			Yes	14	10.53	0.0059	No	Means Glove 13.09%
Wand vs 86\"Non	0.5364	0.4679			Yes	14	5.12	0.0401	No	Means Wand 6.00%
Wand vs W/S	0.5379	0.4633			Yes	14	4.97	0.0427	No	Means Wand 9.12%
86\"Non vs W/S	0.0000	0.9992			Yes	14	2.67	0.1243	Yes	Neither

**Benchmark 3 - Novice Users - Pass 2 - Part 1 Offsets**

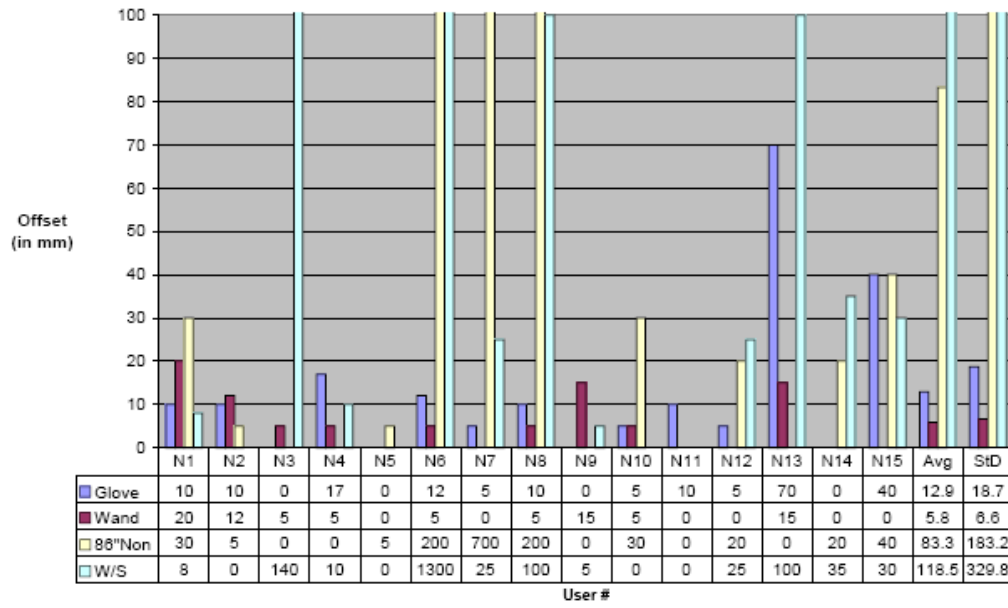


**Figure G- 9: B3Np2-1Off Novice Users Pass 2-Icon 1 Offsets**

**Table G- 9: B3Np2-1Off Novice Users Pass 2-Icon 1 Offsets Statistics**

Benchmark 2, Novice Users, Pass 2, Part 1 Offsets												
B3N1p2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	11.67	9.57	0.00	30.00	0.0802	Yes	22.53%				
Wand	15	6.13	6.15	0.00	20.00	0.0203	No	24.73%				
86\"Non	15	82.67	183.40	0.00	700.00	< 0.0001	No	29.82%				
W/S	15	115.20	330.50	0.00	1300.00	< 0.1000	No	27.39%				
B3N1p2		Homogeneity of Variance				Roy's Greatest Root: F(3,12)			1.40	Pr > F		
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				3.8800	0.0588	Yes	14	3.44	0.0849	No	Means	Wand 47.45%
Glove vs 86\"Non				1.5700	0.2206	Yes	14	2.31	0.1507	Yes	Neither	Neither
Glove vs W/S				1.2000	0.2826	Yes	14	1.49	0.2422	Yes	Neither	Neither
Wand vs 86\"Non				1.5700	0.2199	Yes	14	2.56	0.1317	Yes	Neither	Neither
Wand vs W/S				1.2000	0.2824	Yes	14	1.63	0.2225	Yes	Neither	Neither
86\"Non vs W/S				0.5400	0.4697	Yes	14	0.13	0.7219	Yes	Neither	Neither

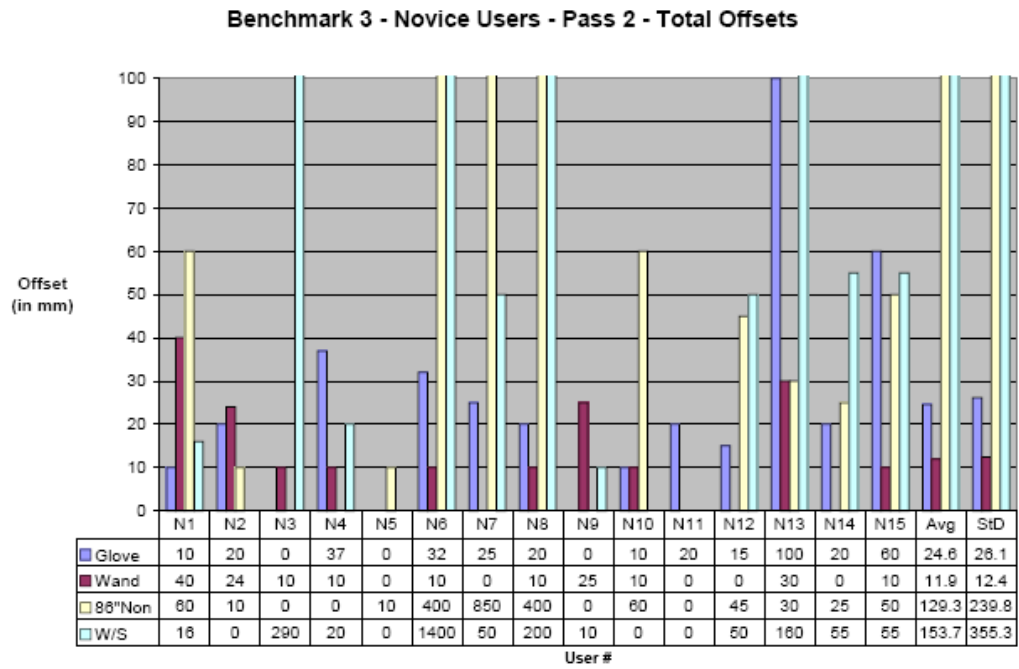
**Benchmark 3 - Novice Users - Pass 2 - Part 2 Offsets**



**Figure G- 10: B3Np2-2Off Novice Users Pass 2-Icon 2 Offsets**

**Table G- 10: B3Np2-2Off Novice Users Pass 2-Icon 2 Offsets Statistics**

Benchmark 2, Novice Users, Pass 2, Part 2 Offsets												
B3N2p2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	12.93	18.71	0.00	70.00	0.0002	No	22.53%				
Wand	15	5.80	6.63	0.00	20.00	0.0023	No	24.73%				
86"Non	15	83.33	183.20	0.00	700.00	< 0.0001	No	29.82%				
W/S	15	118.50	329.80	0.00	1300.00	< 0.0001	No	27.39%				
B3N2p2	Homogeneity of Variance					Roy's Greatest Root: F(3,12)			1.40	Pr > F 0.2899		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			1.7700	0.0588	Yes	14	3.44	0.0849	No	Means	Wand 55.14%	
Glove vs 86"Non			1.5500	0.224	Yes	14	2.31	0.1507	Yes	Neither	Neither	
Glove vs W/S			1.2000	0.2832	Yes	14	1.49	0.2422	Yes	Neither	Neither	
Wand vs 86"Non			1.5700	0.2199	Yes	14	2.56	0.1317	Yes	Neither	Neither	
Wand vs W/S			1.2000	0.2819	Yes	14	1.63	0.2225	Yes	Neither	Neither	
86"Non vs W/S			0.5400	0.4699	Yes	14	0.13	0.7219	Yes	Neither	Neither	



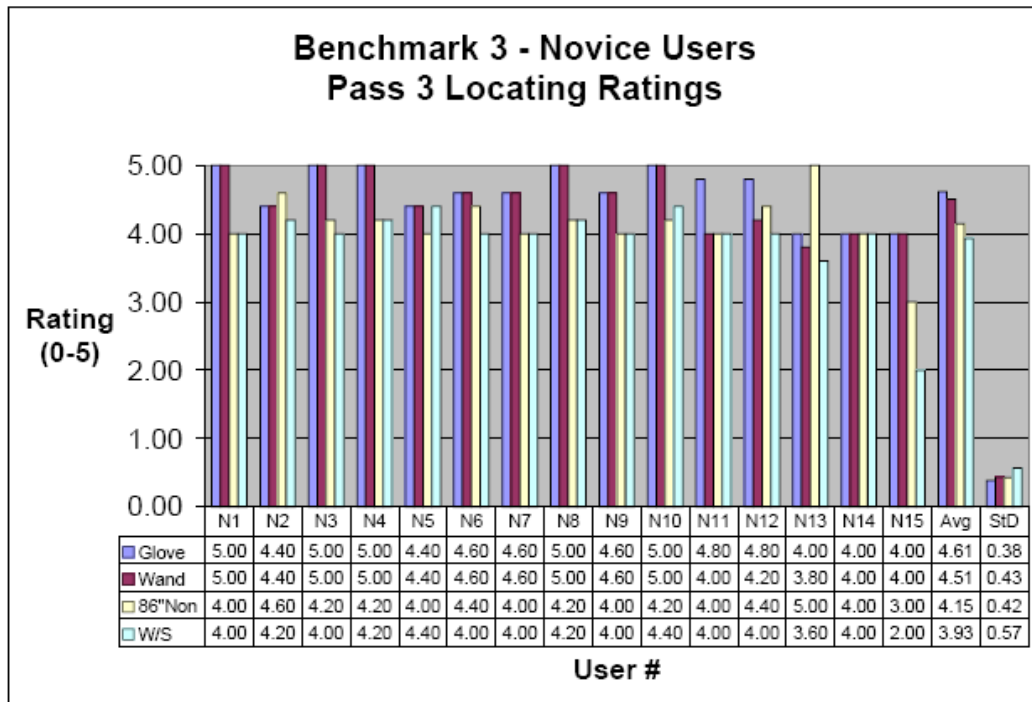
**Figure G- 11: B3Np2TotOff Novice Users Pass 2 Total Offsets**

**Table G- 11: B3Np2TotOff Novice Users Pass 2 Total Offsets Statistics**

Benchmark 2, Novice Users, Pass 2, Total Offsets											
B3Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	24.60	26.14	0.00	100.00	0.0236	No	22.53%			
Wand	15	11.93	12.44	0.00	40.00	0.0010	No	24.73%			
86\"Non	15	129.30	239.80	0.00	850.00	< 0.0001	No	29.82%			
W/S	15	153.70	355.30	0.00	1400.00	< 0.0001	No	27.39%			
B3Np2	Homogeneity of Variance					Roy's Greatest Root:		F(3,12)	0.94	Pr > F	0.5020
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			1.7300	0.1993	Yes	14	0.98	0.3594	Yes	Neither	Neither
Glove vs 86\"Non			2.4700	0.1273	Yes	14	2.07	0.2000	Yes	Neither	Neither
Glove vs W/S			1.3100	0.2629	Yes	14	0.03	0.8657	Yes	Neither	Neither
Wand vs 86\"Non			2.5200	0.1239	Yes	14	0.21	0.6594	Yes	Neither	Neither
Wand vs W/S			1.3200	0.2610	Yes	14	0.94	0.5015	Yes	Neither	Neither
86\"Non vs W/S			0.3500	0.5573	Yes	14	2.73	0.1497	Yes	Neither	Neither



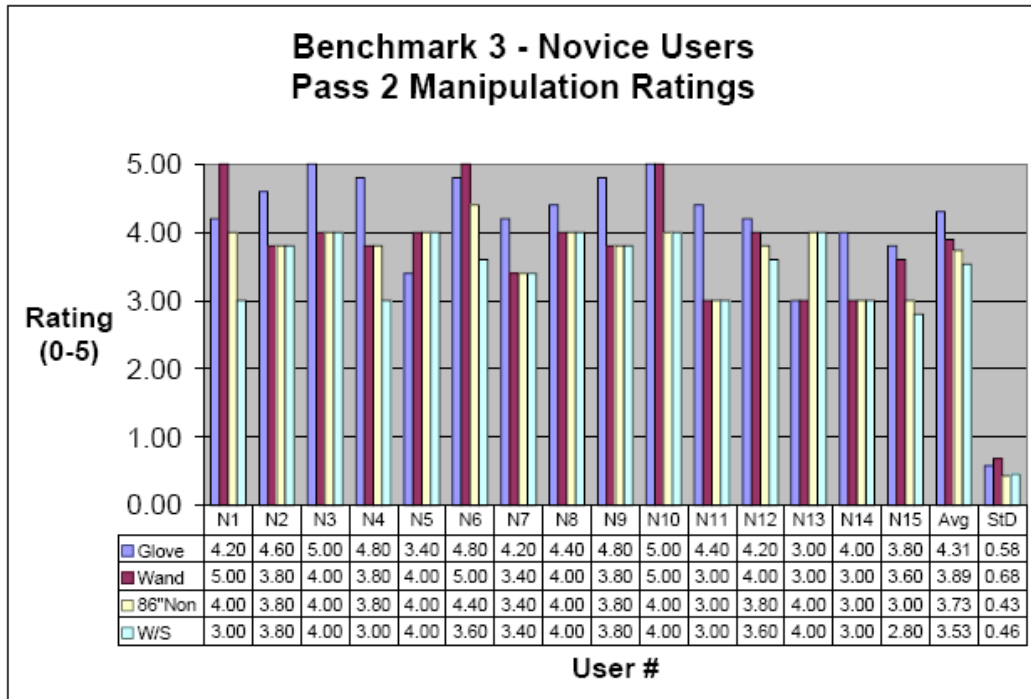
Benchmark 3, Novice Users, Pass 2, Navigation Ratings												
B3Np2	# Users	Mean	ST. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.32	0.5961	3.00	5.00	> 0.1000	Yes	22.53%				
Wand	15	4.10	0.5193	3.29	5.00	0.0062	No	24.73%				
86"Non	15	3.84	0.3851	2.00	4.71	0.0042	No	29.82%				
W/S	15	3.68	0.4752	3.43	4.29	> 0.1000	Yes	27.39%				
B3Np2	Homogeneity of Variance				Roy's Greatest Root: F(3,12)		3.79	Pr > F	0.0217			
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.3200	0.5785	Yes	14	8.60	0.0065	No	Means	Glove	5.30%
Glove vs 86"Non			2.2700	0.1435	Yes	14	1.66	0.2079	Yes	Neither	Neither	
Glove vs W/S	0.6872	0.4071			Yes	14	1.57	0.2204	Yes	Neither	Glove	14.99%
Wand vs 86"Non			1.1000	0.3032	Yes	14	3.24	0.0821	No	Means	Wand	6.28%
Wand vs W/S			0.0900	0.7603	Yes	14	2.38	0.1338	Yes	Neither	Neither	
86"Non vs W/S			0.3700	0.5503	Yes	14	0.26	0.6164	Yes	Neither	Neither	



**Figure G- 13: B3Np2Loc Novice Users Pass 2 Locating Ratings**

**Table G- 13: B3Np2Loc Novice Users Pass 2 Locating Ratings Statistics**

Benchmark 3, Novice Users, Pass 2, Locating Ratings											
B3Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.32	0.5961	3.00	5.00	> 0.1000	Yes	22.53%			
Wand	15	4.10	0.5193	3.29	5.00	0.0062	No	24.73%			
86"Non	15	3.84	0.3851	3.00	4.71	0.0042	No	29.82%			
W/S	15	3.68	0.4752	2.43	4.29	> 0.1000	Yes	27.39%			
B3Np2	Homogeneity of Variance				Roy's Greatest Root: F(3,12)		3.79	Pr > F		0.0217	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.3200	0.5785	Yes	14	8.60	0.0065	No	Means	Glove 5.30%
Glove vs 86"Non			2.2700	0.1435	Yes	14	1.66	0.2079	Yes	Neither	Neither
Glove vs W/S	0.6872	0.4071			Yes	14	1.57	0.2204	Yes	Neither	Neither
Wand vs 86"Non			1.1000	0.3032	Yes	14	3.24	0.0821	No	Means	Wand 6.28%
Wand vs W/S			0.0900	0.7603	Yes	14	2.38	0.1338	Yes	Neither	Neither
86"Non vs W/S			0.3700	0.5503	Yes	14	0.26	0.6164	Yes	Neither	Neither

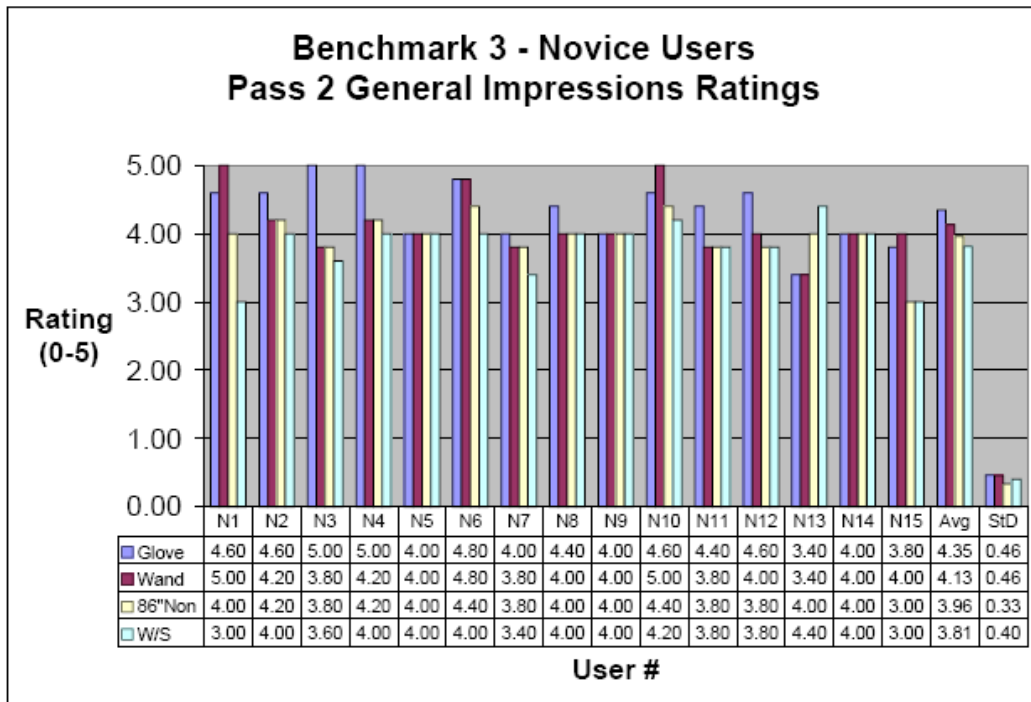


**Figure G- 14: B3Np2Mov Novice Users Pass 2 Manipulation Ratings**

**Table G- 14: B3Np2Mov Novice Users Pass 2 Manipulation Ratings Statistics**

Benchmark 3, Novice Users, Pass 2, Manipulation Ratings												
B3Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.31	0.5800	3.00	5.00	> 0.1000	Yes	22.53%				
Wand	15	3.89	0.6798	3.00	5.00	0.0224	No	24.73%				
86"Non	15	3.73	0.4320	3.00	4.40	0.001	No	29.82%				
W/S	15	3.53	0.4577	2.80	4.00	0.0701	Yes	27.39%				
B3Np2	Homogeneity of Variance					Roy's Greatest Root:		F(3,12)	5.60	Pr > F	0.0123	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.4400	0.5110	Yes	14	6.02	0.0065	No	Means	Glove	9.61%
Glove vs 86"Non			1.1800	0.2875	Yes	14	11.84	0.2079	Yes	Neither	Neither	
Glove vs W/S			0.9400	0.3398	Yes	14	17.28	0.2204	Yes	Neither	Neither	
Wand vs 86"Non			3.2000	0.0846	Yes	14	1.59	0.0821	No	Means	Wand	4.11%
Wand vs W/S			2.9400	0.0977	Yes	14	3.65	0.1338	Yes	Neither	Neither	
86"Non vs W/S			0.1000	0.7521	Yes	14	4.77	0.6164	Yes	Neither	Neither	

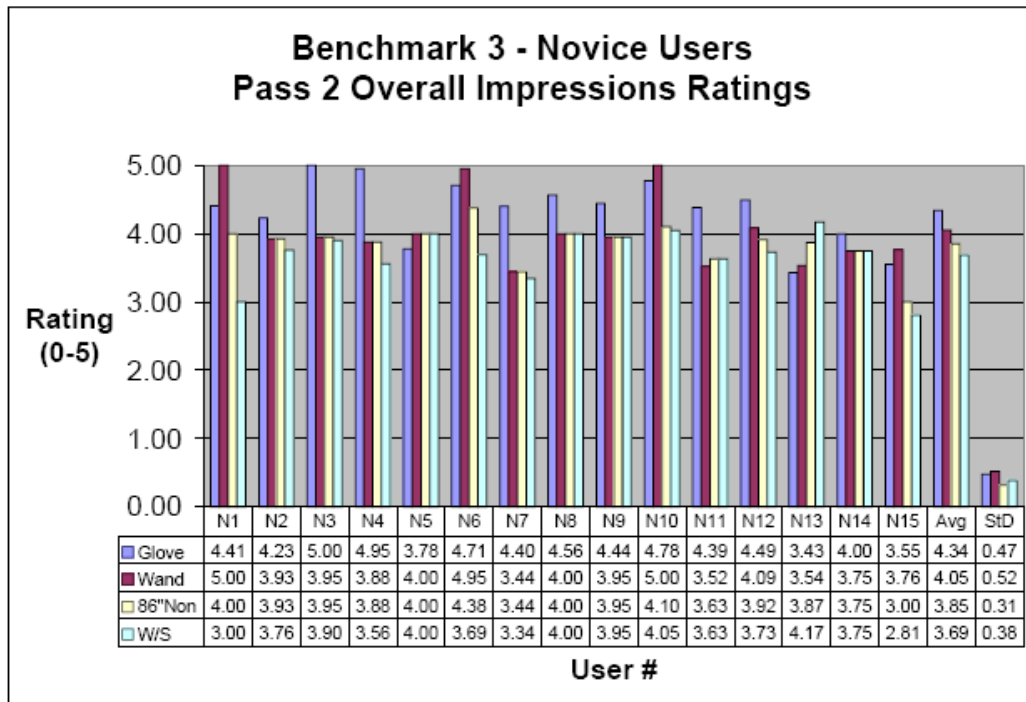




**Figure G- 15: B3Np2Gen Novice User Pass 2 General Impressions Ratings**

**Table G- 15: B3Np2Gen Novice User Pass 2 General Impressions Ratings**

Benchmark 3, Novice Users, Pass 2, General Impressions Ratings												
B3Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.35	0.4627	3.40	5.00	> 0.1000	Yes	22.53%				
Wand	15	4.13	0.4577	3.40	5.00	0.0023	No	24.73%				
86"Non	15	3.96	0.3312	3.00	4.40	0.0137	No	29.82%				
W/S	15	3.81	0.4033	3.00	4.40	0.0027	No	27.39%				
B3Np2		Homogeneity of Variance				Roy's Greatest Root: F(3,12)		3.37	Pr > F		0.0546	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				3.3700	0.0878	Yes	14	0.00	0.9641	Yes	Neither	Neither
Glove vs 86"Non				11.3900	0.0045	No	14	1.29	0.2265	Yes	Var	Glove 8.90%
Glove vs W/S				10.7200	0.0055	No	14	0.33	0.5726	Yes	Var	Glove 12.28%
Wand vs 86"Non				2.5600	0.1323	Yes	14	0.98	0.3303	Yes	Neither	Neither
Wand vs W/S				3.5500	0.0806	Yes	14	0.22	0.6394	Yes	Neither	Neither
86"Non vs W/S				3.4300	0.0853	Yes	14	0.35	0.5616	Yes	Neither	86"Non 3.71%

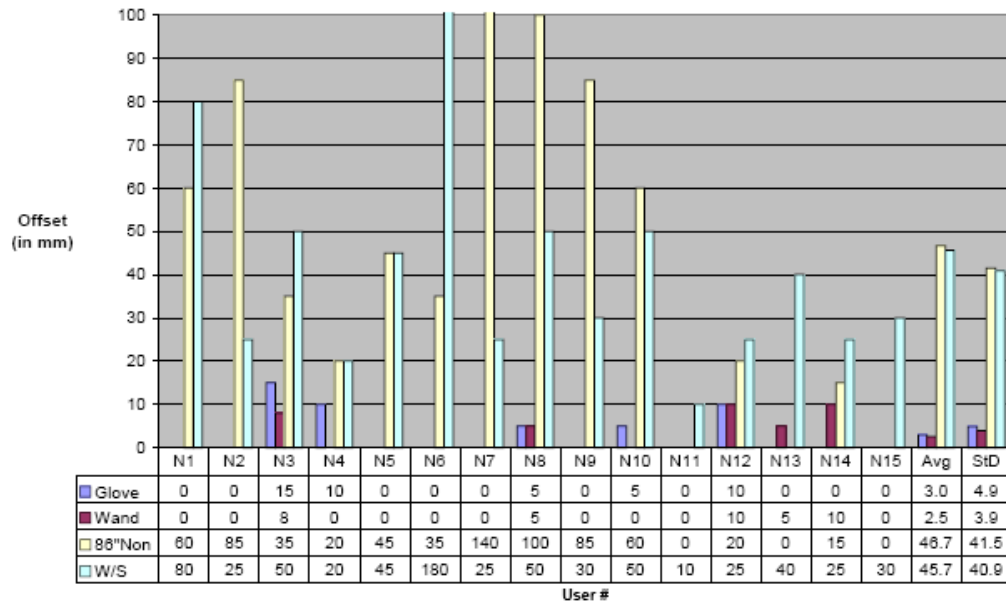


**Figure G- 16: B3Np2Ovr Novice Users Pass 2 Overall Impressions Ratings**

**Table G- 16: B3Np2Ovr Novice Users Pass 2 Overall Impressions Ratings Statistics**

Benchmark 3, Novice Users, Pass 2, Overall Impressions Ratings										
B3Np2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	15	4.34	0.4726	3.43	5.00	0.0940	Yes	22.53%		
Wand	15	4.05	0.5188	3.44	5.00	0.0036	No	24.73%		
86\"Non	15	3.85	0.3149	3.00	4.38	0.0087	No	29.82%		
W/S	15	3.69	0.3847	2.81	4.17	> 0.1000	Yes	27.39%		
B3Np2		Homogeneity of Variance				Roy's Greatest Root:			6.23	
		Bartlett's Test		Levene's Test		df	F(1,14)	Pr > F	Equal Means ?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					
Glove vs Wand				0.1500	0.6984	Yes	14	4.33	0.0563	No
Glove vs 86\"Non				1.9400	0.1748	Yes	14	20.03	0.0005	No
Glove vs W/S		0.5714	0.4497			Yes	14	19.45	0.0006	No
Wand vs 86\"Non				2.4600	0.1280	Yes	14	3.58	0.0792	No
Wand vs W/S				1.1900	0.2850	Yes	14	4.48	0.0527	No
86\"Non vs W/S				0.3700	0.5504	Yes	14	4.10	0.0625	No
										Statistically Better
										Glove 6.68%
										Glove 11.24%
										Glove 15.02%
										Wand 4.89%
										Wand 8.94%
										86\"Non 4.26%

**Benchmark 3 - Novice Users - Pass 3 - Part 1 Offsets**

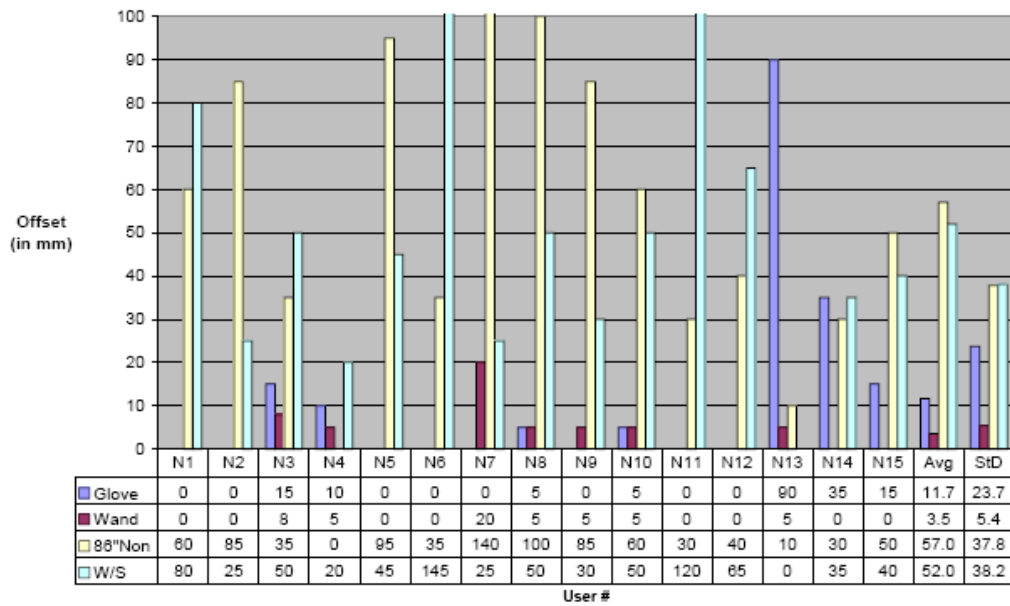


**Figure G- 17: B3Np3-1Off Novice Users Pass 3-Icon 1 Offsets**

**Table G- 17: B3Np3-1Off Novice Users Pass 3-Icon 1 Offsets Statistics**

Benchmark 2, Novice Users, Pass 3, Part 1 Offsets										
B3N1p3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	15	3.00	4.93	0.00	45.00	< 0.0001	No	22.53%		
Wand	15	2.53	3.94	0.00	10.00	< 0.0001	No	24.73%		
86\"Non	15	46.67	41.48	0.00	140.00	> 0.1000	Yes	29.82%		
W/S	15	45.67	40.92	0.00	180.00	0.0002	No	27.39%		
B3N1p3	Homogeneity of Variance				Roy's Greatest Root: F(3,12)		8.60	Pr > F	0.0026	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						Statistically Better
Glove vs Wand			0.6000	0.4437	Yes	14	0.15	0.7001	Yes	Neither
Glove vs 86\"Non			7.9700	0.0086	No	14	15.91	0.0013	No	Var/Mean
Glove vs W/S			1.7000	0.2031	Yes	14	15.61	0.0014	No	Means
Wand vs 86\"Non			8.0600	0.0083	No	14	15.99	0.0013	No	Var/Mean
Wand vs W/S			1.7200	0.2008	Yes	14	15.99	0.0013	No	Means
86\"Non vs W/S			0.0000	0.9742	Yes	14	0.00	0.9475	Yes	Neither

**Benchmark 3 - Novice Users - Pass 3 - Part 2 Offsets**

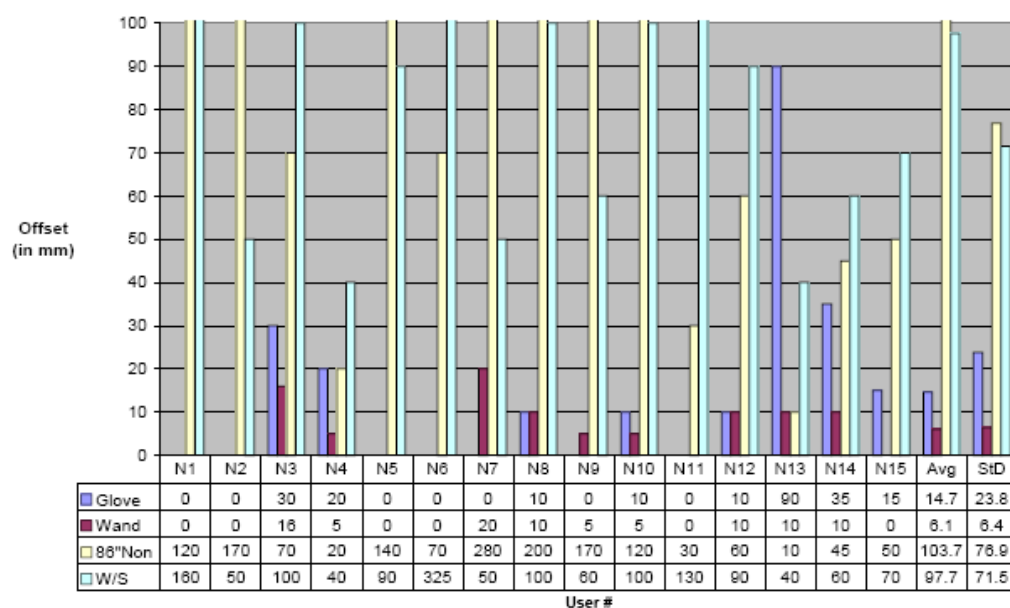


**Figure G- 18: B3Np3-2Off Novice Users Pass 3-Icon 2 Offsets**

**Table G- 18: B3Np3-2Off Novice Users Pass 3-Icon 2 Offsets Statistics**

Benchmark 2, Novice Users, Pass 3, Part 2 Offsets											
B3N2p3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	11.67	23.87	0.00	90.00	0.0004	No	22.53%			
Wand	15	3.53	5.36	0.00	20.00	0.0027	No	24.73%			
86\"Non	15	57.00	37.79	0.00	140.00	> 0.1000	Yes	29.82%			
W/S	15	52.00	38.16	0.00	145.00	0.01	No	27.39%			
B3N2p3		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	8.60	Pr > F	0.0026
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				1.5400	0.2256	Yes	14	0.15	0.7001	Yes	Neither
Glove vs 86\"Non				1.7400	0.1977	Yes	14	15.91	0.0013	No	Means
Glove vs W/S				1.2900	0.2650	Yes	14	15.61	0.0014	No	Means
Wand vs 86\"Non				8.0200	0.0085	No	14	15.99	0.0013	No	Var/Mean
Wand vs W/S				4.7200	0.0384	No	14	15.99	0.0013	No	Var/Mean
86\"Non vs W/S				0.0000	0.9725	Yes	14	0.00	0.9475	Yes	Neither

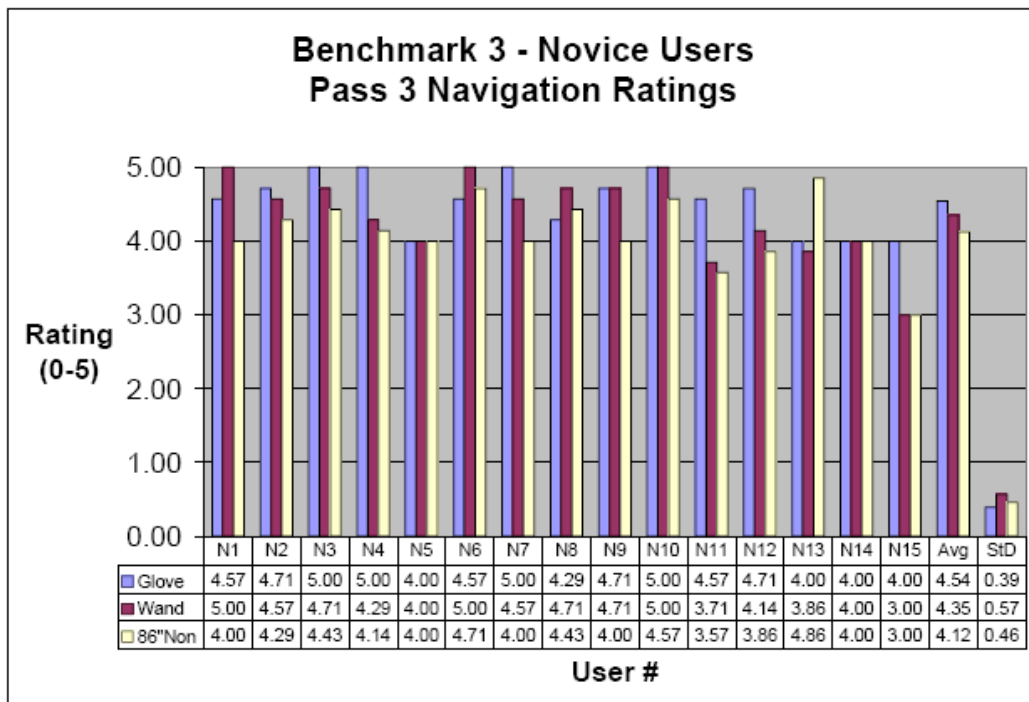
**Benchmark 3 - Novice Users - Pass 3 - Total Offsets**



**Figure G- 19: B3Np3TotOff Novice Users Pass 3 Total Offsets**

**Table G- 19: B3Np3TotOff Novice Users Pass 3 Total Offsets Statistics**

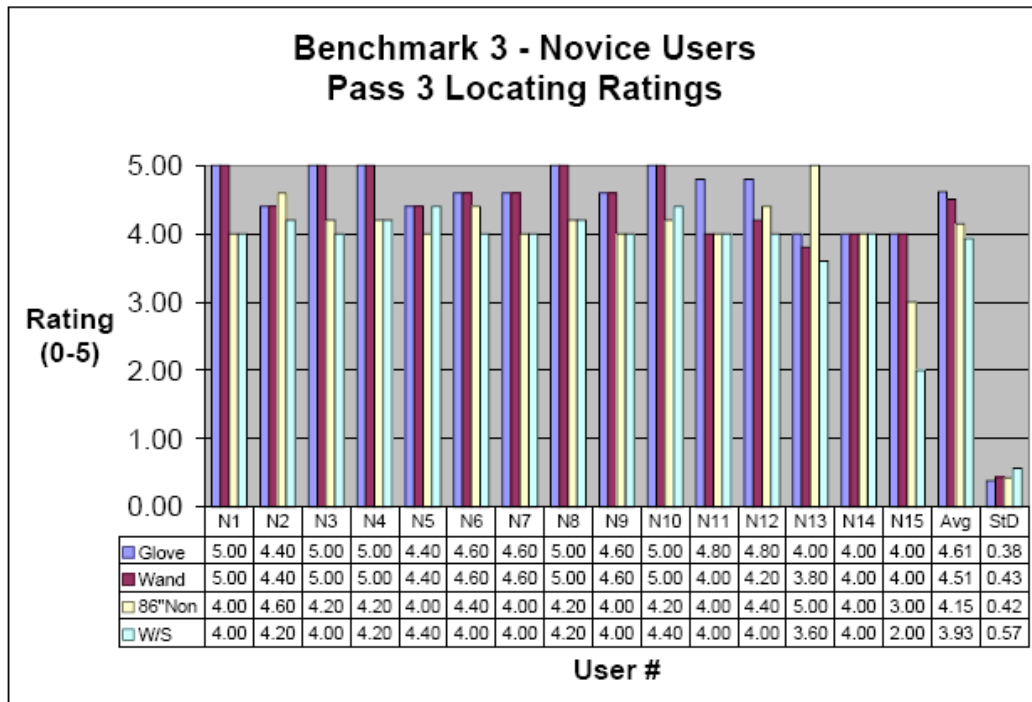
Benchmark 2, Novice Users, Pass 3, Total Offsets													
B3Np3	# Users	Mean	St. Dev.		Low	High		P Value	Normal?	CV			
Glove	15	14.67	23.79		0.00	90.00		0.0046	No	22.53%			
Wand	15	6.07	6.43		0.00	20.00		0.0359	No	24.73%			
86"Non	15	103.70	76.87		10.00	280.00		0.0985	Yes	29.82%			
W/S	15	97.67	71.49		40.00	325.00		0.0017	No	27.39%			
B3Np3		Homogeneity of Variance				Roy's Greatest Root: F(3,12)				34.02	Pr > F	0.0026	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand				2.5900	0.1185	Yes	14	10.52	0.0179	No	Means	Wand	58.64%
Glove vs 86"Non				5.8900	0.0219	No	14	19.70	0.0044	No	Var/Means	Glove	85.85%
Glove vs W/S				1.4900	0.2329	Yes	14	2.53	0.1631	Yes	Neither	Neither	
Wand vs 86"Non				7.6900	0.0098	No	14	0.41	0.5463	Yes	Var	Wand	94.15%
Wand vs W/S				1.9700	0.1712	Yes	14	5.31	0.0607	No	Means	Wand	93.79%
86"Non vs W/S				0.0400	0.8500	Yes	14	7.71	0.0322	No	Means	W/S	5.81%



**Figure G- 20: B3Np3Nav Novice Users Pass 3 Navigation Ratings**

**Table G- 20: B3Np3Nav Novice Users Pass 3 Navigation Ratings Statistics**

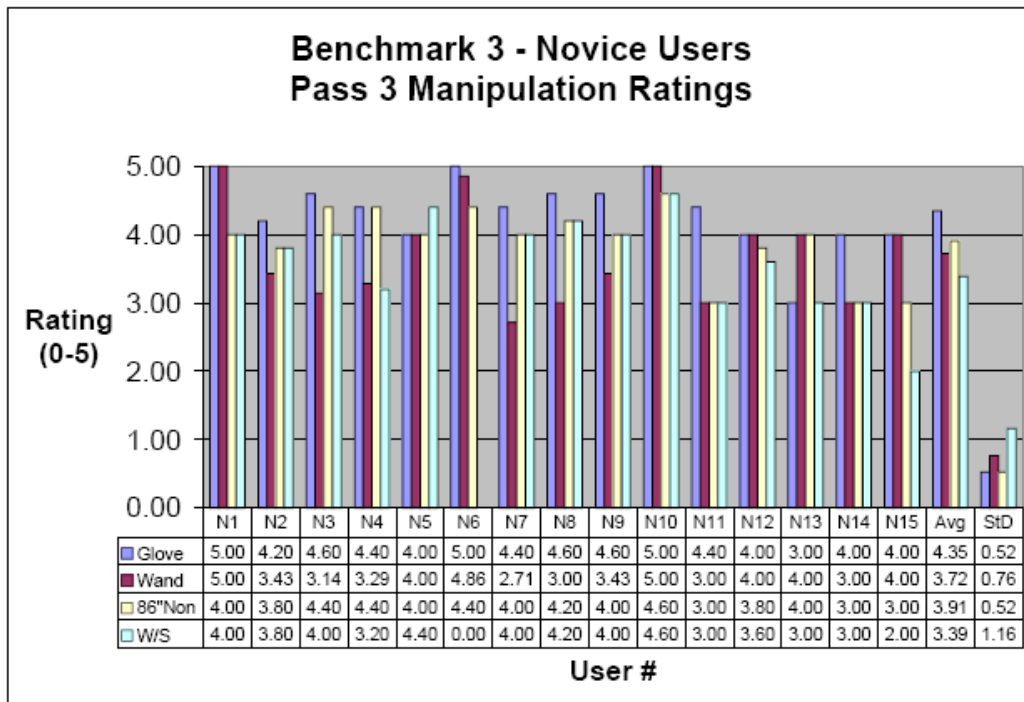
Benchmark 3, Novice Users, Pass 3, Navigation Ratings												
B3Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.54	0.3937	4.00	5.00	> 0.1000	Yes	22.53%				
Wand	15	4.35	0.5686	3.00	5.00	> 0.1000	Yes	24.73%				
86"Non	15	4.12	0.4639	3.00	4.88	> 0.1000	Yes	29.82%				
W/S	15	3.88	0.5736	2.00	4.43	0.0095	No	27.39%				
B3Np3	Homogeneity of Variance				Roy's Greatest Root: F(3,12)				6.72		Pr > F	0.0016
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand	1.7898	0.1810			Yes	14	0.14	0.7148	Yes	Neither	Neither	
Glove vs 86"Non	0.3700	0.5430			Yes	14	14.03	0.0008	No	Means	Glove 9.20%	
Glove vs W/S	1.8928	0.1689			Yes	14	13.68	0.0009	No	Means	Glove 14.66%	
Wand vs 86"Non	0.5482	0.4590			Yes	14	14.69	0.0006	No	Means	Wand 5.24%	
Wand vs W/S			0.0000	0.9801	Yes	14	13.64	0.0009	No	Means	Wand 10.94%	
86"Non vs W/S			0.1900	0.6686	Yes	14	0.00	1.0000	Yes	Neither	Neither	



**Figure G- 21: B3Np3Loc Novice Users Pass 3 Locating Ratings**

**Table G- 21: B3Np3Loc Novice Users Pass 3 Locating Ratings Statistics**

Benchmark 3, Novice Users, Pass 3, Locating Ratings										
B3Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	15	4.54	0.3937	4.00	5.00	> 0.1000	Yes	22.53%		
Wand	15	4.35	0.5686	3.00	5.00	> 0.1000	Yes	24.73%		
86\"Non	15	4.12	0.4639	3.00	4.88	> 0.1000	Yes	29.82%		
W/S	15	3.88	0.5736	2.00	4.43	0.0095	No	27.39%		
Homogeneity of Variance				Roy's Greatest Root: F(3,12)			6.72	Pr > F	0.0016	
B3Np3	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						Statistically Better
Glove vs Wand	1.7898	0.1810			Yes	14	0.14	0.7148	Yes	Neither
Glove vs 86\"Non	0.3700	0.5430			Yes	14	14.03	0.0008	No	Means
Glove vs W/S	1.8928	0.1689			Yes	14	13.68	0.0009	No	Means
Wand vs 86\"Non	0.5482	0.4590			Yes	14	14.69	0.0006	No	Means
Wand vs W/S			0.0000	0.9801	Yes	14	13.64	0.0009	No	Means
86\"Non vs W/S			0.1900	0.6686	Yes	14	0.00	1.0000	Yes	Neither

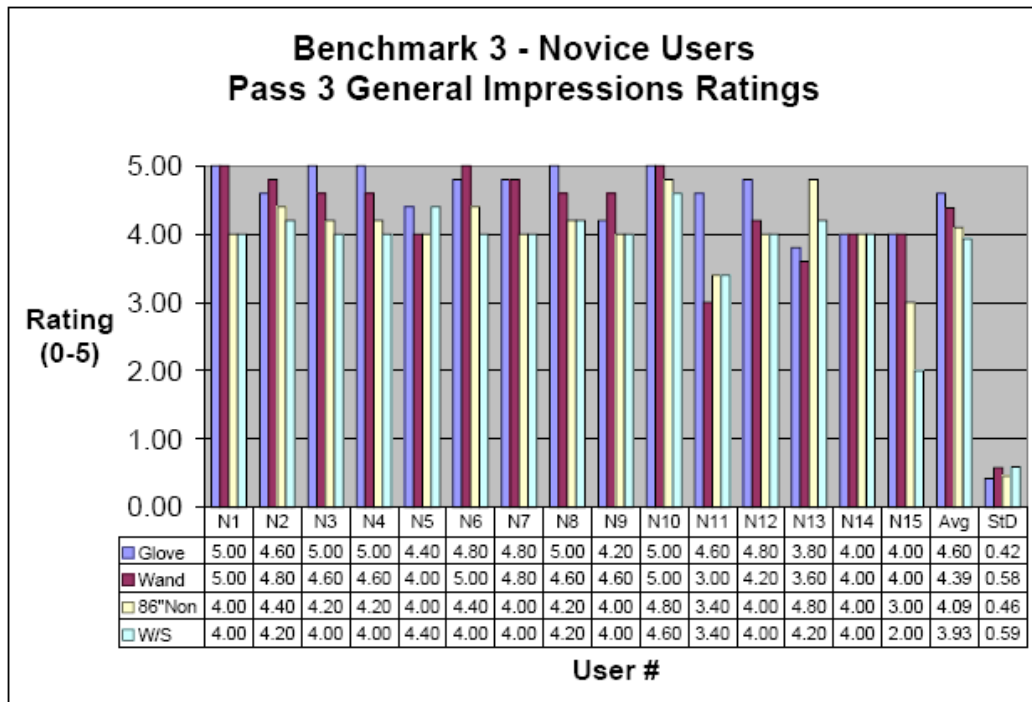


**Figure G- 22: B3Np3Mov Novice Users Pass 3 Manipulation Ratings**

**Table G- 22: B3Np3Mov Novice Users Pass 3 Manipulation Ratings Statistics**

Benchmark 3, Novice Users, Pass 3, Manipulation Ratings											
B3Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.35	0.5208	3.00	5.00	> 0.1000	Yes	22.53%			
Wand	15	3.72	0.7626	2.71	5.00	> 0.1000	Yes	24.73%			
86"Non	15	3.91	0.5230	3.00	4.60	0.0226	No	29.82%			
W/S	15	3.64	0.6812	2.00	4.60	> 0.1000	Yes	27.39%			
B3Np3	Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	5.51	Pr > F	0.013	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	1.9277	0.1650			Yes	14	9.16	0.7148	Yes	Neither	Neither
Glove vs 86"Non			0.0000	0.9884	Yes	14	9.04	0.0008	No	Means	Glove 10.12%
Glove vs W/S	7.8360	0.0051			No	14	5.70	0.0009	No	Var/Means	Glove 16.26%
Wand vs 86"Non			2.8800	0.1010	Yes	14	0.85	0.0006	No	Means	86"Non 4.68%
Wand vs W/S	2.2801	0.1310			Yes	14	0.73	0.0009	No	Means	Wand 2.26%
86"Non vs W/S	7.7609	0.0053			No	14	2.97	1.0000	Yes	Var	86"Non 6.83%

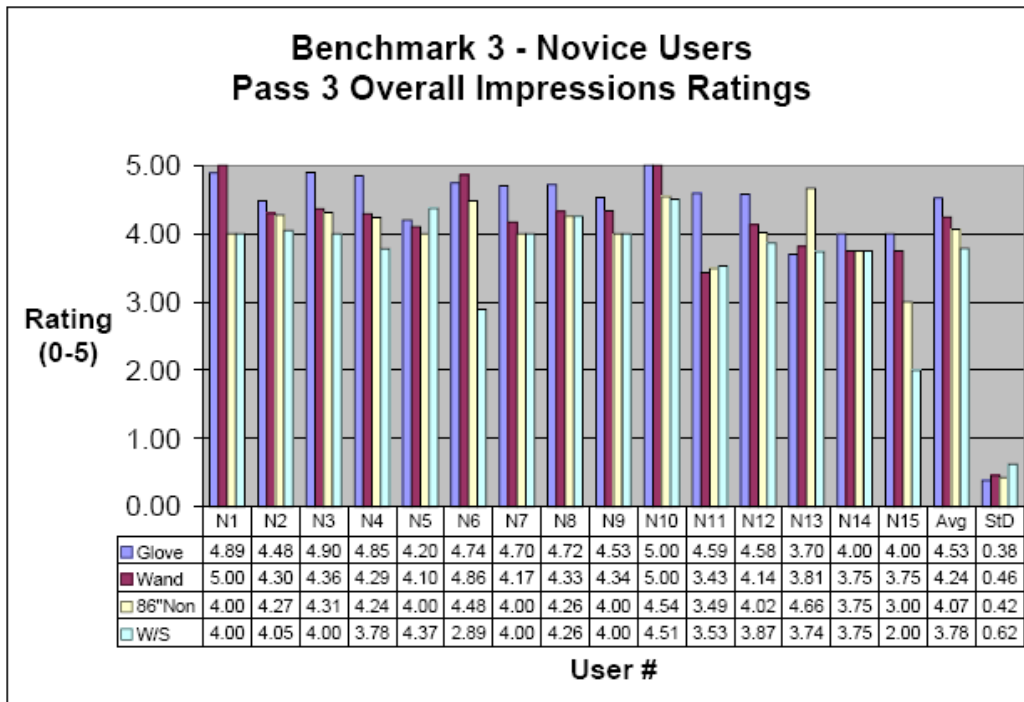




**Figure G- 23: B3Np3Gen Novice users Pass 3 General Impressions Ratings**

**Table G- 23: B3Np3Gen Novice users Pass 3 General Impressions Ratings Statistics**

Benchmark 3, Novice Users, Pass 3, General Impressions Ratings											
B3Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.60	0.4209	3.80	5.00	0.0580	Yes	22.53%			
Wand	15	4.39	0.5780	3.00	5.00	0.0166	No	24.73%			
86\"Non	15	4.09	0.4590	3.00	4.80	0.0017	No	29.82%			
W/S	15	3.93	0.5936	2.00	4.60	< 0.0001	No	27.39%			
B3Np3		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	5.69	Pr > F	0.0116
		Bartlett's Test		Levene's Test		df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				1.2500	0.2727	Yes	14	3.03	0.1038	Yes	Neither
Glove vs 86\"Non				0.1000	0.7518	Yes	14	12.75	0.0031	No	Means
Glove vs W/S				0.4300	0.5780	Yes	14	19.23	0.0006	No	Means
Wand vs 86\"Non				0.5800	0.4514	Yes	14	4.12	0.0619	No	Means
Wand vs W/S				3.4500	0.0741	Yes	14	7.15	0.0182	No	Means
86\"Non vs W/S				0.2600	0.6161	Yes	14	3.69	0.0753	No	Means

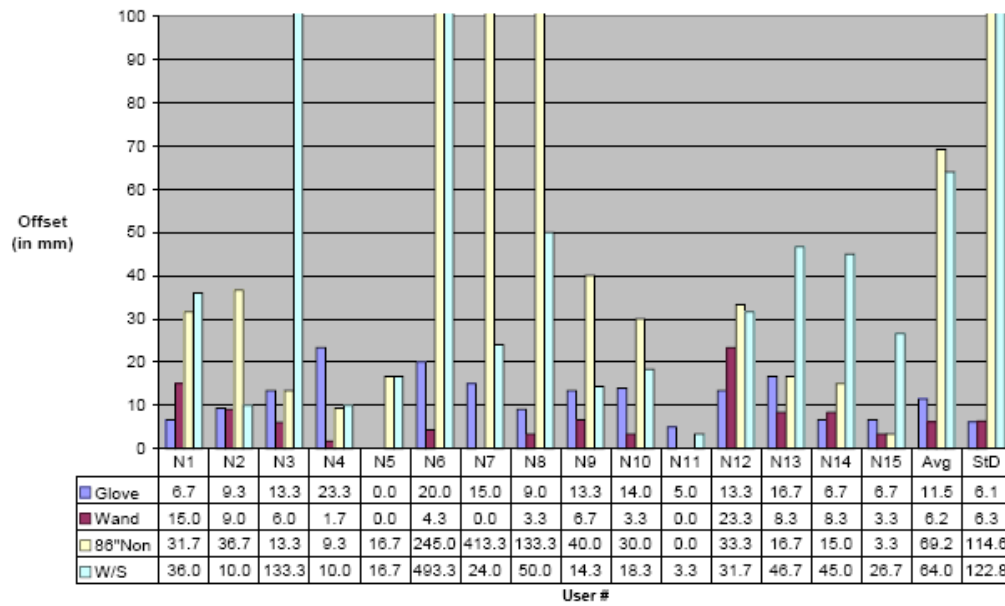


**Figure G- 24: B3Np3Ovr Novice Users Pass 3 Overall Impressions Ratings**

**Table G- 24: B3Np3Ovr Novice Users Pass 3 Overall Impressions Ratings Statistics**

Benchmark 3, Novice Users, Pass 3, Overall Impressions Ratings										
B3Np3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	15	4.53	0.3846	3.70	5.00	> 0.1000	Yes	22.53%		
Wand	15	4.24	0.4582	3.43	5.00	> 0.1000	Yes	24.73%		
86\"Non	15	4.09	0.4231	3.00	4.66	0.0237	No	29.82%		
W/S	15	3.78	0.6211	2.00	4.51	0.0043	No	27.39%		
B3Np3		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	
		Bartlett's Test		Levene's Test		df	F(1,14)	Pr > F	Equal Means ?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					
Glove vs Wand		0.4126	0.5206			Yes	14	10.39	0.0061	No
Glove vs 86\"Non				0.1100	0.7446	Yes	14	13.57	0.0025	No
Glove vs W/S				1.0700	0.3087	Yes	14	23.03	0.0003	No
Wand vs 86\"Non				0.0800	0.7738	Yes	14	2.70	0.1223	Yes
Wand vs W/S				0.5700	0.4573	Yes	14	7.67	0.0150	No
86\"Non vs W/S				0.7500	0.3932	Yes	14	4.65	0.0488	No
						8.43		Pr > F		0.0028
								Statistically Better		
Glove vs Wand								Means		Glove 6.27%
Glove vs 86\"Non								Means		Glove 9.72%
Glove vs W/S								Means		Glove 16.44%
Wand vs 86\"Non								Neither		Neither
Wand vs W/S								Means		Wand 10.84%
86\"Non vs W/S								Means		86\"Non 7.44%

**Benchmark 3 - Novice Users - 3 Pass Avg. - Part 1 Offsets**

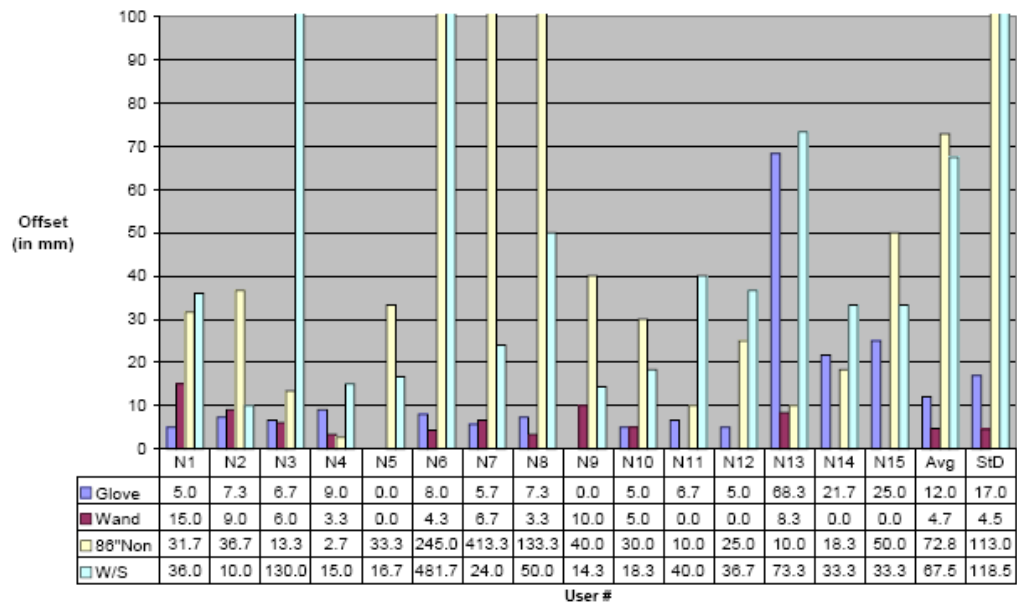


**Figure G- 25: B3N3pA-1Off Novice Users 3 Pass Avg. Icon 1 Offsets**

**Table G- 25: B3N3pA-1Off Novice Users 3 Pass Avg. Icon 1 Offsets Statistics**

Benchmark 2, Novice Users, 3 Pass Avg., Total Offsets											
B3NAvg	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	23.53	19.03	0.00	85.00	0.0095	No	22.53%			
Wand	15	10.91	8.54	0.00	30.00	0.0947	Yes	24.73%			
86\"Non	15	142.00	227.20	0.00	326.70	< 0.0001	No	29.82%			
W/S	15	131.50	241.00	0.00	975.00	< 0.0001	No	27.39%			
B3NAvg		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	2.16	Pr > F	0.2352
		Bartlett's Test	Levene's Test	Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				1.7300	0.1993	Yes	14	1.60	0.2528	Yes	Neither
Glove vs 86\"Non				2.4700	0.1273	Yes	14	5.82	0.0524	No	Means
Glove vs W/S				1.3100	0.2629	Yes	14	0.59	0.4705	Yes	Neither
Wand vs 86\"Non				2.5200	0.1239	Yes	14	0.30	0.6033	Yes	Neither
Wand vs W/S				1.3200	0.2610	Yes	14	1.70	0.2395	Yes	Neither
86\"Non vs W/S				0.3500	0.5573	Yes	14	3.60	0.0494	No	Means
										W/S	7.39%

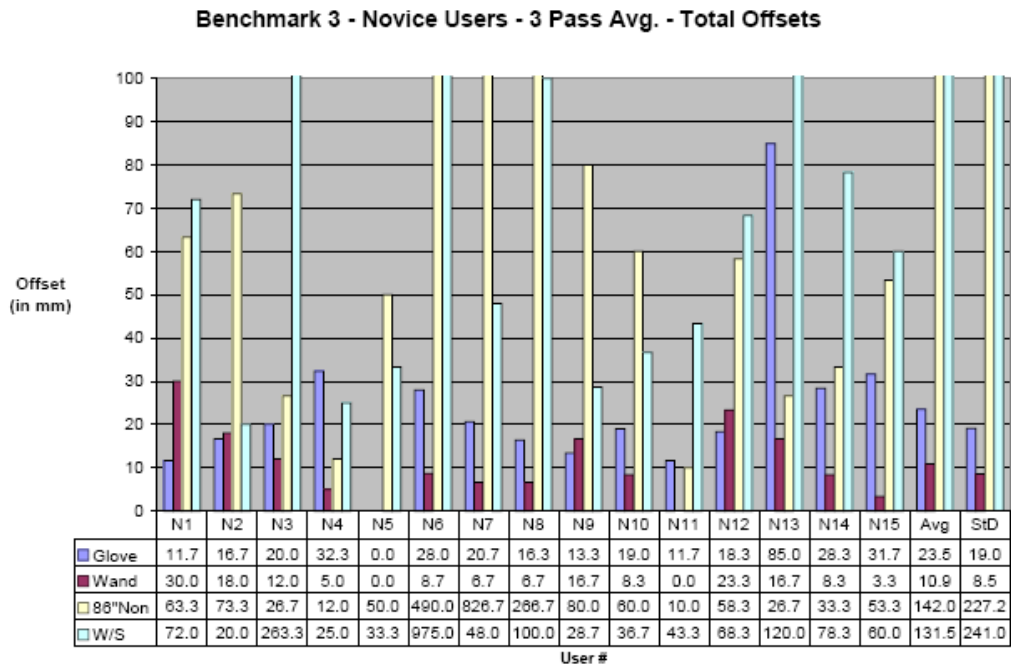
**Benchmark 3 - Novice Users - 3 Pass Avg. - Part 2 Offsets**



**Figure G- 26: B3N3pA-2of Novice Users 3 Pass Avg. Icon 2 Offsets**

**Table G- 26: B3N3pA-2of Novice Users 3 Pass Avg. Icon 2 Offsets Statistics**

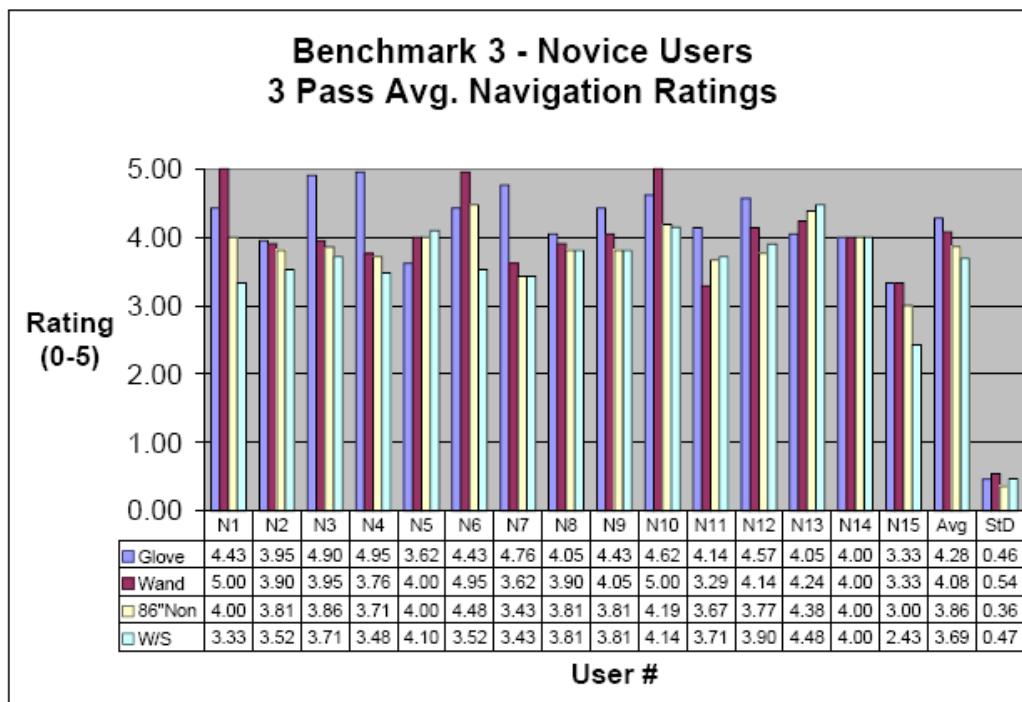
Benchmark 2, Novice Users, 3 Pass Avg., Part 2 Offsets											
B3N2Avg	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	12.04	16.98	0.00	68.33	< 0.0001	No	22.53%			
Wand	15	4.73	4.52	0.00	15.00	> 0.1000	Yes	24.73%			
86*Non	15	72.84	113.00	0.00	413.30	< 0.0001	No	29.82%			
W/S	15	67.51	118.50	3.33	481.70	< 0.0001	No	27.39%			
B3N2Avg	Homogeneity of Variance				Roy's Greatest Root: F(3,12)			2.17	Pr > F 0.1447		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				Yes	14	5.70	0.0120	No	Means	Wand	60.69%
Glove vs 86*Non			2.4400	0.1293	Yes	14	3.92	0.2860	Yes	Neither	Neither
Glove vs W/S			1.3300	0.2583	Yes	14	2.84	0.7610	Yes	Neither	Neither
Wand vs 86*Non			2.4400	0.1293	Yes	14	4.40	0.0547	No	Means	Wand 93.50%
Wand vs W/S			1.3300	0.2574	Yes	14	3.30	0.0908	No	Means	Wand 92.99%
86*Non vs W/S			0.0200	0.9010	Yes	14	0.02	0.8790	Yes	Neither	Neither



**Figure G- 27: B3N3pAtotOff Novice Users 3 Pass Avg. Total Offsets**

**Table G- 27: B3N3pAtotOff Novice Users 3 Pass Avg. Total Offsets Statistics**

Benchmark 2, Novice Users, 3 Pass Avg., Total Offsets												
B3NAvg	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	23.53	19.03	0.00	85.00	0.0095	No	22.53%				
Wand	15	10.91	8.54	0.00	30.00	0.0947	Yes	24.73%				
86"Non	15	142.00	227.20	0.00	326.70	< 0.0001	No	29.82%				
W/S	15	131.50	241.00	0.00	975.00	< 0.0001	No	27.39%				
B3NAvg		Homogeneity of Variance				Roy's Greatest Root: F(3,12)			2.16	Pr > F 0.2352		
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				1.7300	0.1993	Yes	14	1.60	0.2528	Yes	Neither	Neither
Glove vs 86"Non				2.4700	0.1273	Yes	14	5.82	0.0524	No	Means	Glove 83.43%
Glove vs W/S				1.3100	0.2629	Yes	14	0.59	0.4705	Yes	Neither	Neither
Wand vs 86"Non				2.5200	0.1239	Yes	14	0.30	0.6033	Yes	Neither	Neither
Wand vs W/S				1.3200	0.2610	Yes	14	1.70	0.2395	Yes	Neither	Neither
86"Non vs W/S				0.3500	0.5573	Yes	14	3.60	0.0494	No	Means	W/S 7.39%



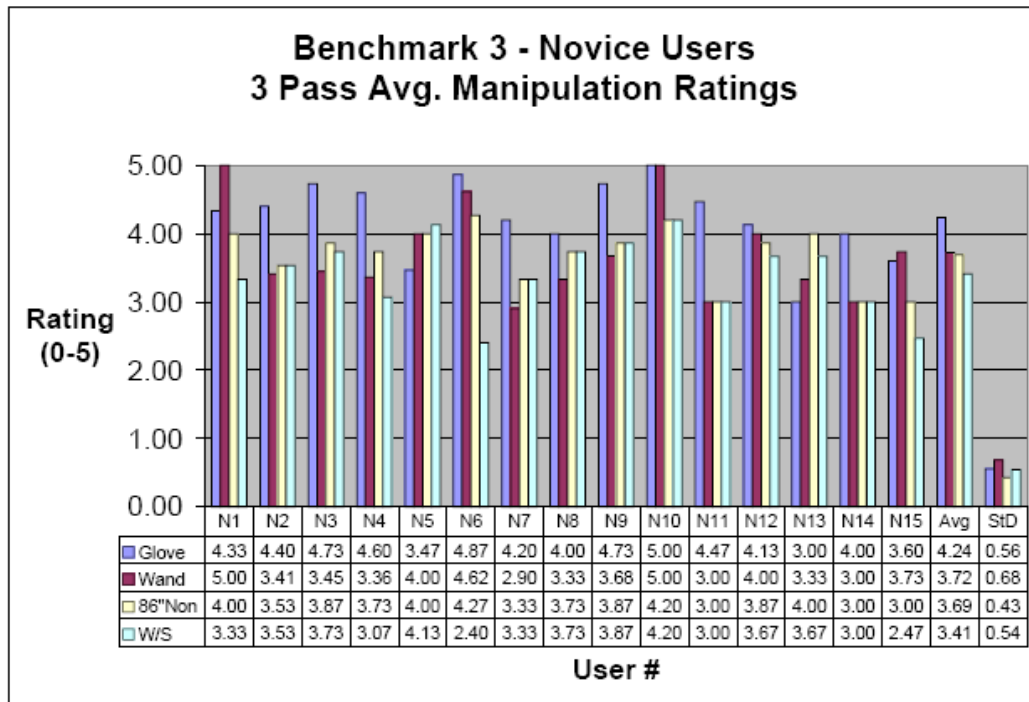
**Figure G- 28: B3N3pANav Novice Users 3 Pass Avg. Navigation Ratings**

**Table G- 28: B3N3pANav Novice Users 3 Pass Avg. Navigation Ratings Statistics**

Benchmark 3, Novice Users, 3 Pass Avg., Navigation Ratings													
B3NAvg	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV					
Glove	15	4.28	0.4618	3.33	4.95	> 0.1000	Yes	22.53%					
Wand	15	4.08	0.5394	3.29	5.00	> 0.1000	Yes	24.73%					
86"Non	15	3.86	0.3605	3.00	4.48	> 0.1000	Yes	29.82%					
W/S	15	3.69	0.4665	2.43	4.48	> 0.1000	Yes	27.39%					
B3NAvg		Homogeneity of Variance					Roy's Greatest Root: F(3,12)		2.38	Pr > F		0.0921	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand		0.3282	0.5667			Yes	14	2.93	0.0976	No	Means	Glove	4.81%
Glove vs 86"Non		0.8131	0.3672			Yes	14	4.23	0.0488	No	Means	Glove	9.83%
Glove vs W/S		0.0018	0.9662			Yes	14	3.43	0.0744	No	Means	Glove	13.78%
Wand vs 86"Non		2.1347	0.1440			Yes	14	5.15	0.0308	No	Means	Wand	5.27%
Wand vs W/S		0.2817	0.5956			Yes	14	4.26	0.0482	No	Means	Wand	9.42%
86"Non vs W/S		0.8901	0.3454			Yes	14	0.01	0.9184	Yes	Neither	Neither	

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	Avg	Std
Glove	4.5	4.1	4.9	4.9	4.1	4.7	4.3	4.4	4.3	4.9	5.0	3.9	3.9	4.6	4.0	4.6	3.3	3.7	3.7	3.8	5.0	4.30	0.49
Wand	5.0	4.0	4.1	3.9	4.0	5.0	3.3	3.9	3.9	5.0	5.0	4.0	4.1	3.9	4.0	5.0	3.5	3.9	3.9	5.0	5.0	4.24	0.57
86°Non	4.0	4.0	3.9	3.7	4.0	4.1	3.3	3.7	3.7	4.1	4.0	4.0	3.9	3.7	4.0	4.1	3.3	3.7	3.7	4.1	3.7	3.85	0.23
W/S	3.3	3.8	3.9	3.7	4.1	3.7	3.3	3.7	3.7	4.1	3.3	3.8	3.9	3.7	4.1	3.7	3.3	3.7	3.7	4.1	3.7	3.75	0.26

Benchmark 3, Novice Users, 3 Pass Avg., Locating Ratings												
B3NAvg	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	4.28	0.4618	3.33	4.95	> 0.1000	Yes	22.53%				
Wand	15	4.08	0.5394	3.29	5.00	> 0.1000	Yes	24.73%				
86"Non	15	3.86	0.3605	3.00	4.48	> 0.1000	Yes	29.82%				
W/S	15	3.69	0.4665	2.43	4.48	> 0.1000	Yes	27.39%				
B3NAvg	Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	2.38	Pr > F		0.0921	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand	0.3282	0.5667			Yes	14	2.93	0.0976	No	Means	Glove	4.81%
Glove vs 86"Non	0.8131	0.3672			Yes	14	4.23	0.0488	No	Means	Glove	9.83%
Glove vs W/S	0.0018	0.9662			Yes	14	3.43	0.0744	No	Means	Glove	13.78%
Wand vs 86"Non	2.1347	0.1440			Yes	14	5.15	0.0308	No	Means	Wand	5.27%
Wand vs W/S	0.2817	0.5956			Yes	14	4.26	0.0482	No	Means	Wand	9.42%
86"Non vs W/S	0.8901	0.3454			Yes	14	0.01	0.9184	Yes	Neither	Neither	

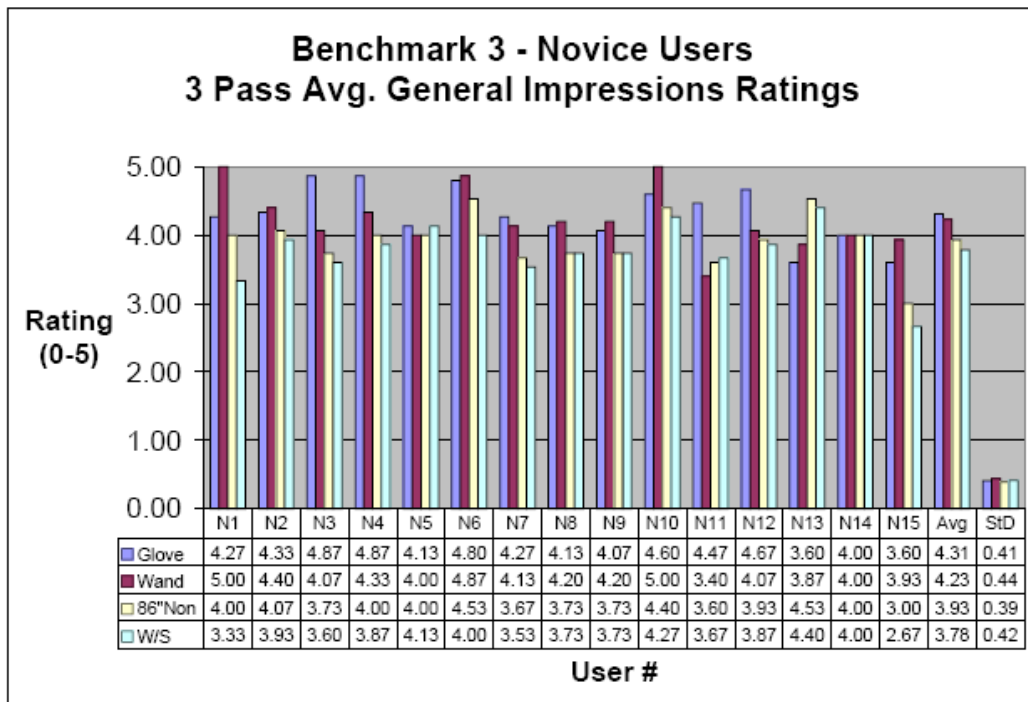


**Figure G- 30: B3N3pAMov Novice Users 3 Pass Avg. Manipulation Ratings**

**Table G- 30: B3N3pAMov Novice Users 3 Pass Avg. Manipulation Ratings Statistics**

Benchmark 3, Novice Users, 3 Pass Avg., Manipulation Ratings											
B3NAvg	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.24	0.5571	3.00	5.00	> 0.1000	Yes	22.53%			
Wand	15	3.72	0.6836	2.91	5.00	> 0.1000	Yes	24.73%			
86\"Non	15	3.69	0.4275	3.00	4.27	0.0937	Yes	29.82%			
W/S	15	3.41	0.5404	2.40	4.20	> 0.1000	Yes	27.39%			
B3NAvg	Homogeneity of Variance					Roy's Greatest Root:		F(3,12)	5.11	Pr > F	0.0165
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.5718	0.4495			Yes	14	7.38	0.0167	No	Means	Glove 12.16%
Glove vs 86\"Non	0.9190	0.3327			Yes	14	11.45	0.0045	No	Means	Glove 12.82%
Glove vs W/S	0.0132	0.9086			Yes	14	16.51	0.0012	No	Means	Glove 19.52%
Wand vs 86\"Non	2.8658	0.0905			Yes	14	0.05	0.8309	Yes	Neither	Wand 0.75%
Wand vs W/S	0.7563	0.3845			Yes	14	2.21	0.1596	Yes	Neither	Neither
86\"Non vs W/S	0.7149	0.3978			Yes	14	4.68	0.0483	No	Means	86\"Non 7.69%

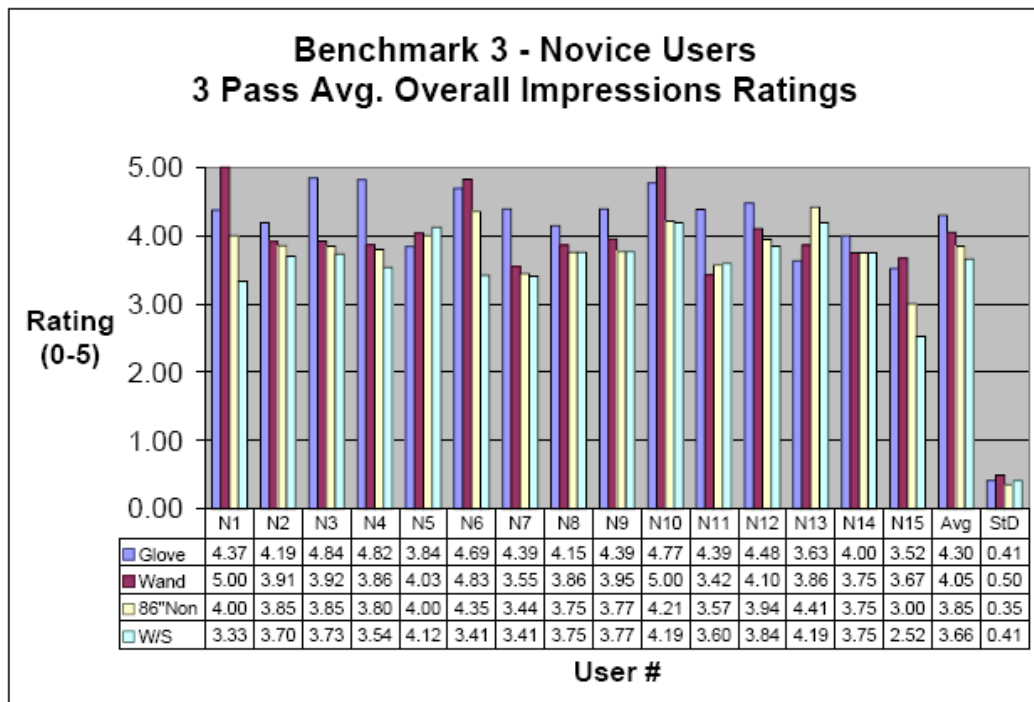




**Figure G- 31: B3N3pAgen Novice Users 3 Pass Avg. General Impressions Ratings**

**Table G- 31: B3N3pAgen Novice Users 3 Pass Avg. General Impressions Ratings Statistics**

Benchmark 3, Novice Users, 3 Pass Avg., General Impressions Ratings											
B3NAvg	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.31	0.4076	3.60	4.87	> 0.1000	Yes	22.53%			
Wand	15	4.23	0.4400	3.40	5.00	> 0.1000	Yes	24.73%			
86\"Non	15	3.93	0.3916	3.00	4.53	> 0.1000	Yes	29.82%			
W/S	15	3.78	0.4167	2.67	4.40	> 0.1000	Yes	27.39%			
B3NAvg	Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	4.88	Pr > F	0.0191	
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.0747	0.7846			Yes	14	0.41	0.0167	No	Means	Glove 1.86%
Glove vs 86\"Non	0.0258	0.8725			Yes	14	9.39	0.0045	No	Means	Glove 8.86%
Glove vs W/S	0.0051	0.9433			Yes	14	15.50	0.0012	No	Means	Glove 12.27%
Wand vs 86\"Non	0.1879	0.6647			Yes	14	7.79	0.8309	Yes	Neither	Neither
Wand vs W/S	0.0409	0.2397			Yes	14	9.38	0.1596	Yes	Neither	Neither
86\"Non vs W/S	0.0536	0.8169			Yes	14	6.94	0.0483	No	Means	86\"Non 3.74%



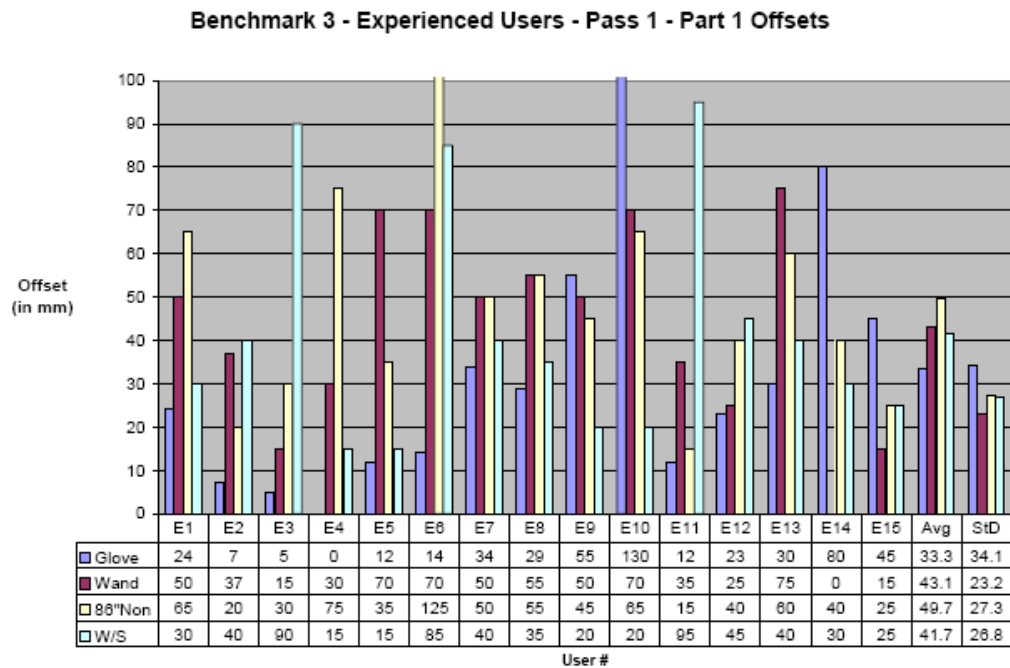
**Figure G- 32: B3N3pAvgOvr Novice User 3 Pass Avg. Overall Impressions Ratings**

**Table G- 32: B3N3pAvgOvr Novice User 3 Pass Avg. Overall Impressions Ratings Statistics**

Benchmark 3, Novice Users, 3 Pass Avg., Overall Impressions Ratings											
B3NAvg	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	4.53	0.3846	3.52	4.84	> 0.1000	Yes	22.53%			
Wand	15	4.24	0.4582	3.42	5.00	0.0090	No	24.73%			
86\"Non	15	4.07	0.4231	3.00	4.41	> 0.1000	Yes	29.82%			
W/S	15	3.78	0.6211	2.52	4.19	> 0.1000	Yes	27.39%			
B3NAvg		Homogeneity of Variance				Roy's Greatest Root:		F(3,12)	6.34	Pr > F	0.008
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						Statistically Better
Glove vs Wand		0.4544	0.5002			Yes	14	3.64	0.0771	No	Means
Glove vs 86\"Non				0.3900	0.5375	Yes	14	14.06	0.0022	No	Means
Glove vs W/S				0.0000	0.9990	Yes	14	21.56	0.0004	No	Means
Wand vs 86\"Non		1.5462	0.2137			Yes	14	4.03	0.0644	No	Means
Wand vs W/S		0.4525	0.5012			Yes	14	6.53	0.0229	No	Means
86\"Non vs W/S				0.1900	0.6648	Yes	14	6.34	0.0246	No	Means

## **APPENDIX H**

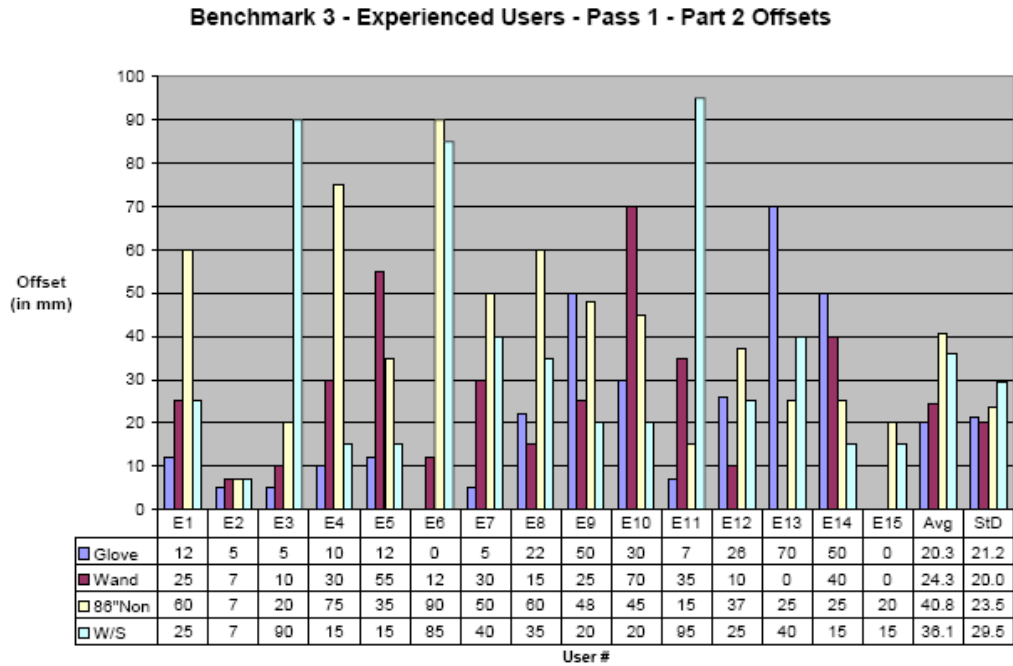
### Experienced User Benchmark 3 (Spatial Awareness) Detail



**Figure H- 1: B3Ep1-1Off Experienced Users Pass 1-Icon 1 Offsets**

**Table H- 1: B3Ep1-1Off Experienced Users Pass 1-Icon 1 Offsets Statistics**

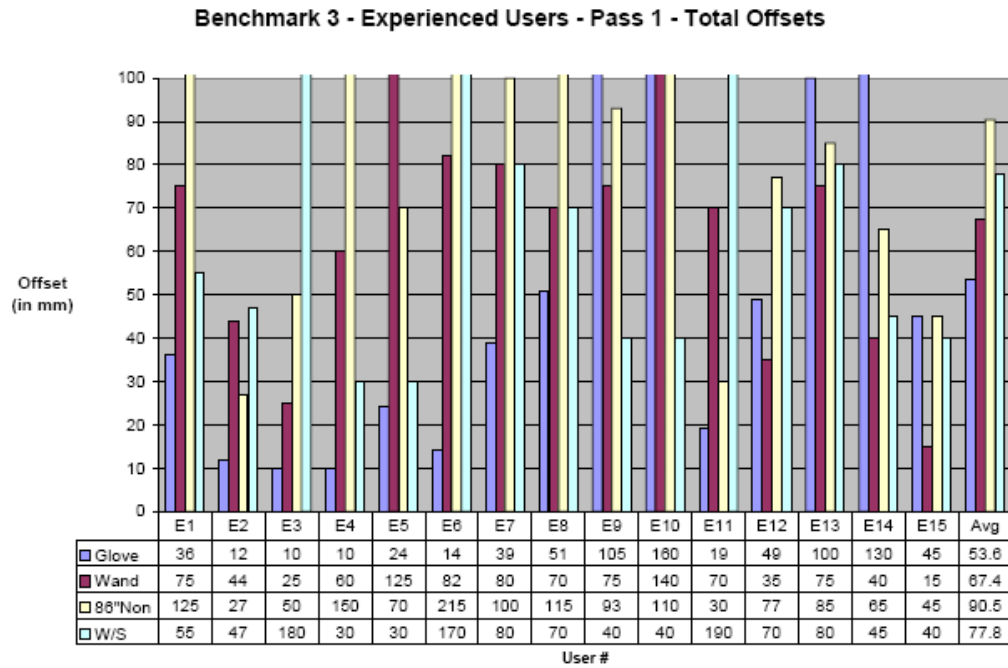
Benchmark 3, Experienced Users, Pass 1, Part 1 Offsets										
B3Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	15	33.33	34.06	0.00	130.00	0.0387	No	102.19%		
Wand	15	43.13	23.16	0.00	75.00	> 0.1000	Yes	53.70%		
86\"Non	15	49.67	27.29	15.00	125.00	> 0.1000	Yes	54.94%		
W/S	15	41.67	26.77	15.00	95.00	0.0082	No	64.24%		
B3Ep1		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	Pr > F
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					Significant?
Glove vs Wand				0.8700	0.3577	Yes	14	0.41	0.3351	Yes
Glove vs 86\"Non				0.3000	0.5892	Yes	14	2.21	0.1597	Yes
Glove vs W/S				0.4000	0.5313	Yes	14	0.40	0.5367	Yes
Wand vs 86\"Non		0.3610	0.5479			Yes	14	1.00	0.3345	Yes
Wand vs W/S				0.3800	0.5430	Yes	14	0.20	0.8815	Yes
86\"Non vs W/S				0.0000	0.9531	Yes	14	0.67	0.4254	Yes
										Statistically Better



**Figure H- 2: B3Ep1-2Off Experienced Users Pass 1-Icon 2 Offsets**

**Table H- 2: B3Ep1-2Off Experienced Users Pass 1-Icon 2 Offsets Statistics**

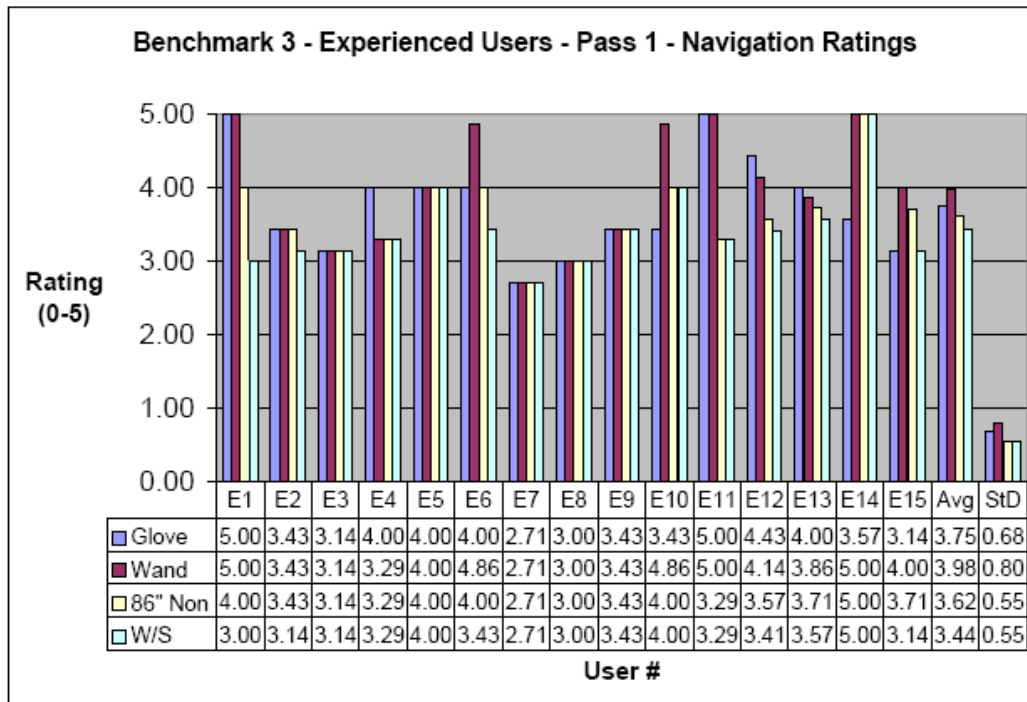
Benchmark 3, Experienced Users, Pass 1, Part 2 Offsets											
B3Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	20.27	21.23	0.00	70.00	0.0115	No	104.74%			
Wand	15	24.27	19.97	0.00	70.00	> 0.1000	Yes	82.28%			
86\"Non	15	40.80	23.46	7.00	90.00	> 0.1000	Yes	57.50%			
W/S	15	36.13	29.51	7.00	95.00	0.0137	No	81.68%			
B3Ep1		Homogeneity of Variance				Roy's Greatest Root:				F(3,12) 0.74 Pr > F 0.551	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				0.0000	1.0000	Yes	14	0.91	0.3551	Yes	Neither
Glove vs 86\"Non				1.8000	0.1892	Yes	14	2.21	0.1597	Yes	Neither
Glove vs W/S				0.6800	0.4172	Yes	14	0.40	0.5367	Yes	Neither
Wand vs 86\"Non		6.8209	0.0090			No	14	1.00	0.3345	Yes	Var
Wand vs W/S				1.1500	0.2918	Yes	14	0.02	0.8815	Yes	Neither
86\"Non vs W/S				0.5400	0.4688	Yes	14	0.67	0.4254	Yes	Neither



**Figure H- 3: B3Ep1TotOff Experienced Users Pass 1 Total Offsets**

**Table H- 3: B3Ep1TotOff Experienced Users Pass 1 Total Offsets Statistics**

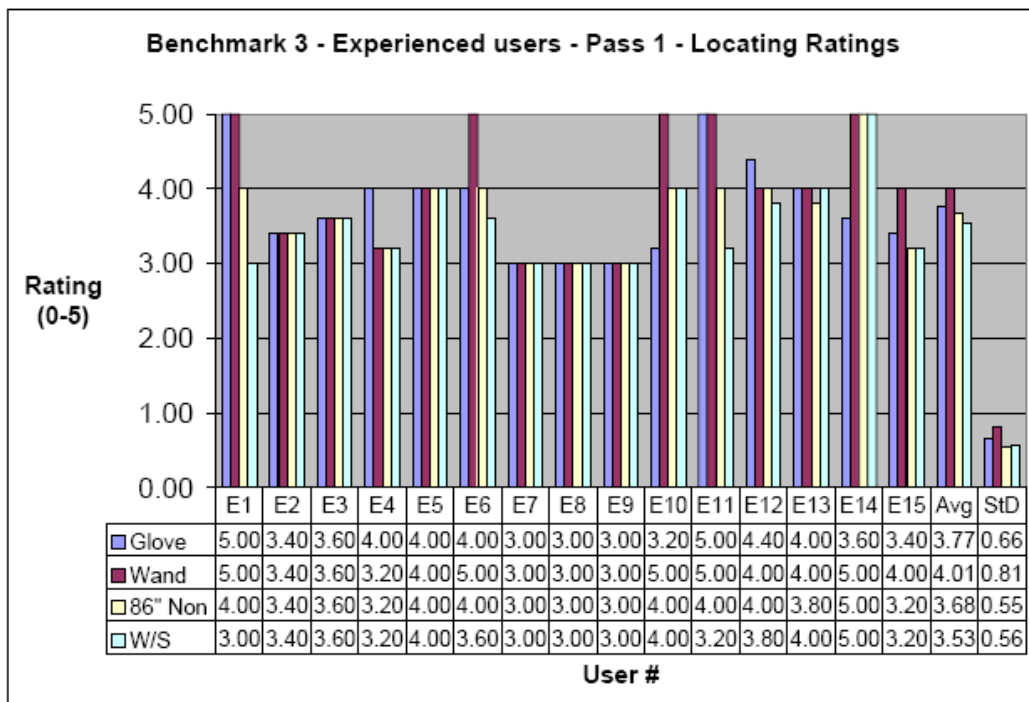
Benchmark 3, Experienced Users, Pass 1, Total Offsets												
B3Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	34.89	17.90	0.00	85.00	> 0.1000	Yes	51.30%				
Wand	15	30.06	26.19	0.00	95.00	> 0.1000	Yes	87.13%				
86"Non	15	106.50	220.80	0.00	1000.00	< 0.0001	No	207.32%				
W/S	15	58.90	75.79	0.00	400.00	0.0008	No	128.68%				
B3Ep1	Homogeneity of Variance					Roy's Greatest Root: F(3,12) 1.7 Pr > F 0.2165						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand	0.0000	1.0000			Yes	14	1.93	0.2800	Yes	Neither	Neither	
Glove vs 86"Non			0.0000	1.0000	Yes	14	1.53	0.2384	Yes	Neither	Neither	
Glove vs W/S			1.1800	0.1892	Yes	14	0.66	0.9090	Yes	Neither	Neither	
Wand vs 86"Non			0.6800	0.4172	Yes	14	0.44	0.0260	No	Means	Wand 254.29%	
Wand vs W/S			2.2400	0.1454	Yes	14	2.93	0.0000	No	Means	Wand 48.96%	
86"Non vs W/S			1.1500	0.2918	Yes	14	3.06	0.2070	Yes	Neither	Neither	



**Figure H- 4: B3Ep1Nav Experienced Users Pass 1 Navigation Ratings**

**Table H- 4: B3Ep1Nav Experienced Users Pass 1 Navigation Ratings Statistics**

Benchmark 3, Experienced Users, Pass 1, Navigation Ratings												
B3Ep1	# Users	Mean	St. Dev.		Low	High			P Value	Normal?	CV	
Glove	15	344.3	73.43		260.0	515.0			> 0.1000	Yes	21.34%	
Wand	15	301.9	158.50		165.0	780.0			0.0895	Yes	52.51%	
86"Non	15	325.9	128.90		185.0	615.0			>0.1000	Yes	39.55%	
W/S	15	346.7	91.28		225.0	525.0			> 0.1000	Yes	26.33%	
B3Ep1		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 0.7 Pr > F 0.5484						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14) Pr > F		Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		7.3102	0.0069			Yes	14	1.20	0.2920	Yes	Neither	Neither
Glove vs 86"Non		4.0610	0.0439			Yes	14	0.29	0.6000	Yes	Neither	Neither
Glove vs W/S		0.6311	0.4270			Yes	14	0.01	0.6300	Yes	Neither	Neither
Wand vs 86"Non		0.5751	0.4482			Yes	14	0.17	0.6827	Yes	Neither	Neither
Wand vs W/S		3.9256	0.0476			Yes	14	1.82	0.1984	Yes	Neither	Neither
86"Non vs W/S		1.5782	0.2090			Yes	14	0.21	0.6534	Yes	Neither	Neither

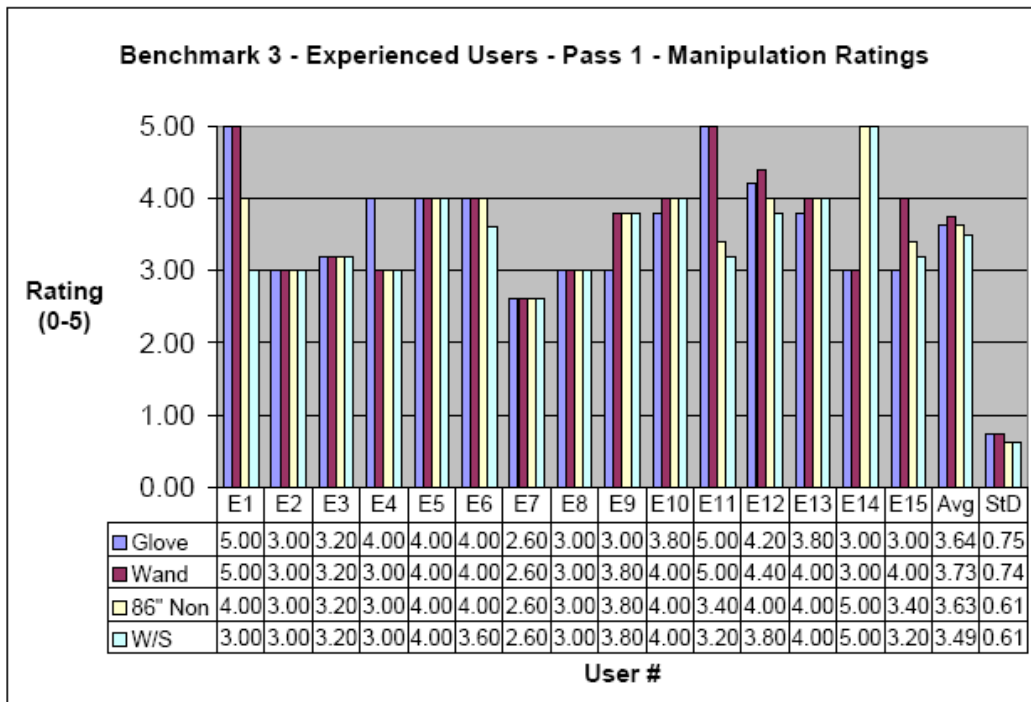


**Figure H- 5: B3Ep1Loc Experienced Users Pass 1 Locating Ratings**

**Table H- 5: B3Ep1Loc Experienced Users Pass 1 Locating Ratings Statistics**

Benchmark 3, Experienced Users, Pass 1, Locating Ratings												
B3Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	344.3	73.43	260.0	515.0	> 0.1000	Yes	21.34%				
Wand	15	301.9	158.50	165.0	780.0	0.0895	Yes	52.51%				
86"Non	15	325.9	128.90	185.0	615.0	>0.1000	Yes	39.55%				
W/S	15	346.7	91.28	225.0	525.0	> 0.1000	Yes	26.33%				
B3Ep1		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 0.74 Pr > F 0.5484						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		7.3102	0.0069			Yes	14	1.20	0.2920	Yes	Neither	Neither
Glove vs 86"Non		4.0610	0.0439			Yes	14	0.29	0.6000	Yes	Neither	Neither
Glove vs W/S		0.6311	0.4270			Yes	14	0.01	0.6300	Yes	Neither	Neither
Wand vs 86"Non		0.5751	0.4482			Yes	14	0.17	0.6827	Yes	Neither	Neither
Wand vs W/S		3.9256	0.0476			Yes	14	1.82	0.1984	Yes	Neither	Neither
86"Non vs W/S		1.5782	0.2090			Yes	14	0.21	0.6534	Yes	Neither	Neither

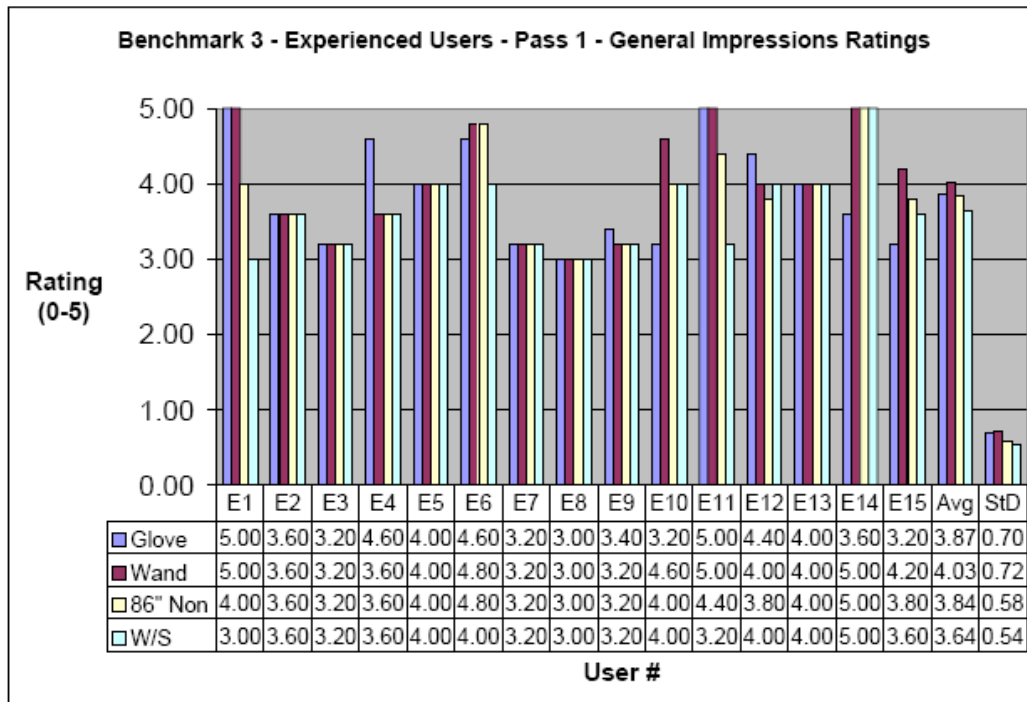




**Figure H- 6: B3Ep1Mov Experienced Users Pass 1 Manipulation Ratings**

**Table H- 6: B3Ep1Mov Experienced Users Pass 1 Manipulation Ratings Statistics**

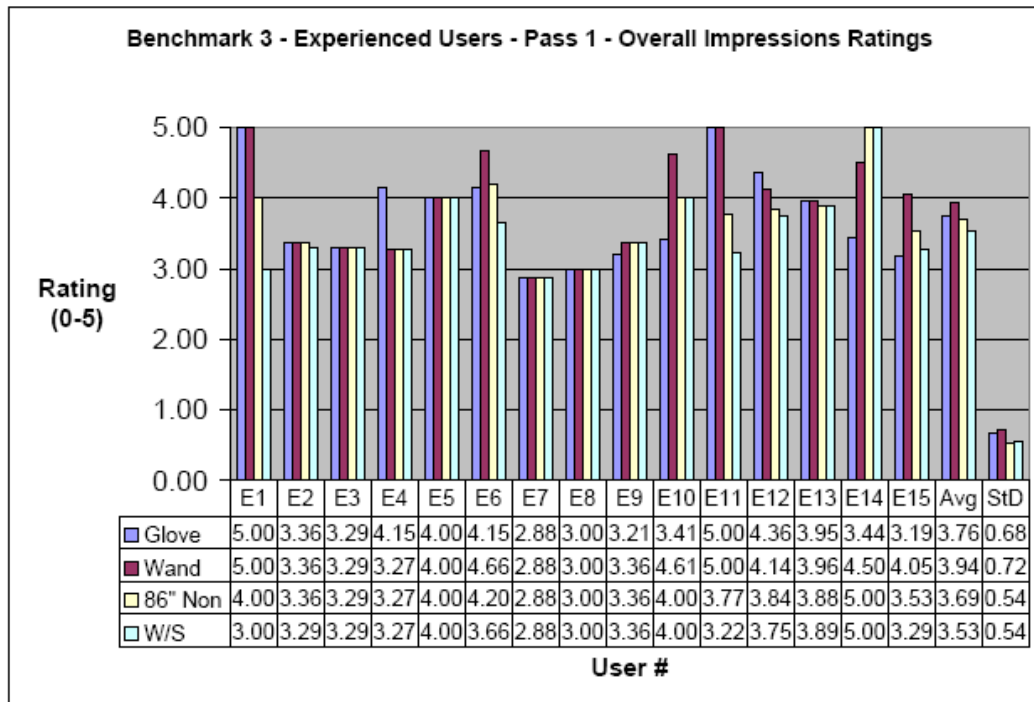
Benchmark 3. Experienced Users, Pass 1, Manipulation Ratings												
B3Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	344.3	73.43	260.0	515.0	> 0.1000	Yes	21.34%				
Wand	15	301.9	158.50	165.0	780.0	0.0895	Yes	52.51%				
86"Non	15	325.9	128.90	185.0	615.0	>0.1000	Yes	39.55%				
W/S	15	346.7	91.28	225.0	525.0	> 0.1000	Yes	26.33%				
B3Ep1		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	0.74	Pr > F	0.5484
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		7.3102	0.0069			Yes	14	1.20	0.2920	Yes	Neither	Neither
Glove vs 86"Non		4.0610	0.0439			Yes	14	0.29	0.6000	Yes	Neither	Neither
Glove vs W/S		0.6311	0.4270			Yes	14	0.01	0.6300	Yes	Neither	Neither
Wand vs 86"Non		0.5751	0.4482			Yes	14	0.17	0.6827	Yes	Neither	Neither
Wand vs W/S		3.9256	0.0476			Yes	14	1.82	0.1984	Yes	Neither	Neither
86"Non vs W/S		1.5782	0.2090			Yes	14	0.21	0.6534	Yes	Neither	Neither



**Figure H- 7: B3Ep1Gen Experienced Users Pass 1 General Impressions Ratings**

**Table H- 7: B3Ep1Gen Experienced Users Pass 1 General Impressions Ratings Statistics**

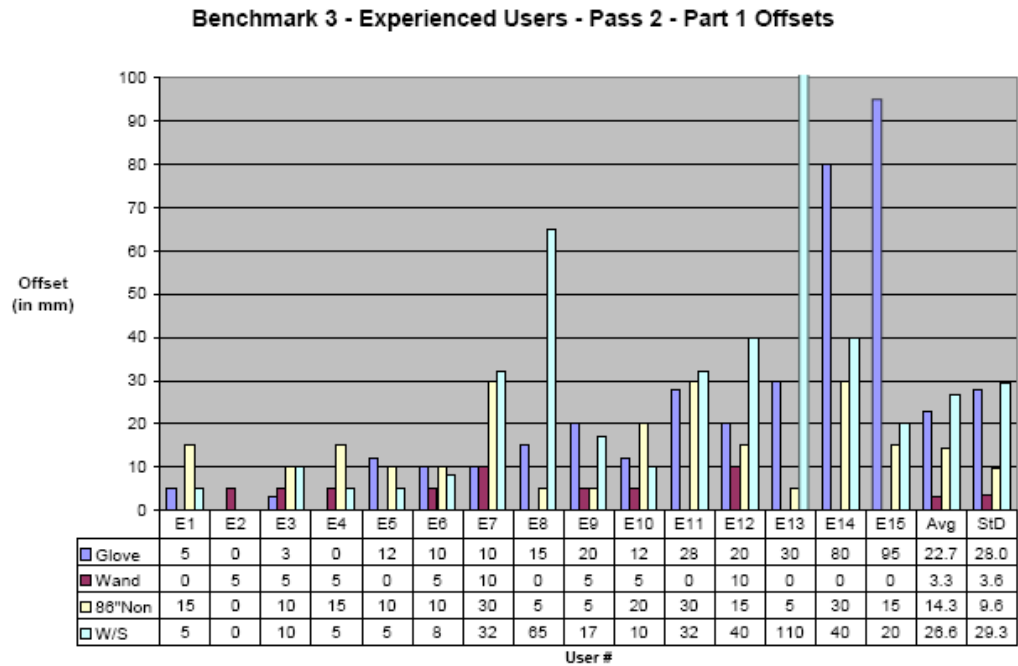
Benchmark 3, Experienced Users, Pass 1, General Impressions Ratings										
B3Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	15	344.3	73.43	260.0	515.0	> 0.1000	Yes	21.34%		
Wand	15	301.9	158.50	165.0	780.0	0.0895	Yes	52.51%		
86\" Non	15	325.9	128.90	185.0	615.0	>0.1000	Yes	39.55%		
W/S	15	346.7	91.28	225.0	525.0	> 0.1000	Yes	26.33%		
B3Ep1	Homogeneity of Variance				Roy's Greatest Root:				F(3,12)	Pr > F
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand	7.3102	0.0069			Yes	14	1.20	0.2920	Yes	Neither
Glove vs 86\" Non	4.0610	0.0439			Yes	14	0.29	0.6000	Yes	Neither
Glove vs W/S	0.6311	0.4270			Yes	14	0.01	0.6300	Yes	Neither
Wand vs 86\" Non	0.5751	0.4482			Yes	14	0.17	0.6827	Yes	Neither
Wand vs W/S	3.9256	0.0476			Yes	14	1.82	0.1984	Yes	Neither
86\" Non vs W/S	1.5782	0.2090			Yes	14	0.21	0.6534	Yes	Neither



**Figure H- 8: B3Ep1Ovr Experienced Users Pass 1 Overall Ratings**

**Table H- 8: B3Ep1Ovr Experienced Users Pass 1 Overall Ratings Statistics**

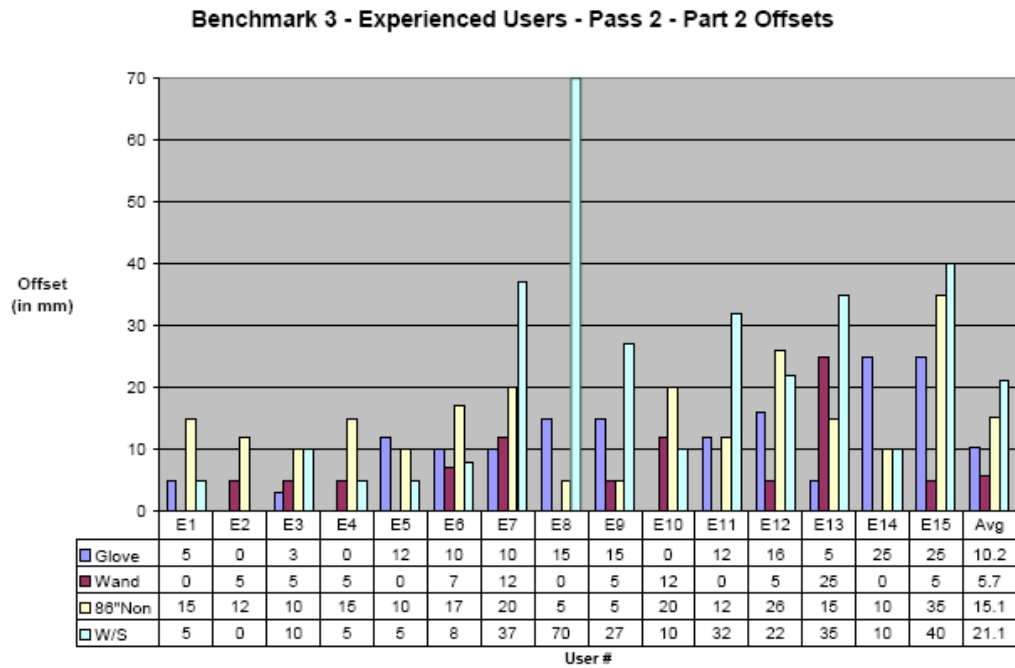
Benchmark 3, Experienced Users, Pass 1, Overall Impressions Ratings												
B3Ep1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	344.3	73.43	260.0	515.0	> 0.1000	Yes	21.34%				
Wand	15	301.9	158.50	165.0	780.0	0.0895	Yes	52.51%				
86"Non	15	325.9	128.90	185.0	615.0	>0.1000	Yes	39.55%				
W/S	15	346.7	91.28	225.0	525.0	> 0.1000	Yes	26.33%				
B3Ep1		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 0.7 Pr > F 0.5484						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		7.3102	0.0069			Yes	14	1.20	0.2920	Yes	Neither	Neither
Glove vs 86"Non		4.0610	0.0439			Yes	14	0.29	0.6000	Yes	Neither	Neither
Glove vs W/S		0.6311	0.4270			Yes	14	0.01	0.6300	Yes	Neither	Neither
Wand vs 86"Non		0.5751	0.4482			Yes	14	0.17	0.6827	Yes	Neither	Neither
Wand vs W/S		3.9256	0.0476			Yes	14	1.82	0.1984	Yes	Neither	Neither
86"Non vs W/S		1.5782	0.2090			Yes	14	0.21	0.6534	Yes	Neither	Neither



**Figure H- 9: B3Ep2-1Off Experienced Users Pass 2-Icon 1 Offsets**

**Table H- 9: B3Ep2-1Off Experienced Users Pass 2-Icon 1 Offsets Statistics**

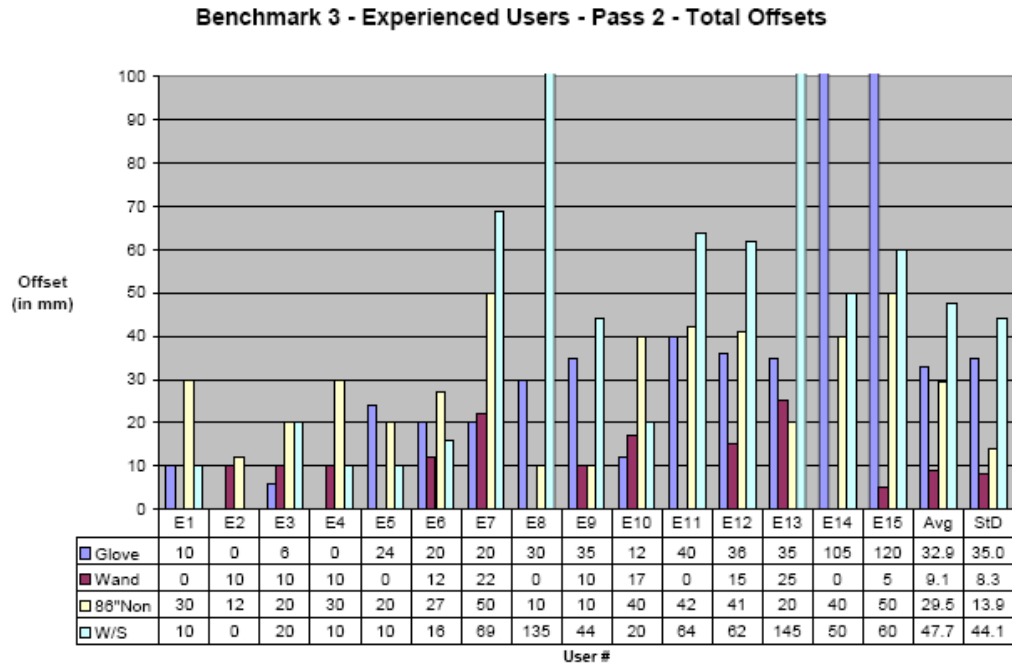
Benchmark 3, Experienced Users, Pass 2, Part 1 Offsets												
B3Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	22.67	27.97	0.00	95.00	0.0040	No	123.38%				
Wand	15	3.33	3.69	0.00	10.00	0.0015	No	110.74%				
86"Non	15	14.33	9.61	0.00	30.00	0.0879	Yes	67.08%				
W/S	15	26.60	29.26	0.00	110.00	> 0.1000	Yes	110.00%				
B3Ep2	Homogeneity of Variance					Roy's Greatest Root:			F(3,12)	7.17	Pr > F	0.0052
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			3.44	0.0725	Yes	14	6.36	0.0244	No	Means	Wand	580.17%
Glove vs 86"Non			2.29	0.1060	Yes	14	1.51	0.2399	Yes	Neither	Neither	
Glove vs W/S			0.01	0.9077	Yes	14	0.19	0.6680	Yes	Neither	Neither	
Wand vs 86"Non			8.2800	0.0076	No	14	18.02	0.0008	No	Var/Mean	Wand	329.94%
Wand vs W/S			3.0600	0.0911	Yes	14	8.82	0.0101	No	Means	Wand	698.08%
86"Non vs W/S	14.1292	0.0052			No	14	2.32	0.1497	Yes	Var	86"Non	85.62%



**Figure H- 10: B3Ep2-2Off Experienced Users Pass 2-Icon 2 Offsets**

**Table H- 10: B3Ep2-2Off Experienced Users Pass 2-Icon 2 Offsets Statistics**

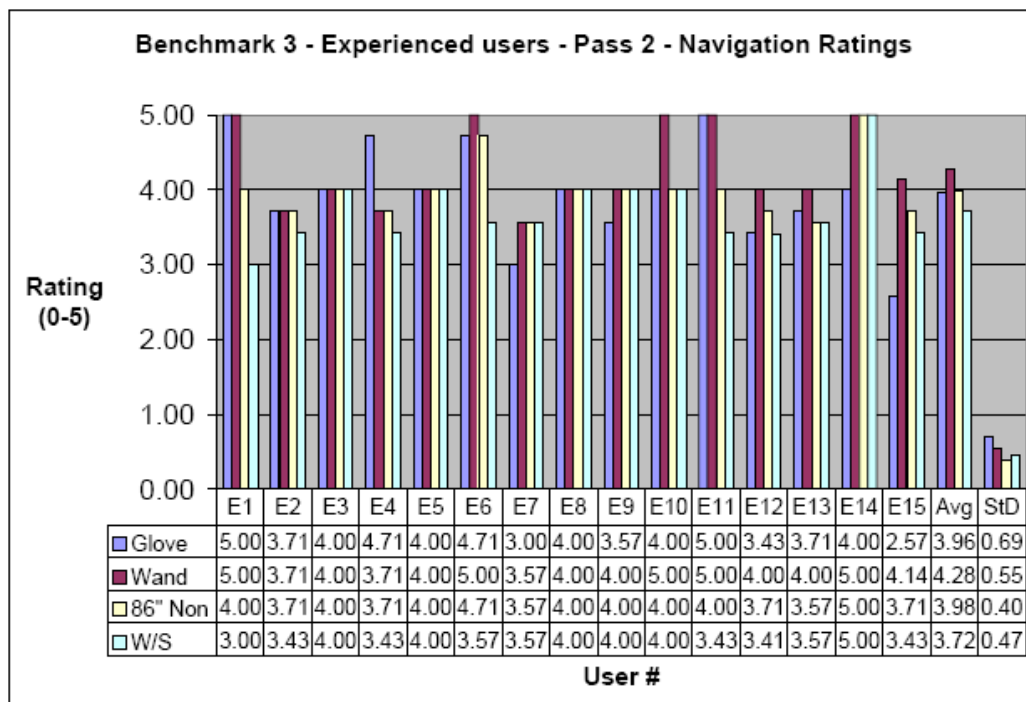
Benchmark 3, Experienced Users, Pass 2, Part 2 Offsets												
B3Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	10.20	8.03	0.00	25.00	> 0.1000	Yes	78.68%				
Wand	15	5.73	6.65	0.00	25.00	0.0029	No	115.98%				
86"Non	15	15.13	7.87	5.00	35.00	> 0.1000	Yes	52.04%				
W/S	15	21.07	19.11	0.00	70.00	0.0112	No	90.70%				
B3Ep2		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 7.17 Pr > F 0.0052						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				3.0900	0.0899	Yes	14	6.36	0.0244	No	Means	Wand 77.92%
Glove vs 86"Non		0.0002	0.9876			Yes	14	1.51	0.2399	Yes	Neither	Neither
Glove vs W/S				1.4500	0.1739	Yes	14	0.19	0.6680	Yes	Neither	Neither
Wand vs 86"Non				0.9600	0.3359	Yes	14	18.02	0.0008	No	Means	Wand 163.91%
Wand vs W/S				2.2200	0.1475	No	14	8.82	0.0101	No	Var/Means	Wand 267.52%
86"Non vs W/S				1.9400	0.1747	Yes	14	2.32	0.1497	Yes	Neither	Neither



**Figure H- 11: B3Ep2TotOff Experienced Users Pass 2 Total Offsets**

**Table H- 11: B3Ep2TotOff Experienced Users Pass 2 Total Offsets Statistics**

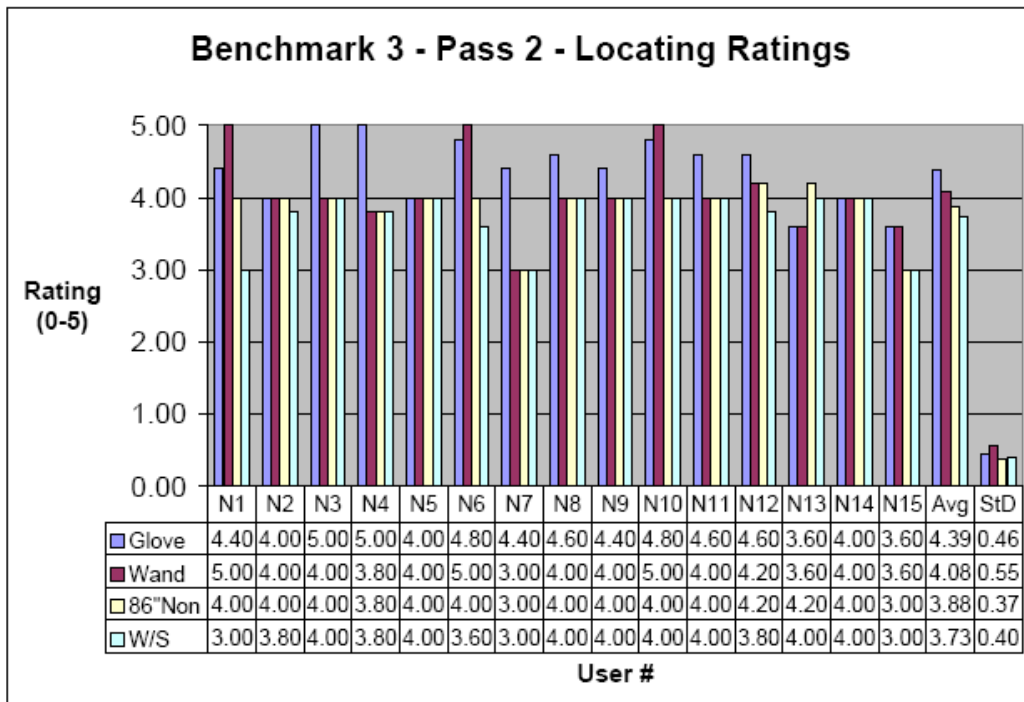
Benchmark 3, Experienced Users, Pass 2, Total Offsets												
B3Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	17.40	20.40	0.00	100.00	0.0006	No	117.24%				
Wand	15	8.83	10.29	0.00	40.00	0.0006	No	116.49%				
86"Non	15	72.23	176.50	0.00	850.00	< 0.0001	No	244.36%				
W/S	15	87.40	256.20	0.00	1400.00	< 0.0001	No	293.14%				
B3Ep2		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	3.2	Pr > F	0.0605
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.5400	0.4688	Yes	14	1.26	0.2808	Yes	Neither	Neither
Glove vs 86"Non				3.0900	0.0899	Yes	14	5.08	0.0408	No	Means	Glove 315.11%
Glove vs W/S				0.0000	0.9482	Yes	14	0.58	0.4607	Yes	Neither	Neither
Wand vs 86"Non				1.9500	0.1739	Yes	14	0.36	0.5560	Yes	Neither	Neither
Wand vs W/S				0.9600	0.3359	Yes	14	1.36	0.2638	Yes	Neither	Neither
86"Non vs W/S				1.9400	0.1747	Yes	14	7.22	0.0177	No	Means	86"Non 21.00%



**Figure H- 12: B3Ep2Nav Experienced Users Pass 2 Navigation Ratings**

**Table H- 12: B3Ep2Nav Experienced Users Pass 2 Navigation Ratings Statistics**

Benchmark 3, Experienced Users, Pass 2, Navigation Ratings												
B3Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	267.0	60.14	190.0	365.0	> 0.1000	Yes	22.53%				
Wand	15	191.5	47.37	115.0	300.0	0.0929	Yes	24.73%				
86"Non	15	237.3	70.78	150.0	440.0	>0.1000	Yes	29.82%				
W/S	15	263.7	72.24	165.0	445.0	> 0.1000	Yes	27.39%				
B3Ep2	Homogeneity of Variance					Roy's Greatest Root: F(3,12) 23 Pr > F <0.0001						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand	7.6330	0.3823			Yes	14	13.12	0.2800	Yes	Neither	Neither	
Glove vs 86"Non	0.3571	0.5501			Yes	14	1.52	0.2384	Yes	Neither	Neither	
Glove vs W/S	0.4514	0.5017			Yes	14	0.01	0.9090	Yes	Neither	Neither	
Wand vs 86"Non	2.1244	0.1450			Yes	14	6.19	0.0260	No	Means	Wand 23.92%	
Wand vs W/S	2.3389	0.1262			Yes	14	35.27	0.0000	No	Means	Wand 23.92%	
86"Non vs W/S	0.0056	0.9404			Yes	14	1.75	0.2070	Yes	Neither	Neither	

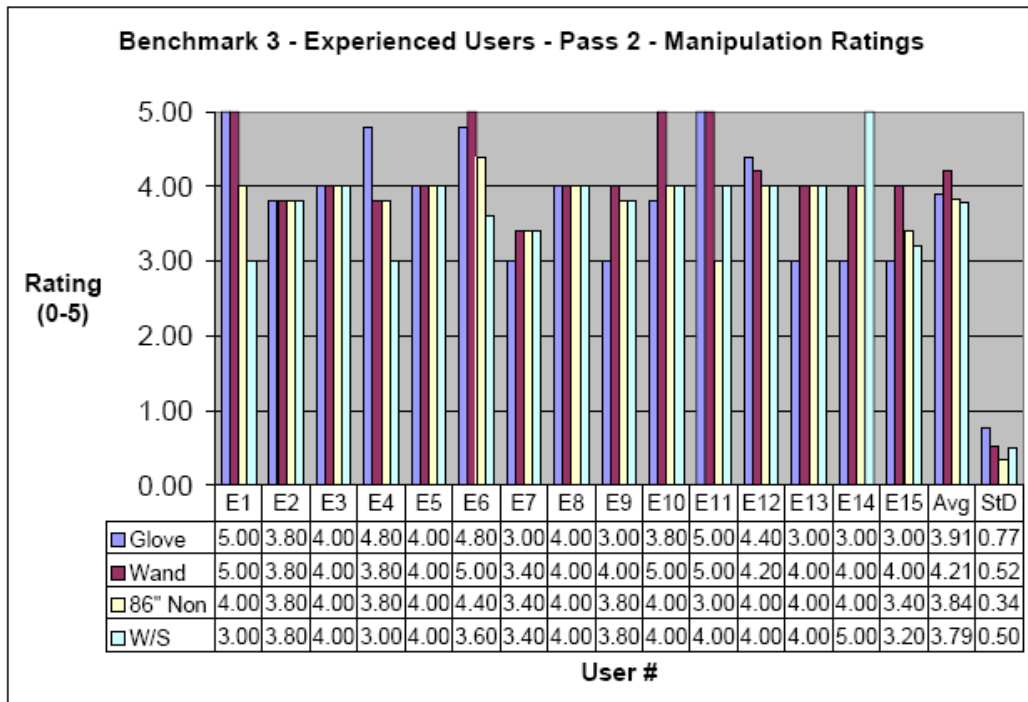


**Figure H- 13: B3Ep2Loc Experienced Users Pass 2 Locating Ratings**

**Table H- 13: B3Ep2Loc Experienced Users Pass 2 Locating Ratings Statistics**

Benchmark 3, Experienced Users, Pass 2, Locating Ratings												
B3Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	267.0	60.14	190.0	365.0	> 0.1000	Yes	22.53%				
Wand	15	191.5	47.37	115.0	300.0	0.0929	Yes	24.73%				
86"Non	15	237.3	70.78	150.0	440.0	>0.1000	Yes	29.82%				
W/S	15	263.7	72.24	165.0	445.0	> 0.1000	Yes	27.39%				
B3Ep2	Homogeneity of Variance				Roy's Greatest Root: F(3,12) 23 Pr > F <0.0001							
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand	7.6330	0.3823			Yes	14	13.12	0.2800	Yes	Neither	Neither	
Glove vs 86"Non	0.3571	0.5501			Yes	14	1.52	0.2384	Yes	Neither	Neither	
Glove vs W/S	0.4514	0.5017			Yes	14	0.01	0.9090	Yes	Neither	Neither	
Wand vs 86"Non	2.1244	0.1450			Yes	14	6.19	0.0260	No	Means	Wand 23.92%	
Wand vs W/S	2.3389	0.1262			Yes	14	35.27	0.0000	No	Means	Wand 23.92%	
86"Non vs W/S	0.0056	0.9404			Yes	14	1.75	0.2070	Yes	Neither	Neither	

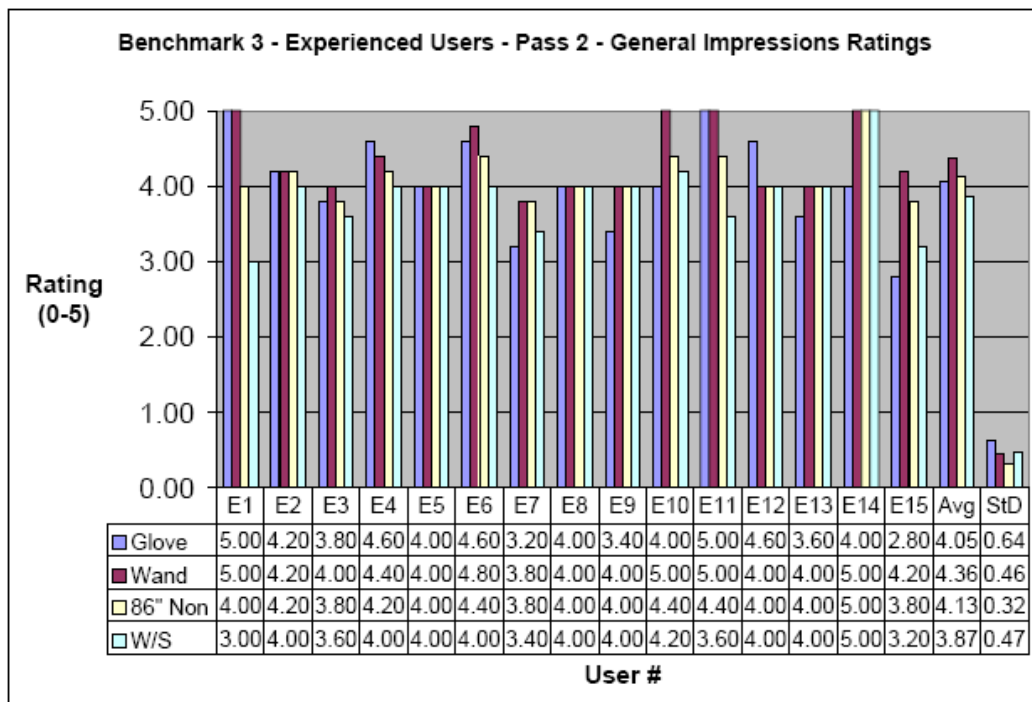




**Figure H- 14: B3Ep2Mov Experienced Users Pass 2 Manipulation Ratings**

**Table H- 14: B3Ep2Mov Experienced Users Pass 2 Manipulation Ratings Statistics**

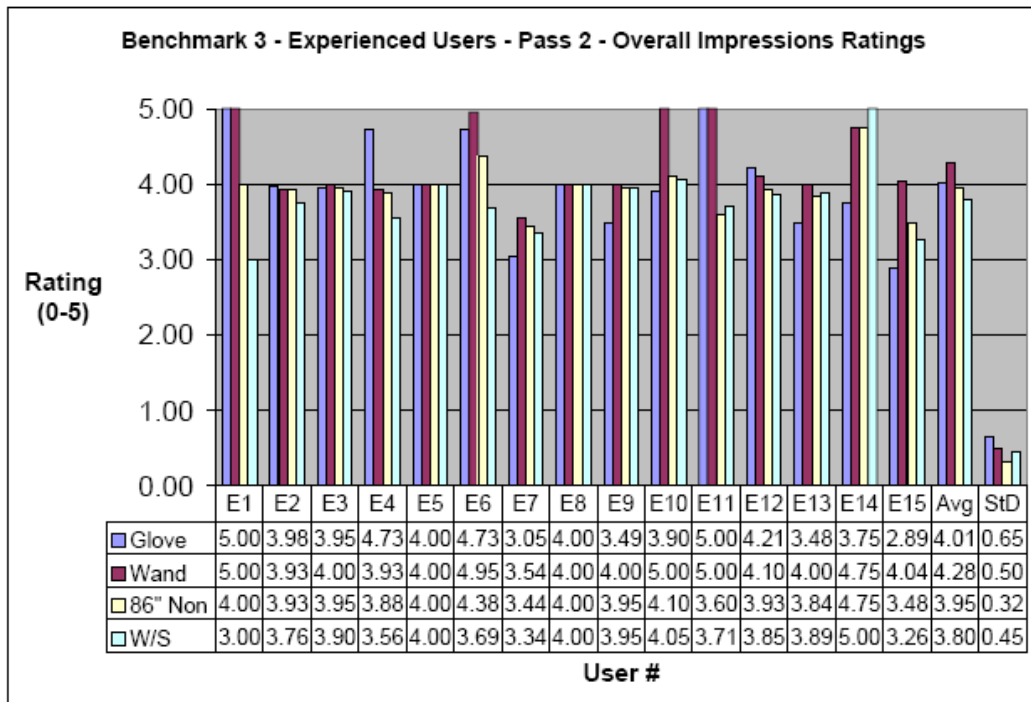
Benchmark 3, Experienced Users, Pass 2, Manipulation Ratings												
B3Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	267.0	60.14	190.0	365.0	> 0.1000	Yes	22.53%				
Wand	15	191.5	47.37	115.0	300.0	0.0929	Yes	24.73%				
86"Non	15	237.3	70.78	150.0	440.0	>0.1000	Yes	29.82%				
W/S	15	263.7	72.24	165.0	445.0	> 0.1000	Yes	27.39%				
B3Ep2		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	23	Pr > F	<0.0001
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		7.6330	0.3823			Yes	14	13.12	0.2800	Yes	Neither	Neither
Glove vs 86"Non		0.3571	0.5501			Yes	14	1.52	0.2384	Yes	Neither	Neither
Glove vs W/S		0.4514	0.5017			Yes	14	0.01	0.9090	Yes	Neither	Neither
Wand vs 86"Non		2.1244	0.1450			Yes	14	6.19	0.0260	No	Means	Wand 23.92%
Wand vs W/S		2.3389	0.1262			Yes	14	35.27	0.0000	No	Means	Wand 23.92%
86"Non vs W/S		0.0056	0.9404			Yes	14	1.75	0.2070	Yes	Neither	Neither



**Figure H- 15: B3Ep2Gen Experienced User Pass 2 General Impressions Ratings**

**Table H- 15: B3Ep2Gen Experienced User Pass 2 General Impressions Ratings Statistics**

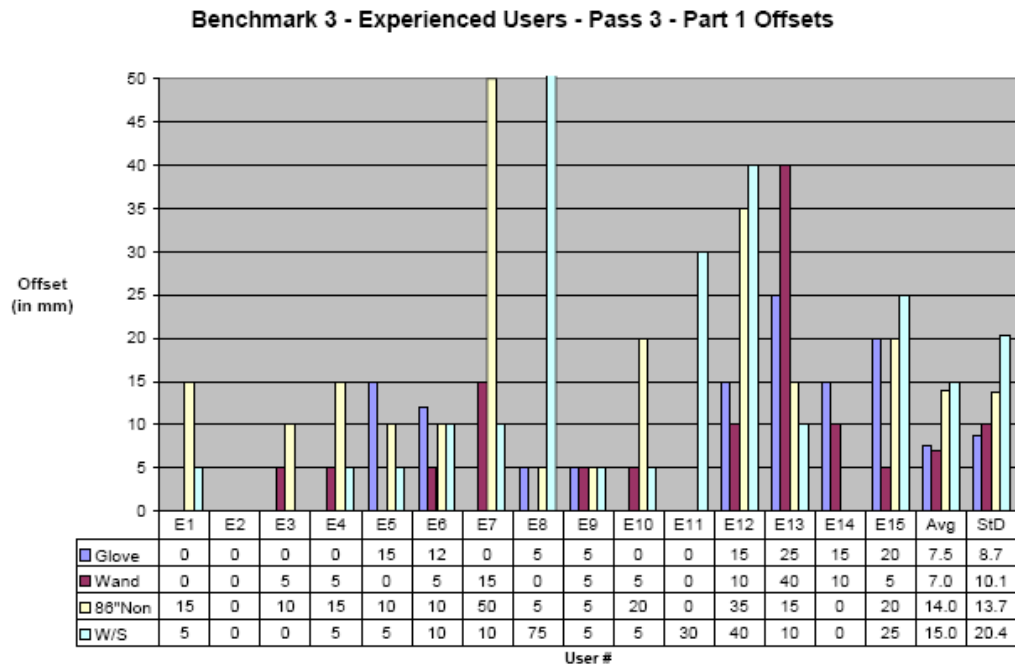
Benchmark 3, Experienced Users, Pass 2, General Impressions Ratings													
B3Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV					
Glove	15	267.0	60.14	190.0	365.0	> 0.1000	Yes	22.53%					
Wand	15	191.5	47.37	115.0	300.0	0.0929	Yes	24.73%					
86"Non	15	237.3	70.78	150.0	440.0	>0.1000	Yes	29.82%					
W/S	15	263.7	72.24	165.0	445.0	> 0.1000	Yes	27.39%					
B3Ep2		Homogeneity of Variance				Roy's Greatest Root:				F(3,12)	23	Pr > F	<0.0001
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand		7.6330	0.3823			Yes	14	13.12	0.2800	Yes	Neither	Neither	
Glove vs 86"Non		0.3571	0.5501			Yes	14	1.52	0.2384	Yes	Neither	Neither	
Glove vs W/S		0.4514	0.5017			Yes	14	0.01	0.9090	Yes	Neither	Neither	
Wand vs 86"Non		2.1244	0.1450			Yes	14	6.19	0.0260	No	Means	Wand 23.92%	
Wand vs W/S		2.3389	0.1262			Yes	14	35.27	0.0000	No	Means	Wand 23.92%	
86"Non vs W/S		0.0056	0.9404			Yes	14	1.75	0.2070	Yes	Neither	Neither	



**Figure H- 16: B3Ep2Ovr Experienced Users Pass 2 Overall Impressions Ratings**

**Table H- 16: B3Ep2Ovr Experienced Users Pass 2 Overall Impressions Ratings Statistics**

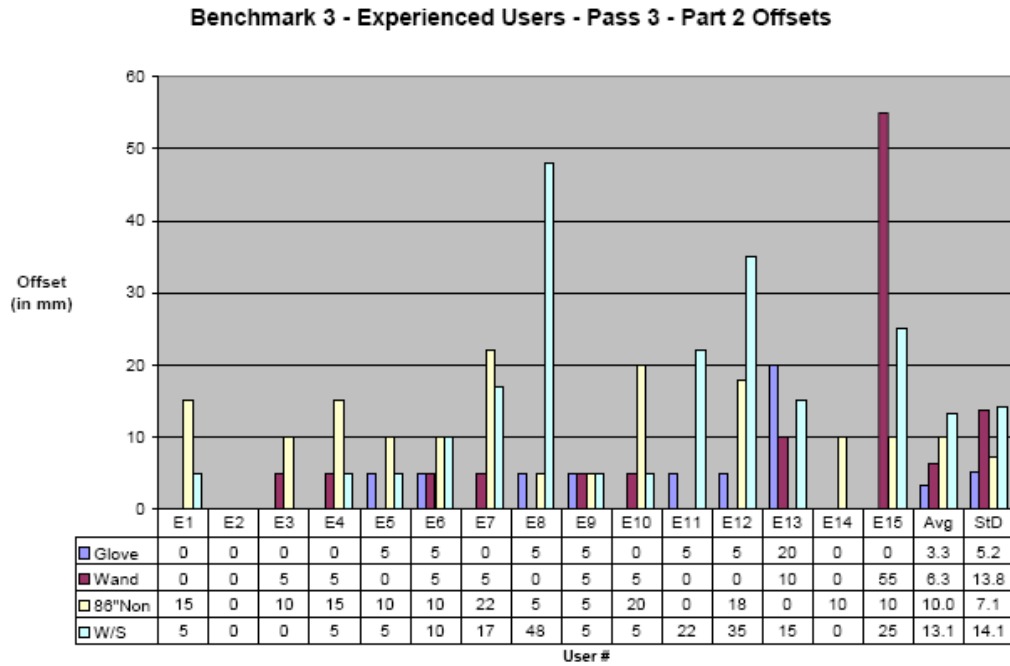
Benchmark 3, Experienced Users, Pass 2, Overall Impressions Ratings												
B3Ep2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	267.0	60.14	190.0	365.0	> 0.1000	Yes	22.53%				
Wand	15	191.5	47.37	115.0	300.0	0.0929	Yes	24.73%				
86"Non	15	237.3	70.78	150.0	440.0	>0.1000	Yes	29.82%				
W/S	15	263.7	72.24	165.0	445.0	> 0.1000	Yes	27.39%				
B3Ep2		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 23 Pr > F <0.0001						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		7.6330	0.3823			Yes	14	13.12	0.2800	Yes	Neither	Neither
Glove vs 86"Non		0.3571	0.5501			Yes	14	1.52	0.2384	Yes	Neither	Neither
Glove vs W/S		0.4514	0.5017			Yes	14	0.01	0.9090	Yes	Neither	Neither
Wand vs 86"Non		2.1244	0.1450			Yes	14	6.19	0.0260	No	Means	Wand 23.92%
Wand vs W/S		2.3389	0.1262			Yes	14	35.27	0.0000	No	Means	Wand 23.92%
86"Non vs W/S		0.0056	0.9404			Yes	14	1.75	0.2070	Yes	Neither	Neither



**Figure H- 17: B3Ep3-1Off Experienced Users Pass 3-Icon 1 Offsets**

**Table H- 17: B3Ep3-1Off Experienced Users Pass 3-Icon 1 Offsets Statistics**

Benchmark 3, Experienced Users, Pass 3, Part 1 Offsets												
B3Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	7.47	8.69	0.00	25.00	0.0040	No	116.41%				
Wand	15	7.00	10.14	0.00	40.00	0.0004	No	144.86%				
86*Non	15	14.00	13.65	0.00	50.00	0.0929	Yes	97.50%				
W/S	15	15.00	20.35	0.00	75.00	0.0001	No	135.67%				
B3Ep3		Homogeneity of Variance				Roy's Greatest Root:			F(3,12)	1.26	Pr > F	0.3317
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.1200	0.7321	Yes	14	0.04	0.8411	Yes	Neither	Neither
Glove vs 86*Non				1.3800	0.2498	Yes	14	2.52	0.1351	Yes	Neither	Neither
Glove vs W/S				1.8300	0.1866	Yes	14	1.88	0.1915	Yes	Neither	Neither
Wand vs 86*Non				0.4900	0.4904	Yes	14	3.69	0.0754	No	Means	Wand 100.00%
Wand vs W/S				1.4300	0.2423	Yes	14	1.68	0.2156	Yes	Neither	Neither
86*Non vs W/S				1.4300	0.2433	Yes	14	0.03	0.8737	Yes	Neither	Neither



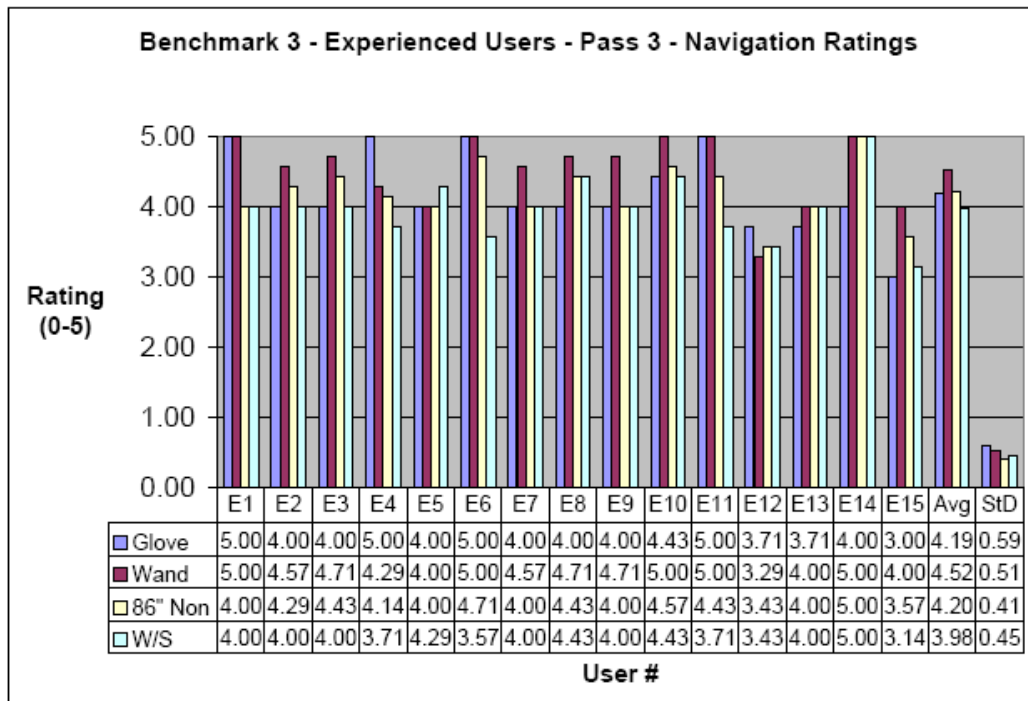
**Figure H- 18: B3Ep3-2Off Experienced Users Pass 3-Icon 2 Offsets**

**Table H- 18: B3Ep3-2Off Experienced Users Pass 3-Icon 2 Offsets Statistics**

Benchmark 3, Experienced Users, Pass 3, Part 2 Offsets												
B3Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	3.33	5.23	0.00	20.00	0.0004	No	157.01%				
Wand	15	6.33	13.82	0.00	55.00	< 0.0001	No	218.22%				
86*Non	15	10.00	7.11	0.00	22.00	> 0.1000	Yes	71.11%				
W/S	15	13.13	14.12	0.00	48.00	0.0117	No	107.54%				
B3Ep3		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 1.26 Pr > F 0.3317						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				1.0400	0.3175	Yes	14	0.04	0.8411	Yes	Neither	Neither
Glove vs 86*Non				1.1700	0.2892	Yes	14	2.52	0.1351	Yes	Neither	Neither
Glove vs W/S				1.4200	0.2427	Yes	14	0.88	0.1915	Yes	Neither	Neither
Wand vs 86*Non				0.1100	0.7431	Yes	14	3.69	0.0754	No	Means	Wand 57.90%
Wand vs W/S				0.9200	0.3445	Yes	14	1.68	0.2156	Yes	Neither	Neither
86*Non vs W/S				0.6700	0.4189	Yes	14	0.03	0.9737	Yes	Neither	Neither

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14	E15	Avg	Std
Glove	0	0	0	0	20	17	0	10	10	0	5	20	45	15	20	10.8	12.6
Wand	0	0	10	10	0	10	20	0	10	10	0	10	50	10	60	13.3	18.0
88"Non	30	0	20	30	20	20	72	10	10	40	0	53	15	10	30	24.0	19.6
W/S	10	0	0	10	10	20	27	123	10	10	52	75	25	0	50	28.1	34.2

Benchmark 3, Experienced Users, Pass 3, Total Offsets												
B3Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	9.00	17.88	0.00	90.00	< 0.0001	No	198.67%				
Wand	15	6.20	10.59	0.00	55.00	< 0.0001	No	170.81%				
86"Non	15	56.83	71.74	0.00	280.00	< 0.0001	No	126.24%				
W/S	15	55.40	66.42	0.00	325.00	0.0030	No	119.89%				
B3Ep3	Homogeneity of Variance					Roy's Greatest Root:			F(3,12)	11	Pr > F	0.001
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			1.0400	0.3175	Yes	14	6.41	0.0239	No	Means	Wand	31.11%
Glove vs 86"Non			1.1700	0.2892	Yes	14	14.59	0.0019	No	Means	Glove	531.44%
Glove vs W/S			1.4200	0.2427	Yes	14	13.67	0.0022	No	Means	Glove	515.56%
Wand vs 86"Non			0.1100	0.7432	Yes	14	0.00	0.9759	Yes	Neither	Neither	
Wand vs W/S			0.9200	0.3445	Yes	14	1.84	0.1367	Yes	Neither	Neither	
86"Non vs W/S			0.6700	0.4189	Yes	14	3.22	0.0942	No	Means	W/S	2.58%



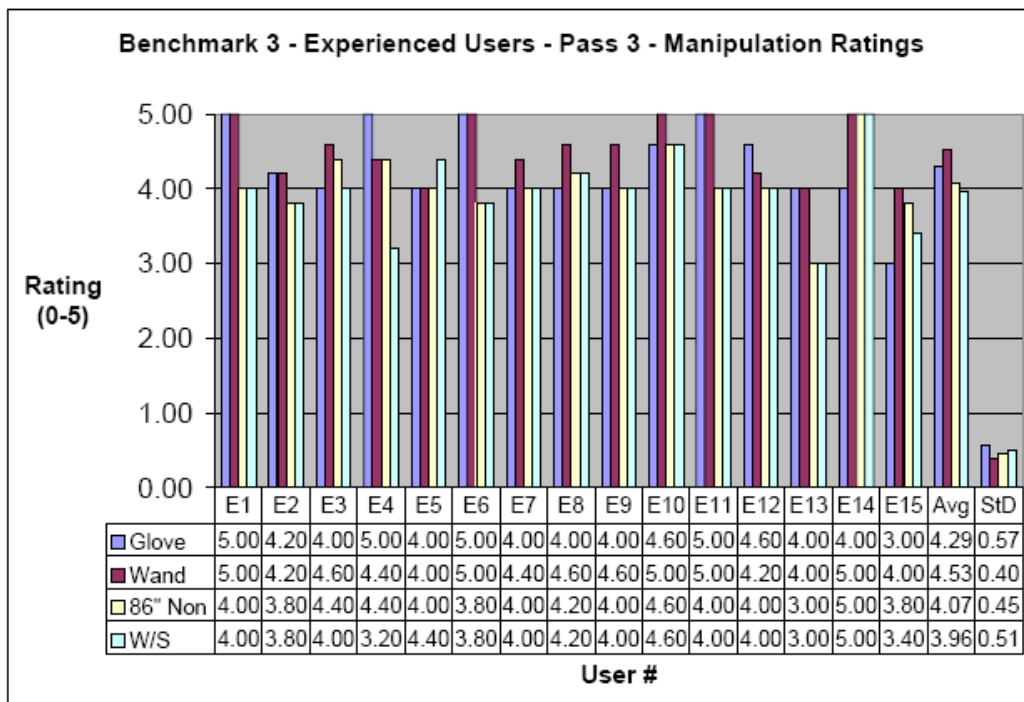
**Figure H- 20: B3Ep3Nav Experienced Users Pass 3 Navigation Ratings**

**Table H- 20: B3Ep3Nav Experienced Users Pass 3 Navigation Ratings Statistics**

Benchmark 3, Experienced Users, Pass 3, Navigation Ratings													
B3Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV					
Glove	15	212.3	30.64	175.0	270.0	> 0.1000	Yes	14.43%					
Wand	15	171.9	37.11	115.0	245.0	>0.1000	Yes	21.58%					
86"Non	15	194.5	48.14	95.0	260.0	>0.1000	Yes	24.75%					
W/S	15	196.4	36.80	125.0	250.0	> 0.1000	Yes	18.83%					
B3Ep3		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 4.4 Pr > F 0.0186							
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand		0.4924	0.4828			Yes	14	10.23	0.0064	No	Means	Wand	23.50%
Glove vs 86"Non		2.7107	0.0997			Yes	14	1.75	0.2069	Yes	Neither	Neither	
Glove vs W/S		0.4509	0.5019			Yes	14	1.17	0.2969	Yes	Neither	Neither	
Wand vs 86"Non		0.9305	0.3347			Yes	14	3.76	0.0729	No	Means	Wand	13.15%
Wand vs W/S		0.0009	0.9757			Yes	14	6.72	0.0213	No	Means	Wand	13.15%
86"Non vs W/S		0.9893	0.3199			Yes	14	6.72	0.0213	No	Means	86"Non	0.98%



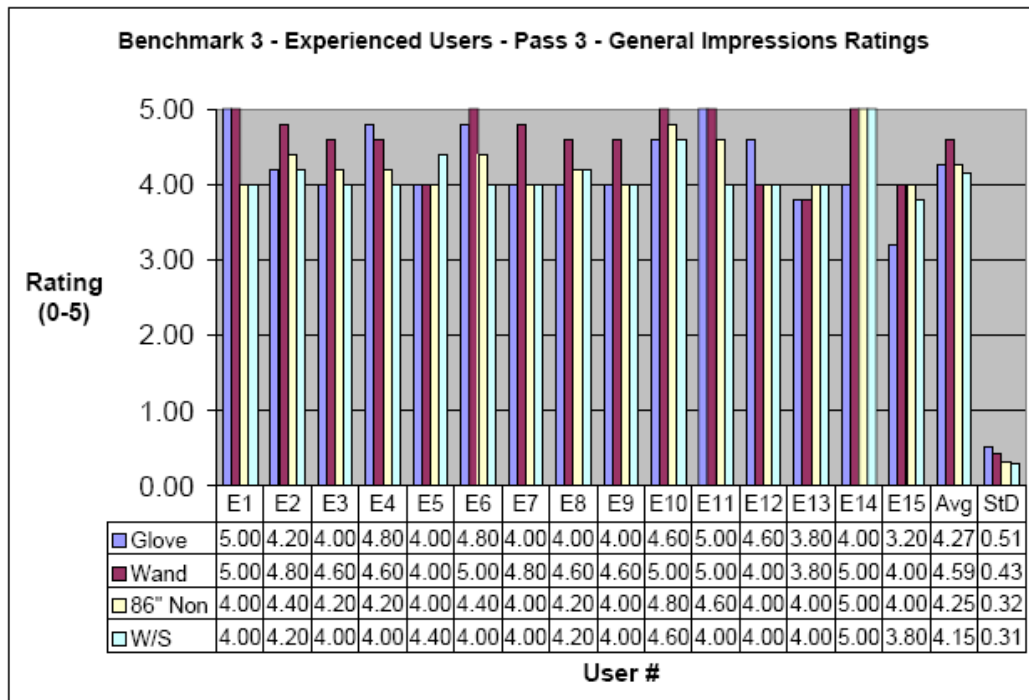




**Figure H- 22: B3Ep3Mov Experienced Users Pass 3 Manipulation Ratings**

**Table H- 22: B3Ep3Mov Experienced Users Pass 3 Manipulation Ratings Statistics**

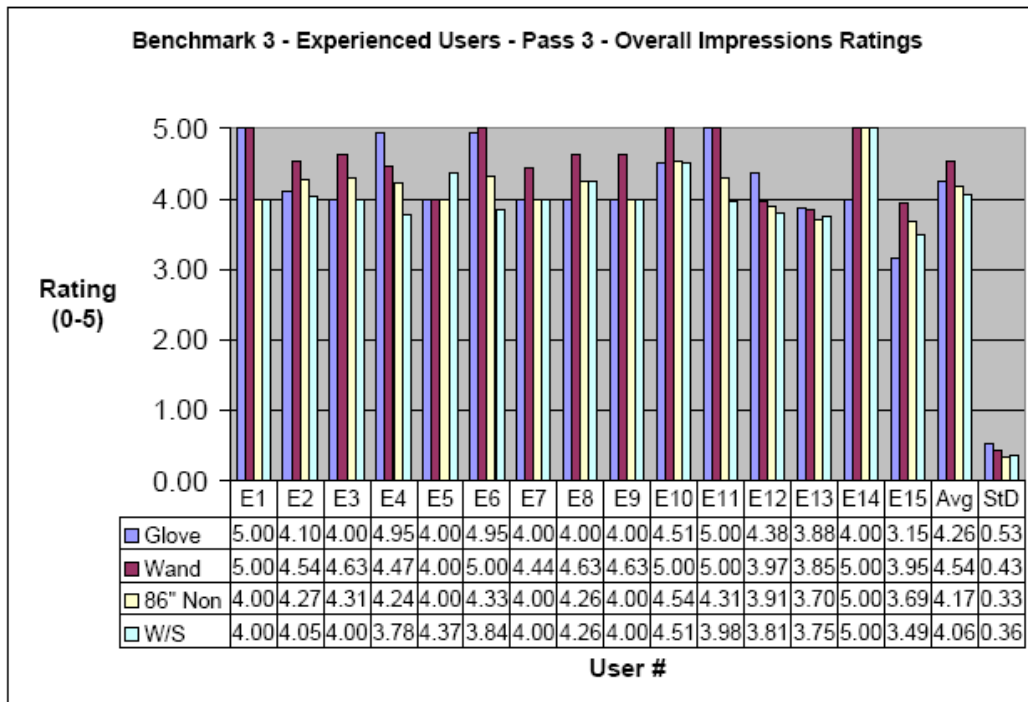
Benchmark 3, Experienced Users, Pass 3, Manipulation Ratings											
B3Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	212.3	30.64	175.0	270.0	> 0.1000	Yes	14.43%			
Wand	15	171.9	37.11	115.0	245.0	>0.1000	Yes	21.58%			
86"Non	15	194.5	48.14	95.0	260.0	>0.1000	Yes	24.75%			
W/S	15	196.4	36.80	125.0	250.0	> 0.1000	Yes	18.83%			
B3Ep3	Homogeneity of Variance				Roy's Greatest Root: F(3,12) 4.39 Pr > F 0.0186						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	0.4924	0.4828			Yes	14	10.23	0.0064	No	Means	Wand 23.50%
Glove vs 86"Non	2.7107	0.0997			Yes	14	1.75	0.2069	Yes	Neither	Neither
Glove vs W/S	0.4509	0.5019			Yes	14	1.17	0.2969	Yes	Neither	Neither
Wand vs 86"Non	0.9305	0.3347			Yes	14	3.76	0.0729	No	Means	Wand 13.15%
Wand vs W/S	0.0009	0.9757			Yes	14	6.72	0.0213	No	Means	Wand 13.15%
86"Non vs W/S	0.9893	0.3199			Yes	14	6.72	0.0213	No	Means	86"Non 0.98%



**Figure H- 23: B3Ep3Gen Experienced Users Pass 3 General Impressions Ratings**

**Table H- 23: B3Ep3Gen Experienced Users Pass 3 General Impressions Ratings Statistics**

Benchmark 3, Experienced Users, Pass 3, General Impressions Ratings												
B3Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	212.3	30.64	175.0	270.0	> 0.1000	Yes	14.43%				
Wand	15	171.9	37.11	115.0	245.0	>0.1000	Yes	21.58%				
86"Non	15	194.5	48.14	95.0	260.0	>0.1000	Yes	24.75%				
W/S	15	196.4	36.80	125.0	250.0	> 0.1000	Yes	18.83%				
B3Ep3		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 4.39 Pr > F 0.0186						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		0.4924	0.4828			Yes	14	10.23	0.0064	No	Means	Wand 23.50%
Glove vs 86"Non		2.7107	0.0997			Yes	14	1.75	0.2069	Yes	Neither	Neither
Glove vs W/S		0.4509	0.5019			Yes	14	1.17	0.2969	Yes	Neither	Neither
Wand vs 86"Non		0.9305	0.3347			Yes	14	3.76	0.0729	No	Means	Wand 13.15%
Wand vs W/S		0.0009	0.9757			Yes	14	6.72	0.0213	No	Means	Wand 13.15%
86"Non vs W/S		0.9893	0.3199			Yes	14	6.72	0.0213	No	Means	86"Non 0.98%

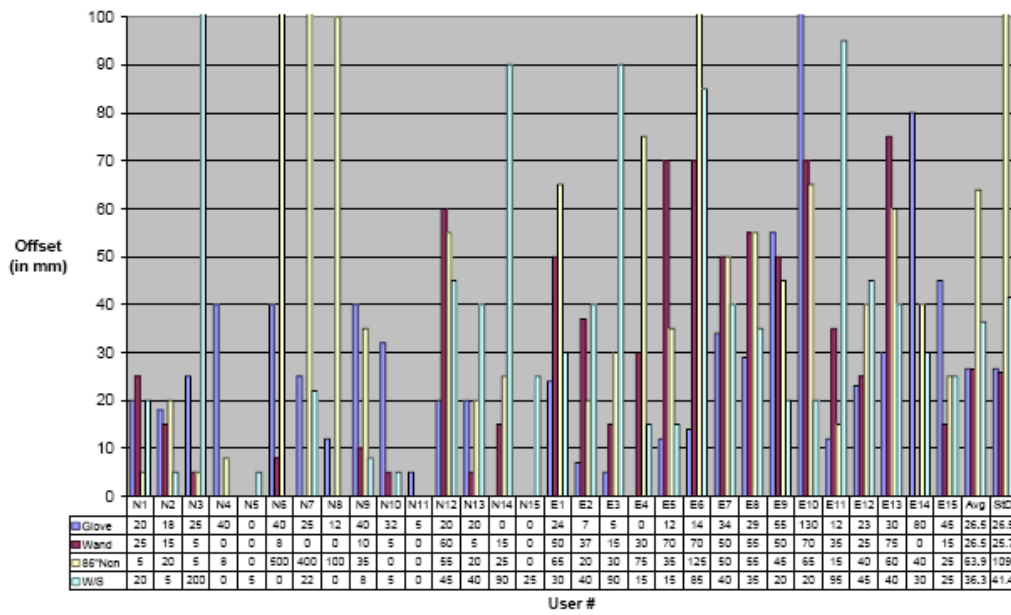


**Figure H- 24: B3Ep3Ovr Experienced Users Pass 3 Overall Impressions Ratings**

**Table H- 24: B3Ep3Ovr Experienced Users Pass 3 Overall Impressions Ratings Statistics**

Benchmark 3, Experienced Users, Pass 3, Overall Impressions Ratings												
B3Ep3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	212.3	30.64	175.0	270.0	> 0.1000	Yes	14.43%				
Wand	15	171.9	37.11	115.0	245.0	>0.1000	Yes	21.58%				
86"Non	15	194.5	48.14	95.0	260.0	>0.1000	Yes	24.75%				
W/S	15	196.4	36.80	125.0	250.0	> 0.1000	Yes	18.83%				
B3Ep3		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 4.4 Pr > F 0.0186						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		0.4924	0.4828			Yes	14	10.23	0.0064	No	Means	Wand 23.50%
Glove vs 86"Non		2.7107	0.0997			Yes	14	1.75	0.2069	Yes	Neither	Neither
Glove vs W/S		0.4509	0.5019			Yes	14	1.17	0.2969	Yes	Neither	Neither
Wand vs 86"Non		0.9305	0.3347			Yes	14	3.76	0.0729	No	Means	Wand 13.15%
Wand vs W/S		0.0009	0.9757			Yes	14	6.72	0.0213	No	Means	Wand 13.15%
86"Non vs W/S		0.9893	0.3199			Yes	14	6.72	0.0213	No	Means	86"Non 0.98%

**Benchmark 3 - All Users - Pass 1 - Part 1 Offsets**

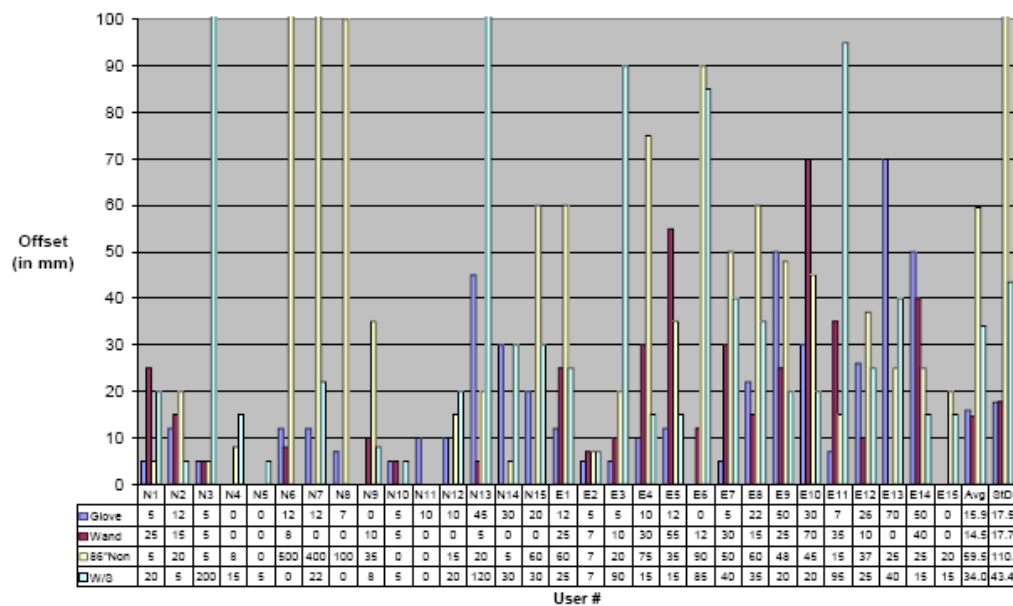


**Figure H- 25: B3E3pA-1Off Experienced Users 3 Pass Avg. Icon 1 Offsets**

**Table H- 25: B3E3pA-1Off Experienced Users 3 Pass Avg. Icon 1 Offsets Statistics**

Benchmark 3, Experienced Users, 3 Pass Avg., Part 1 Offsets										
B3E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	15	20.76	14.80	0.00	47.33	> 0.1000	Yes	71.29%		
Wand	15	17.82	9.14	3.33	38.33	> 0.1000	Yes	51.31%		
86\"Non	15	26.00	11.25	6.67	48.33	> 0.1000	Yes	43.27%		
W/S	15	27.76	17.15	8.33	58.33	> 0.1000	Yes	61.78%		
B3E3pA		Homogeneity of Variance			Roy's Greatest Root:			F(3,12) 3.60	Pr > F	0.0460
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					Significant?
Glove vs Wand		3.0241	0.0820			Yes	14	0.46	0.5106	Yes
Glove vs 86\"Non		1.0145	0.3128			Yes	14	1.20	0.2924	Yes
Glove vs W/S		29.4000	0.5893			Yes	14	1.83	0.1971	Yes
Wand vs 86\"Non		0.5706	0.4507			Yes	14	8.37	0.0118	No
Wand vs W/S		5.0277	0.0249			No	14	4.31	0.0521	No
86\"Non vs W/S		2.3472	0.1255			Yes	14	0.11	0.7481	Yes
										Statistically Better
										Wand 45.90%
										Wand 55.78%
										Neither

**Benchmark 3 - All Users - Pass 1 - Part 2 Offsets**

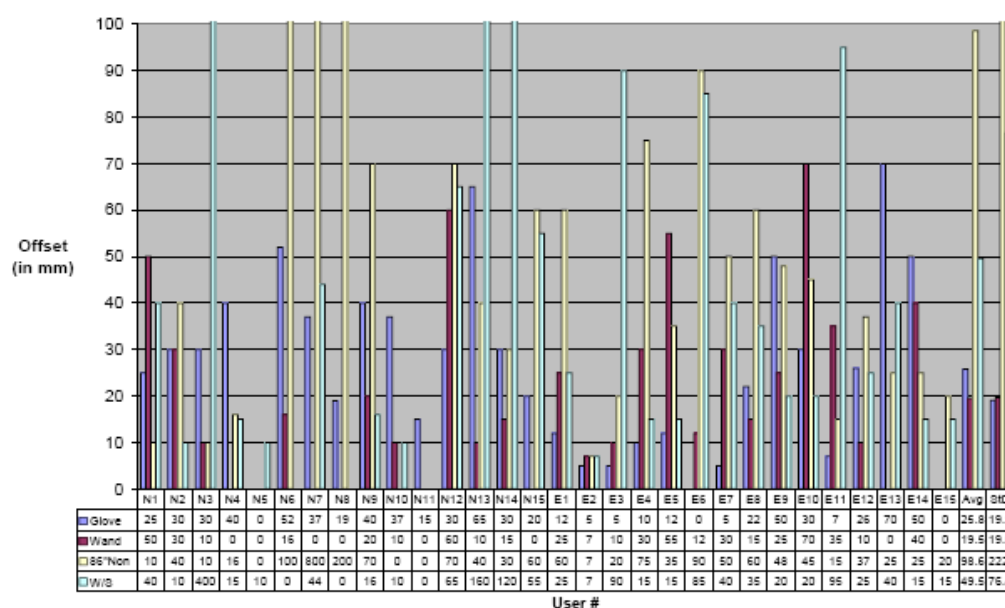


**Figure H- 26: B3E3pA-2of Experienced Users 3 Pass Avg. Icon 2 Offsets**

**Table H- 26: B3E3pA-2of Experienced Users 3 Pass Avg. Icon 2 Offsets Statistics**

Benchmark 3, Experienced Users, 3 Pass Avg., Part 2 Offsets													
B3E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV					
Glove	15	11.27	9.02	1.67	31.67	0.0442	No	80.05%					
Wand	15	12.11	6.71	4.00	29.00	> 0.1000	Yes	55.40%					
86"Non	15	21.98	9.63	6.33	39.00	> 0.1000	Yes	43.81%					
W/S	15	23.44	15.21	2.33	51.00	> 0.1000	Yes	64.89%					
B3E3pA		Homogeneity of Variance				Roy's Greatest Root:				F(3,12)	3.6	Pr > F	0.046
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand				0.9700	0.3322	Yes	14	0.46	0.5106	Yes	Neither	Neither	
Glove vs 86"Non				2.4000	0.1328	Yes	14	1.20	0.2924	Yes	Neither	Neither	
Glove vs W/S				2.5200	0.1234	Yes	14	1.83	0.1971	Yes	Neither	Neither	
Wand vs 86"Non		8.9386	0.0028			No	14	8.37	0.0118	No	Var/Mean	Wand 81.50%	
Wand vs W/S		13.4373	0.0002			No	14	4.51	0.0521	No	Var/Mean	Wand 93.56%	
86"Non vs W/S		0.6647	0.4149			Yes	14	0.11	0.7481	Yes	Neither	Neither	

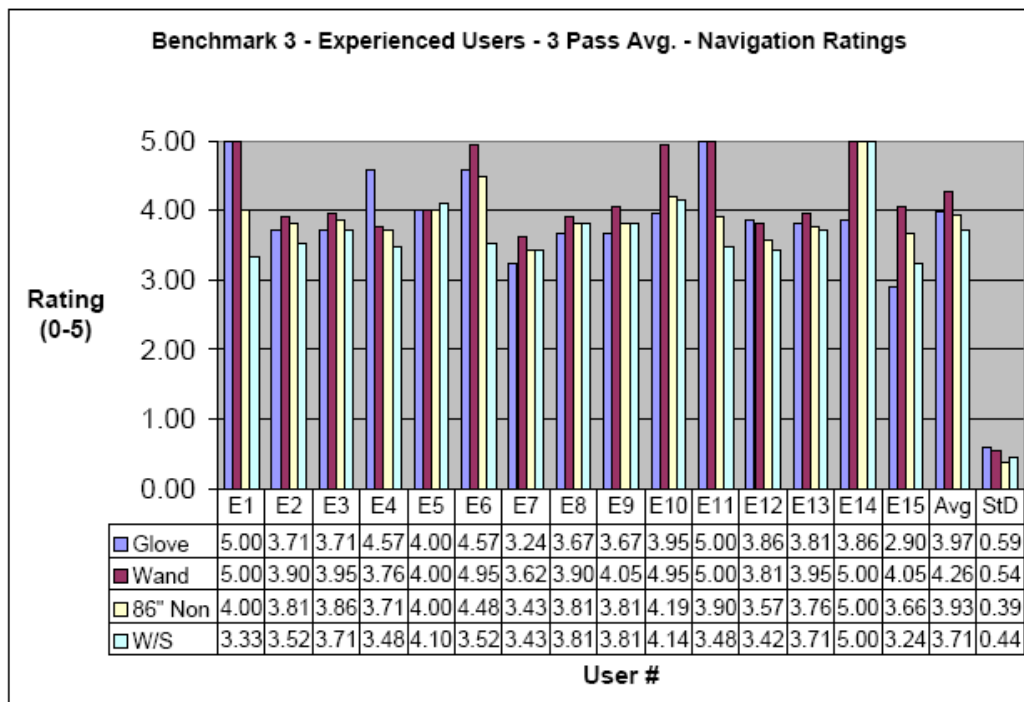
**Benchmark 3 - All Users - Pass 1 - Total Offsets**



**Figure H- 27: B3E3pAtotOff Experienced Users 3 Pass Avg. Total Offsets**

**Table H- 27: B3E3pAtotOff Experienced Users 3 Pass Avg. Total Offsets Statistics**

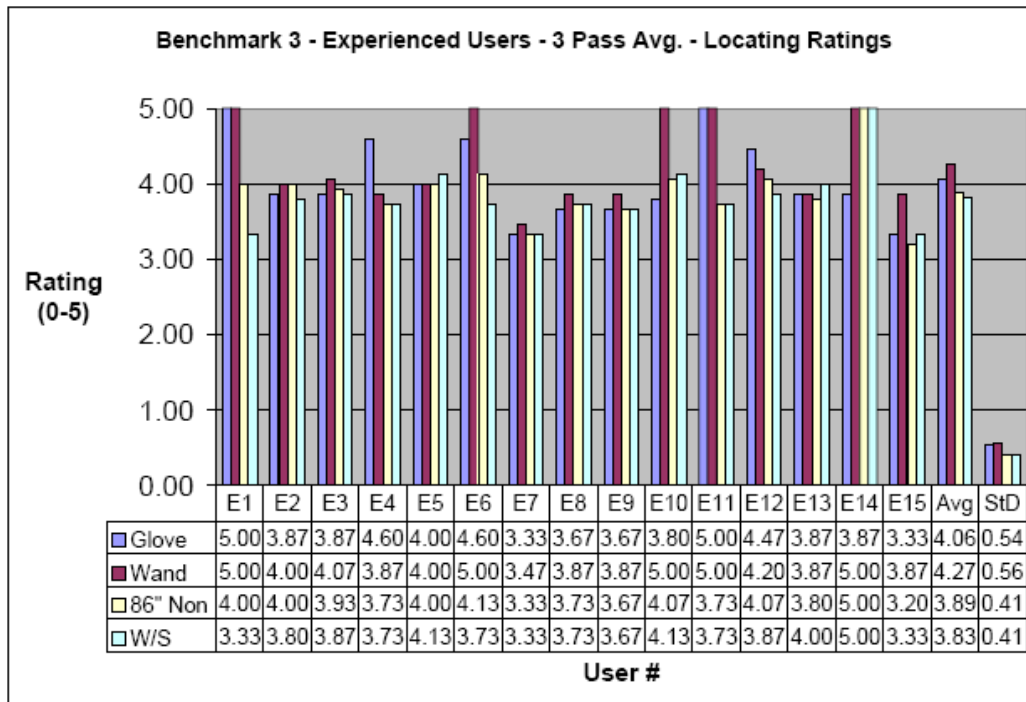
Benchmark 3, Experienced Users, 3 Pass Avg., Total Offsets												
B3E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	20.43	14.86	0.00	85.00	0.0188	No	72.74%				
Wand	15	15.03	9.03	0.00	23.33	< 0.0001	No	60.05%				
86"Non	15	84.65	168.50	10.00	826.70	< 0.0001	No	199.05%				
W/S	15	80.57	175.70	2.33	975.00	< 0.0001	No	218.07%				
B3E3pA		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 3.6 Pr > F 0.0466						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.9700	0.3322	Yes	14	2.05	0.1737	Yes	Neither	
Glove vs 86"Non				2.4000	0.1320	Yes	14	6.94	0.0196	No	Means	Glove 314.34%
Glove vs W/S				2.5200	0.1234	Yes	14	0.88	0.3653	Yes	Neither	Neither
Wand vs 86"Non				3.7200	0.0640	Yes	14	0.33	0.5744	Yes	Neither	Neither
Wand vs W/S				3.1900	0.0849	Yes	14	2.44	0.1409	Yes	Neither	Neither
86"Non vs W/S				0.4100	0.5281	Yes	14	8.59	0.0110	No	Means	W/S 5.06%



**Figure H- 28: B3E3pANav Experienced Users 3 Pass Avg. Navigation Ratings**

**Table H- 28: B3E3pANav Experienced Users 3 Pass Avg. Navigation Ratings Statistics**

Benchmark 3, Experienced Users, 3 Pass Avg., Navigation Ratings										
B3E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	15	274.5	44.01	213.3	348.3	> 0.1000	Yes	16.03%		
Wand	15	221.8	69.17	153.3	403.3	>0.1000	Yes	31.19%		
86\" Non	15	252.6	69.45	167.7	436.7	0.0135	No	27.49%		
W/S	15	268.6	54.36	191.7	376.7	> 0.1000	Yes	20.24%		
B3E3pA		Homogeneity of Variance				Roy's Greatest Root:			F(3,12) 6.6 Pr > F 0.0071	
		Bartlett's Test		Levene's Test		df	F(1,14)	Pr > F	Equal Means ?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					
Glove vs Wand		2.6742	0.1020			Yes	14	8.30	0.0120	No
Glove vs 86\" Non				1.4300	0.2413	Yes	14	1.23	0.2860	Yes
Glove vs W/S		0.5987	0.4391			Yes	14	0.10	0.7610	Yes
Wand vs 86\" Non				0.0001	0.9909	Yes	14	1.32	0.2690	Yes
Wand vs W/S		0.7771	0.3780			Yes	14	11.81	0.0040	No
86\" Non vs W/S				0.5400	0.4683	Yes	14	0.51	0.4900	Yes
										Neither
										Neither
										Neither
										Wand 23.76%
										Neither
										Neither
										Wand 13.89%
										Neither

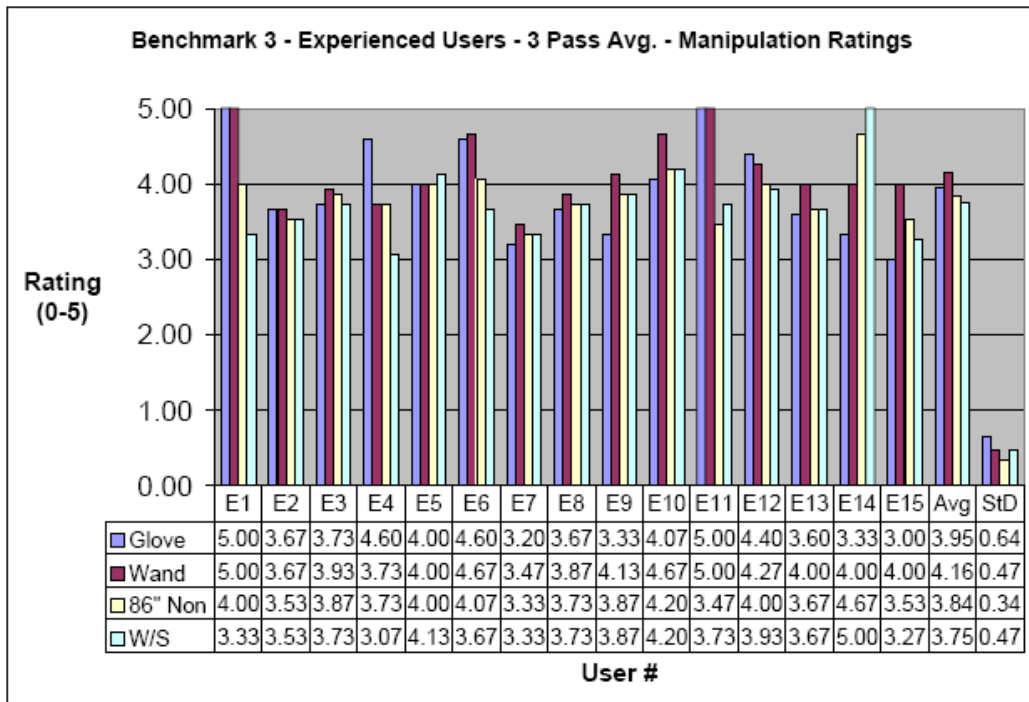


**Figure H- 29: B3E3pALoc Experienced Users 3 Pass Avg. Locating Ratings**

**Table H- 29: B3E3pALoc Experienced Users 3 Pass Avg. Locating Ratings Statistics**

Benchmark 3, Experienced Users, 3 Pass Avg., Locating Ratings												
B3E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	15	274.5	44.01	213.3	348.3	> 0.1000	Yes	16.03%				
Wand	15	221.8	69.17	153.3	403.3	>0.1000	Yes	31.19%				
86\" Non	15	252.6	69.45	167.7	436.7	0.0135	No	27.49%				
W/S	15	268.6	54.36	191.7	376.7	> 0.1000	Yes	20.24%				
B3E3pA		Homogeneity of Variance				Roy's Greatest Root:				F(3,12) 6.57 Pr > F 0.0071		
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		2.6742	0.1020			Yes	14	8.30	0.0120	No	Means	Wand 23.76%
Glove vs 86\" Non				1.4300	0.2413	Yes	14	1.23	0.2860	Yes	Neither	Neither
Glove vs W/S		0.5987	0.4391			Yes	14	0.10	0.7610	Yes	Neither	Neither
Wand vs 86\" Non				0.0001	0.9909	Yes	14	1.32	0.2690	Yes	Neither	Wand 13.89%
Wand vs W/S		0.7771	0.3780			Yes	14	11.81	0.0040	No	Means	Wand 13.89%
86\" Non vs W/S				0.5400	0.4683	Yes	14	0.51	0.4900	Yes	Neither	Neither

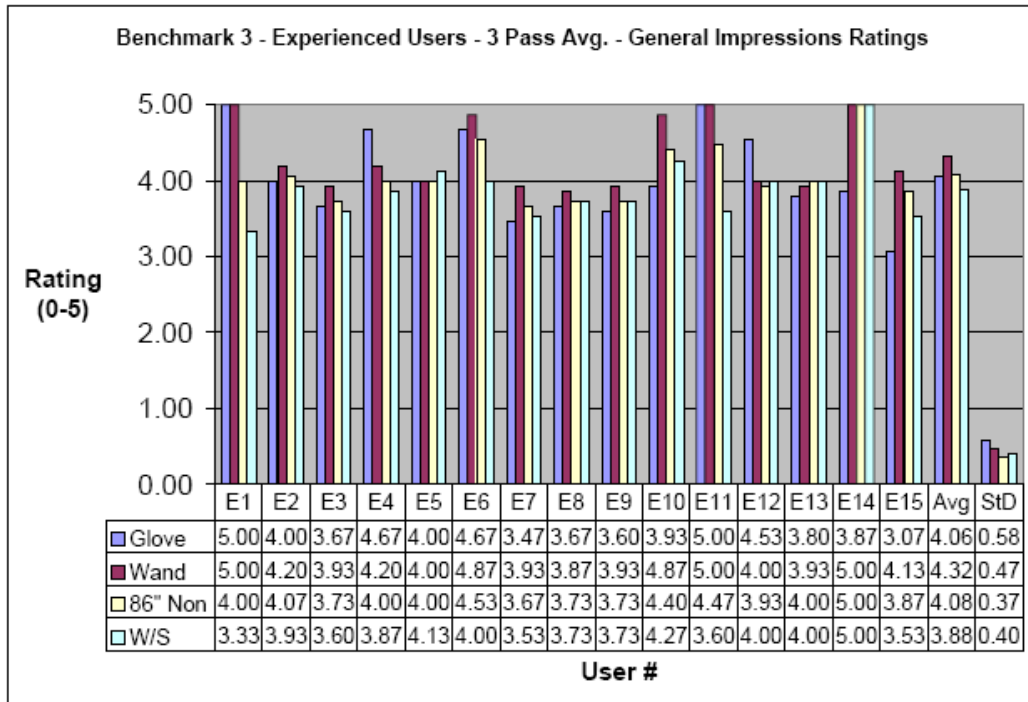




**Figure H- 30: B3E3pAMov Experienced Users 3 Pass Avg. Manipulation Ratings**

**Table H- 30: B3E3pAMov Experienced Users 3 Pass Avg. Manipulation Ratings Statistics**

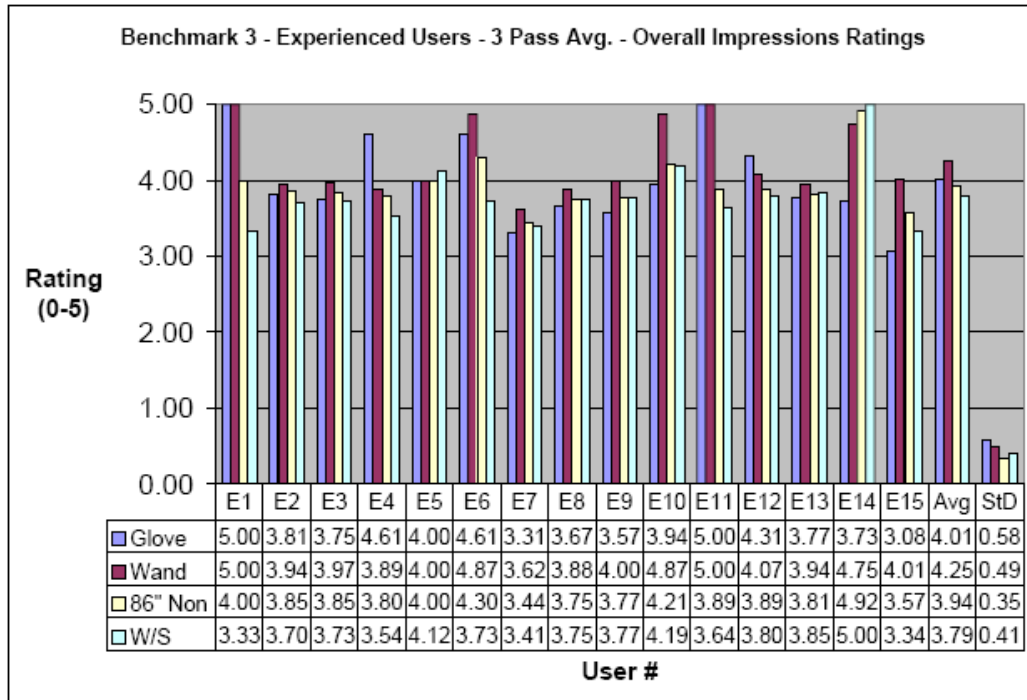
Benchmark 3, Experienced Users, 3 Pass Avg., Manipulation Ratings											
B3E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	15	274.5	44.01	213.3	348.3	> 0.1000	Yes	16.03%			
Wand	15	221.8	69.17	153.3	403.3	>0.1000	Yes	31.19%			
86"Non	15	252.6	69.45	167.7	436.7	0.0135	No	27.49%			
W/S	15	268.6	54.36	191.7	376.7	> 0.1000	Yes	20.24%			
B3E3pA	Homogeneity of Variance				Roy's Greatest Root: F(3,12) 6.57 Pr > F 0.0071						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand	2.6742	0.1020			Yes	14	8.30	0.0120	No	Means	Wand 23.76%
Glove vs 86"Non			1.4300	0.2413	Yes	14	1.23	0.2860	Yes	Neither	Neither
Glove vs W/S	0.5987	0.4391			Yes	14	0.10	0.7610	Yes	Neither	Neither
Wand vs 86"Non			0.0001	0.9909	Yes	14	1.32	0.2690	Yes	Neither	Neither
Wand vs W/S	0.7771	0.3780			Yes	14	11.81	0.0040	No	Means	Wand 13.89%
86"Non vs W/S			0.5400	0.4683	Yes	14	0.51	0.4900	Yes	Neither	Neither



**Figure H- 31: B3E3pAGen Experienced Users 3 Pass Avg. General Ratings**

**Table H- 31: B3E3pAGen Experienced Users 3 Pass Avg. General Ratings Statistics**

Benchmark 3, Experienced Users, 3 Pass Avg., General Impressions Ratings												
B3E3pA	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	CV
Glove	15	274.5		44.01		213.3		348.3		> 0.1000	Yes	16.03%
Wand	15	221.8		69.17		153.3		403.3		>0.1000	Yes	31.19%
86"Non	15	252.6		69.45		167.7		436.7		0.0135	No	27.49%
W/S	15	268.6		54.36		191.7		376.7		> 0.1000	Yes	20.24%
B3E3pA		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 6.57 Pr > F 0.0071						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)	Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand		2.6742	0.1020			Yes	14	8.30	0.0120	No	Means	Wand 23.76%
Glove vs 86"Non				1.4300	0.2413	Yes	14	1.23	0.2860	Yes	Neither	Neither
Glove vs W/S		0.5987	0.4391			Yes	14	0.10	0.7610	Yes	Neither	Neither
Wand vs 86"Non				0.0001	0.9909	Yes	14	1.32	0.2690	Yes	Neither	Neither
Wand vs W/S		0.7771	0.3780			Yes	14	11.81	0.0040	No	Means	Wand 13.89%
86"Non vs W/S				0.5400	0.4683	Yes	14	0.51	0.4900	Yes	Neither	Neither



**Figure H- 32: B3E3pAvgOvr Experienced User 3 Pass Avg. Overall Ratings**

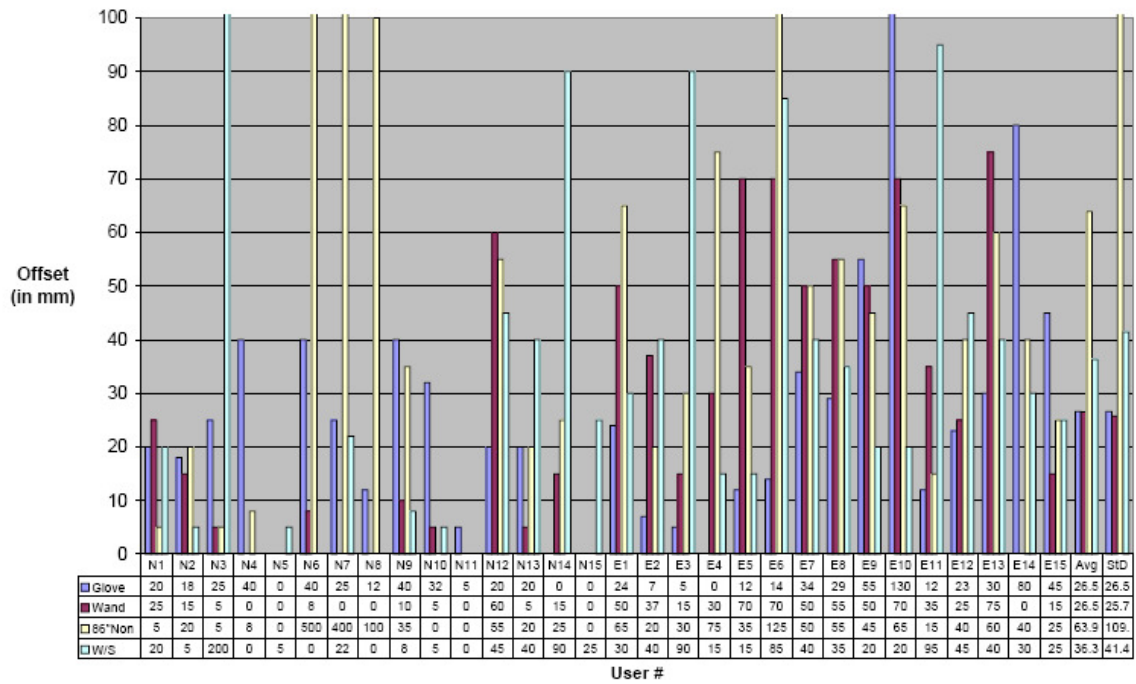
**Table H- 32: B3E3pAvgOvr Experienced User 3 Pass Avg. Overall Ratings Statistics**

Benchmark 3, Experienced Users, 3 Pass Avg., Overall Impressions Ratings													
B3E3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV					
Glove	15	274.5	44.01	213.3	348.3	> 0.1000	Yes	16.03%					
Wand	15	221.8	69.17	153.3	403.3	>0.1000	Yes	31.19%					
86"Non	15	252.6	69.45	167.7	436.7	0.0135	No	27.49%					
W/S	15	268.6	54.36	191.7	376.7	> 0.1000	Yes	20.24%					
B3E3pA		Homogeneity of Variance				Roy's Greatest Root: F(3,12) 6.6 Pr > F 0.0071							
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,14)		Pr > F	Equal Means ?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand		2.6742	0.1020			Yes	14	8.30	0.0120	No	Means	Wand	23.76%
Glove vs 86"Non				1.4300	0.2413	Yes	14	1.23	0.2860	Yes	Neither	Neither	
Glove vs W/S		0.5987	0.4391			Yes	14	0.10	0.7610	Yes	Neither	Neither	
Wand vs 86"Non				0.0001	0.9909	Yes	14	1.32	0.2690	Yes	Neither	Neither	
Wand vs W/S		0.7771	0.3780			Yes	14	11.81	0.0040	No	Means	Wand	13.89%
86"Non vs W/S				0.5400	0.4683	Yes	14	0.51	0.4900	Yes	Neither	Neither	

## **Appendix I**

### **All Users Benchmark 3 (Spatial Awareness) Detail**

**Benchmark 3 - All Users - Pass 1 - Part 1 Offsets**

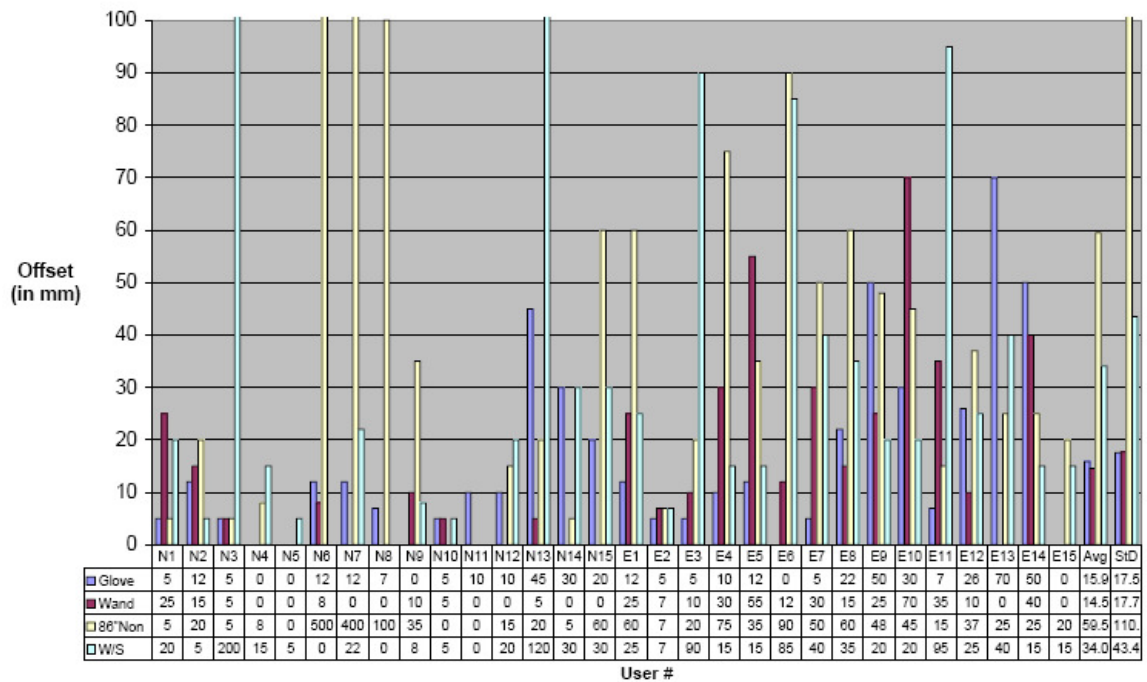


**Figure I- 1: B3Ap1-1off All Users Pass 1-Icon 1 Offsets**

**Table I- 1: B3Ap1-1off All Users Pass 1-Icon 1 Offsets Statistics**

Benchmark 3, All Users, Pass 1, Part 1 Offsets												
B3Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	26.57	26.59	0.00	130.00	0.0218	No	100.08%				
Wand	30	26.50	25.77	0.00	75.00	0.0023	Yes	97.25%				
86*Non	30	63.93	110.00	0.00	500.00	< 0.0001	No	172.06%				
W/S	30	36.33	41.46	0.00	200.00	< 0.0001	No	114.12%				
B3Ap1	Homogeneity of Variance					Roy's Greatest Root:			F(3,27)	1.57	Pr > F	0.2207
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.01	0.9131	Yes	29	0.00	0.2800	Yes	Neither	Neither	
Glove vs 86*Non			2.3400	0.1318	Yes	29	3.45	0.2384	Yes	Neither	Neither	
Glove vs W/S			1.0500	0.3096	Yes	29	1.03	0.9090	Yes	Neither	Neither	
Wand vs 86*Non			2.3600	0.1299	Yes	29	3.20	0.0260	No	Means	Wand	58.55%
Wand vs W/S			1.3100	0.2575	Yes	29	1.33	0.0000	No	Means	Wand	27.06%
86*Non vs W/S			1.9200	0.1716	Yes	29	1.49	0.2070	Yes	Neither	Neither	

**Benchmark 3 - All Users - Pass 1 - Part 2 Offsets**

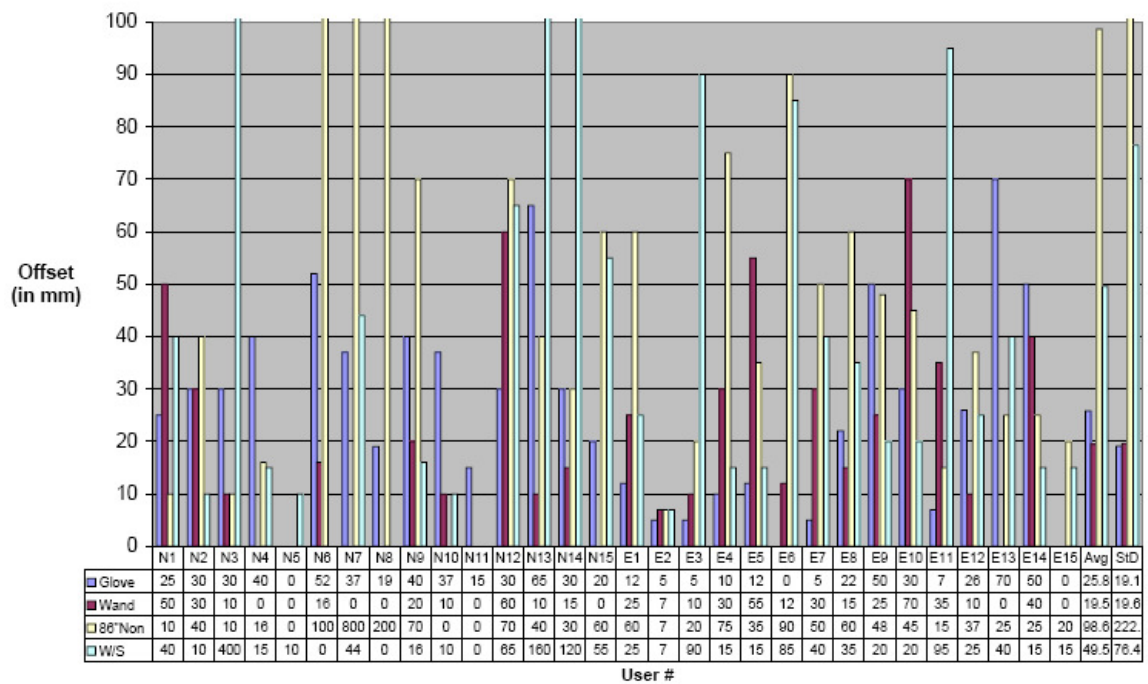


**Figure I- 2: B3Ap1-2off All Users Pass 2-Icon 2 Offsets**

**Table I- 2: B3Ap1-2off All Users Pass 2-Icon 2 Offsets Statistics**

Benchmark 3, All Users, Pass 1, Part 2 Offsets										
B3Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	30	15.90	17.59	0.00	70.00	< 0.0001	No	110.63%		
Wand	30	14.57	17.75	0.00	70.00	0.0022	No	121.83%		
86\"Non	30	59.50	110.20	0.00	500.00	< 0.0001	No	185.21%		
W/S	30	34.07	43.47	0.00	200.00	< 0.0001	No	127.59%		
B3Ap1		Homogeneity of Variance				Roy's Greatest Root:			F(3,27) 1.57 Pr > F 0.2207	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F					
Glove vs Wand				2.09	0.1537	Yes	29	0.00	0.2800	Yes
Glove vs 86\"Non				4.0000	0.0236	No	29	3.45	0.2384	Yes
Glove vs W/S				2.8400	0.0975	Yes	29	1.03	0.9090	Yes
Wand vs 86\"Non				3.9100	0.0257	No	29	3.20	0.0260	No
Wand vs W/S				1.8100	0.1842	Yes	29	1.33	0.0000	No
86\"Non vs W/S				3.5400	0.0356	No	29	1.49	0.2070	Yes
										Significant?
										Statistically Better
										Neither
										Neither
										Var
										Wand
										308.37%
										Means
										Wand
										57.24%
										W/S
										74.64%

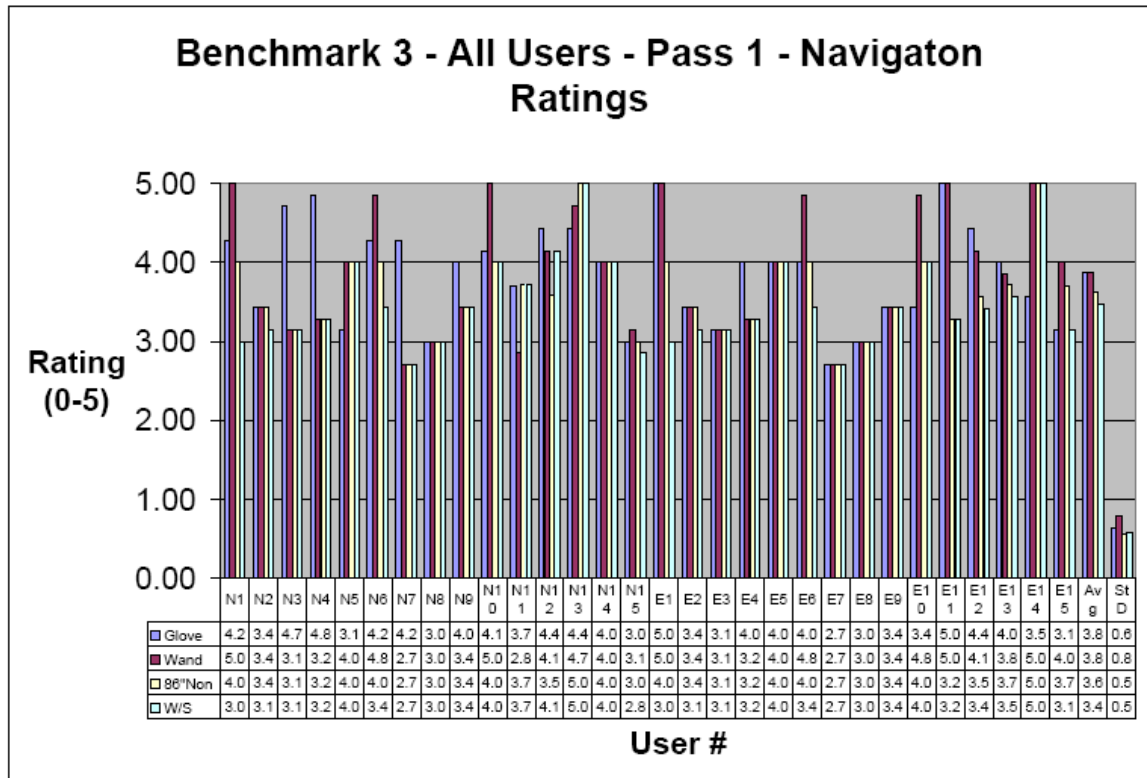
**Benchmark 3 - All Users - Pass 1 - Total Offsets**



**Figure I- 3: B3Ap1TotOff All Users Pass 1 Total Offsets**

**Table I- 3: B3Ap1TotOff All Users Pass 1 Total Offsets Statistics**

Benchmark 3, All Users, Pass 1, Total Offsets												
B3Ap1	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	CV
Glove	30	38.17		23.10		0.00		105.00		0.0297	No	60.52%
Wand	30	37.50		31.44		0.00		95.00		0.0713	Yes	83.84%
86*Non	30	116.40		220.90		0.00		1000.00		< 0.0001	No	189.78%
W/S	30	63.73		77.09		0.00		400.00		0.0010	No	120.96%
B3Ap1		Homogeneity of Variance				Roy's Greatest Root: F(3,27) 1.72 Pr > F 0.2165						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.0000	1.0000	Yes	29	1.93	0.2800	Yes	Neither	Neither
Glove vs 86*Non				0.0000	1.0000	Yes	29	1.53	0.2384	Yes	Neither	Neither
Glove vs W/S				1.1800	0.1892	Yes	29	0.66	0.9090	Yes	Neither	Neither
Wand vs 86*Non				0.6800	0.4172	Yes	29	0.44	0.0260	No	Means	Wand 210.40%
Wand vs W/S				2.2400	0.1454	Yes	29	2.93	0.0000	No	Means	Neither
86*Non vs W/S				1.1500	0.2918	Yes	29	3.06	0.2070	Yes	Neither	Neither

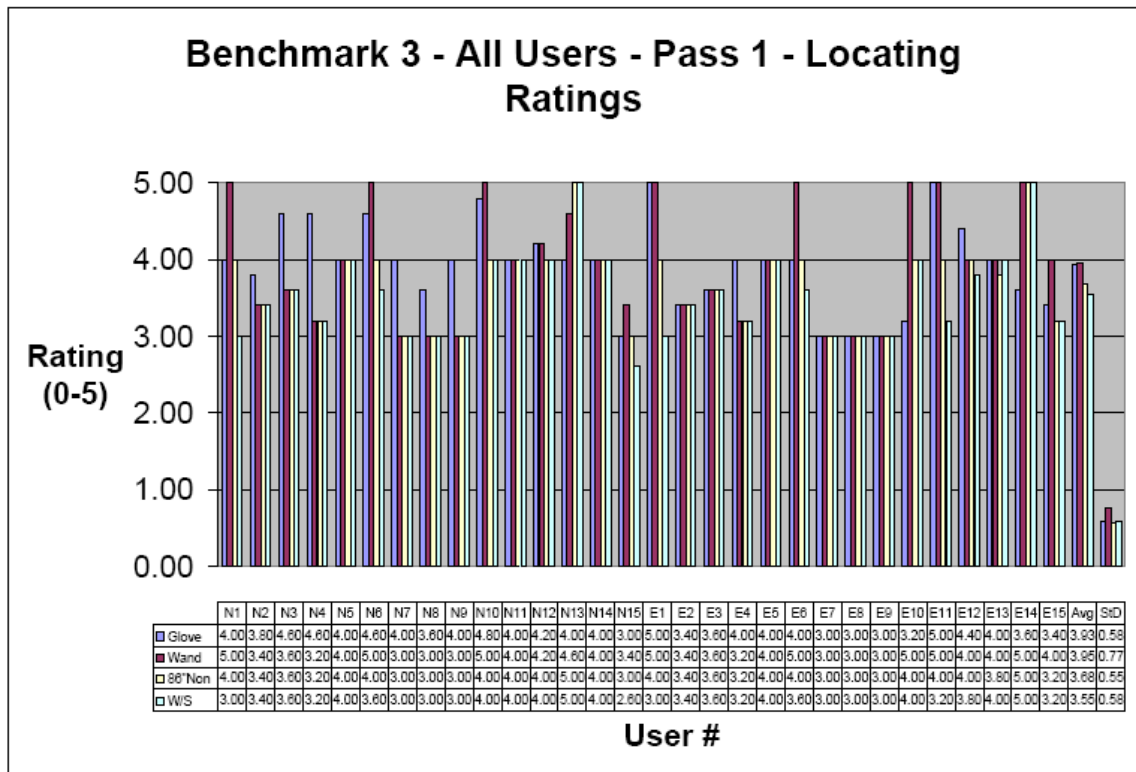


**Figure I- 4: B3Ap1Nav All Users Pass 1 Navigation Ratings**

**Table I- 4: B3Ap1Nav All Users Pass 1 Navigation Ratings Statistics**

Benchmark 3, All Users, Pass 1, Navigation Ratings												
B3Ap1	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	CV
Glove	30	3.87		0.64		2.71		5.00		0.0879	Yes	16.61%
Wand	30	3.88		0.80		2.71		5.00		0.0133	No	20.57%
86"Non	30	3.69		0.55		2.71		5.00		0.0154	No	14.98%
W/S	30	3.48		0.58		2.71		5.00		0.0030	No	16.58%
B3Ap1	Homogeneity of Variance					Roy's Greatest Root: F(3,27) 2.06 Pr > F 1291						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand				3.4600	0.0684	Yes	29	1.48	0.2434	Yes	Neither	Neither
Glove vs 86"Non				0.8000	0.3741	Yes	29	2.64	0.1262	Yes	Neither	Neither
Glove vs W/S				0.3700	0.5452	Yes	29	1.15	0.3027	Yes	Neither	Neither
Wand vs 86"Non				6.5000	0.0155	No	29	3.81	0.0713	No	Var/Mean	Wand 4.90%
Wand vs W/S				4.6500	0.0352	No	29	3.23	0.0940	No	Var/Mean	Wand 10.33%
86"Non vs W/S				0.0400	0.8480	Yes	29	5.21	0.0386	No	Means	86"Non 5.72%

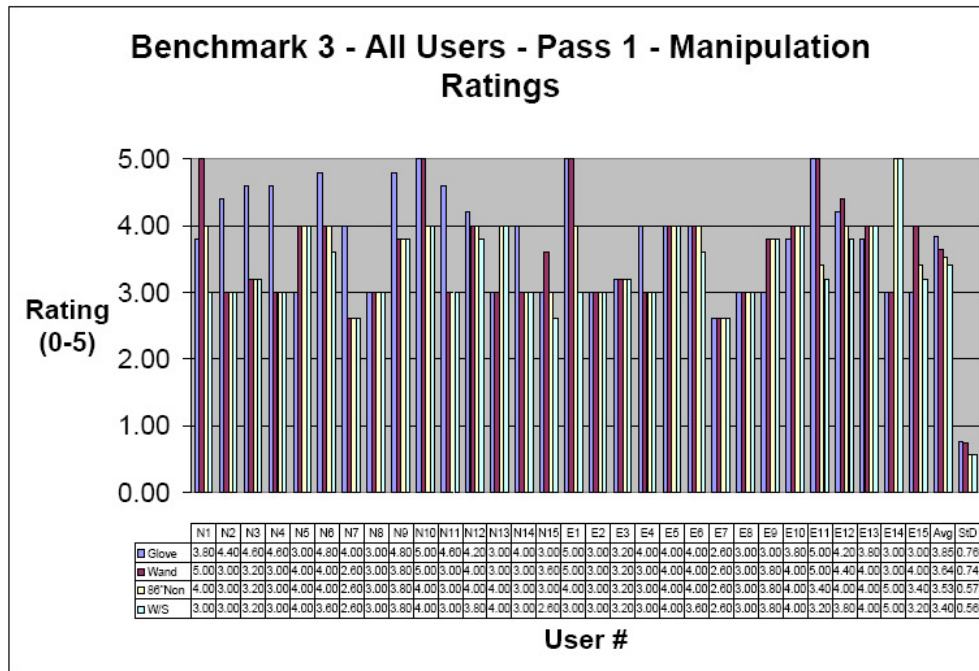




**Figure I- 5: B3Ap1Loc All Users Pass 1 Locating Ratings**

**Table I- 5: B3Ap1Loc All Users Pass 1 Locating Ratings Statistics**

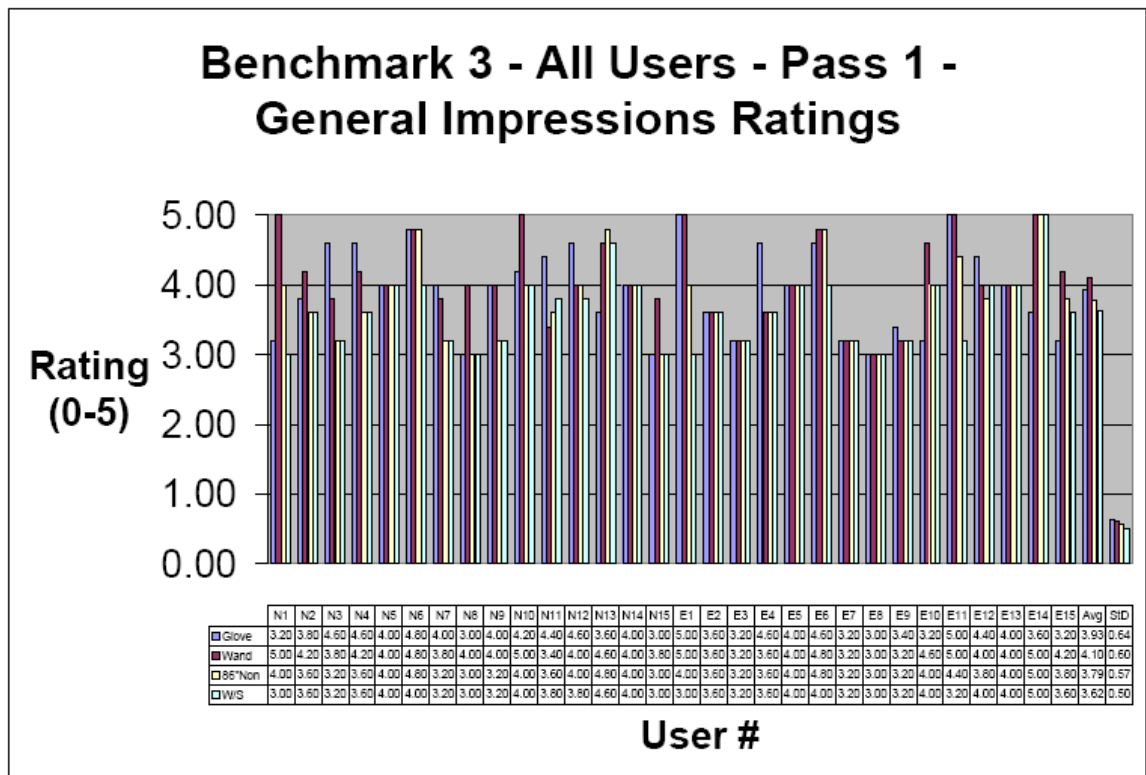
Benchmark 3, All Users, Pass 1, Locating Ratings											
B3Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	3.97	0.58	3.00	5.00	0.0110	No	14.51%			
Wand	30	3.95	0.77	3.00	5.00	0.0136	No	19.38%			
86\"Non	30	3.68	0.55	3.00	5.00	0.0009	No	15.08%			
W/S	30	3.55	0.58	2.60	5.00	0.0532	Yes	16.35%			
B3Ap1		Homogeneity of Variance				Roy's Greatest Root:				F(3,27) 2.58	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				5.1200	0.0274	Yes	29	3.09	0.1007	Yes	Neither
Glove vs 86\"Non				0.0500	0.8210	Yes	29	3.80	0.0714	No	Means
Glove vs W/S				0.0000	0.9737	Yes	29	1.66	0.2183	Yes	Neither
Wand vs 86\"Non				5.5100	0.0212	Yes	29	3.41	0.0859	No	Means
Wand vs W/S				3.7200	0.0588	Yes	29	5.30	0.0371	No	Means
86\"Non vs W/S				0.0500	0.8217	Yes	29	7.60	0.0154	No	Means
										Pr > F 0.1021	
										Statistically Better	
										Glove 7.35%	
										Wand 6.91%	
										86\"Non 10.27%	
										W/S 3.61%	



**Figure I- 6: B3Ap1Mov All Users Pass 1 Manipulation Ratings**

**Table I- 6: B3Ap1Mov All Users Pass 1 Manipulation Ratings Statistics**

Benchmark 3, All Users, Pass 1, Manipulation Ratings											
B3Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	3.85	0.76	2.60	5.00	0.0031	No	19.68%			
Wand	30	3.64	0.74	2.60	5.00	0.0019	No	20.22%			
86"Non	30	3.53	0.57	2.60	5.00	0.0004	No	16.17%			
W/S	30	3.40	0.56	2.60	5.00	0.0022	No	16.49%			
B3Ap1	Homogeneity of Variance				Roy's Greatest Root: F(3,27) 4.32 Pr > F 0.0131						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.0500	0.8256	Yes	29	2.07	0.1610	Yes	Neither	Neither
Glove vs 86"Non			4.6800	0.0347	No	29	3.67	0.0652	No	Var/Mean	Glove 8.16%
Glove vs W/S			4.4400	0.0394	No	29	6.78	0.0146	No	Var/Mean	Glove 11.62%
Wand vs 86"Non			2.6000	0.1142	Yes	29	0.89	0.3528	Yes	Neither	Neither
Wand vs W/S			2.6000	0.1125	Yes	29	2.80	0.1048	Yes	Neither	Neither
86"Non vs W/S			0.0100	0.9172	Yes	29	7.34	0.0112	No	Means	86"Non 3.76%



**Figure I- 7: B3Ap1Loc All Users Pass 1 General Impressions Ratings**

**Table I- 7: B3Ap1Loc All Users Pass 1 General Impressions Ratings Statistics**

Benchmark 3, All Users, Pass 1, General Impressions Ratings												
B3Ap1	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	CV
Glove	30	3.93		0.64		3.00		5.00		> 0.1000	Yes	16.19%
Wand	30	4.10		0.60		3.00		5.00		0.0345	No	14.70%
86"Non	30	3.79		0.57		3.00		5.00		0.0086	No	15.06%
W/S	30	3.62		0.50		3.00		5.00		0.0043	No	13.94%
B3Ap1	Homogeneity of Variance					Roy's Greatest Root: F(3,27) 6.18 Pr > F 0.0025						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand				0.1800	0.6741	Yes	29	1.85	0.1610	Yes	Neither	
Glove vs 86"Non				0.6400	0.4276	Yes	29	1.36	0.0652	No	Means	Glove 3.57%
Glove vs W/S				2.4500	0.1226	Yes	29	5.20	0.0146	No	Means	Glove 7.82%
Wand vs 86"Non				0.1400	0.7090	Yes	29	17.89	0.3528	Yes	Neither	Neither
Wand vs W/S				1.2200	0.2742	Yes	29	17.82	0.1048	Yes	Neither	Neither
86"Non vs W/S				0.4900	0.4868	Yes	29	5.91	0.0112	No	Means	86"Non 4.41%

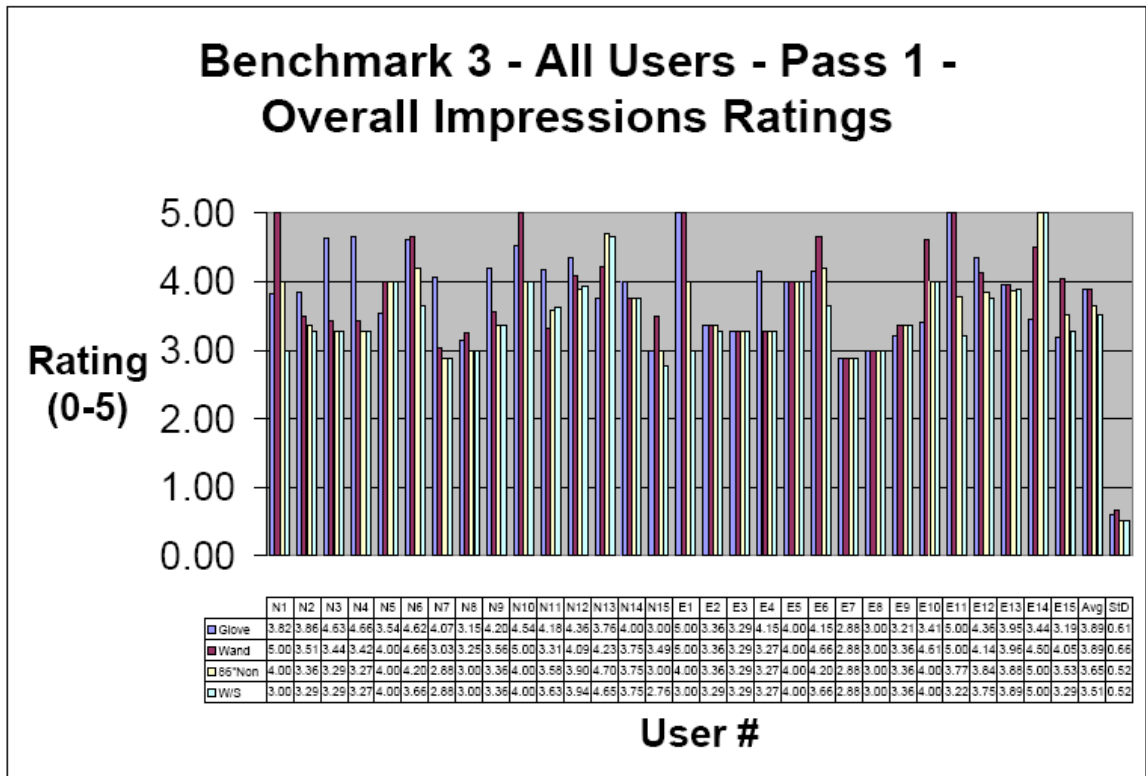
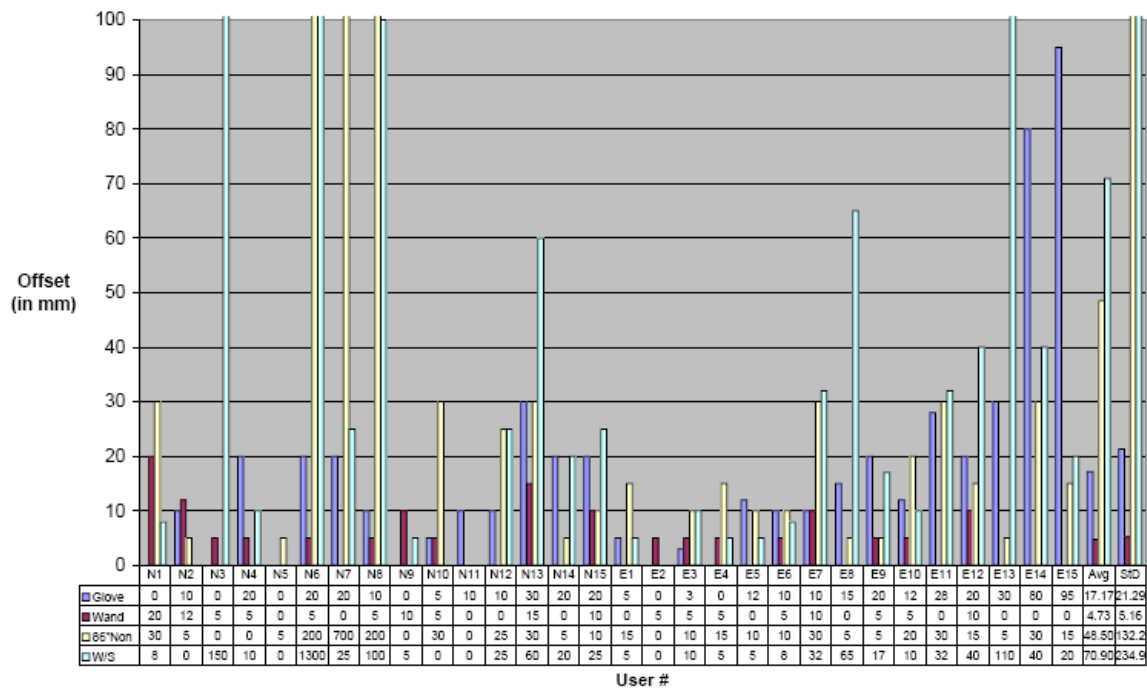


Figure I- 8: B3Ap1Ovr All Users Pass 1 Overall Ratings

Table I- 8: B3Ap1Ovr All Users Pass 1 Overall Ratings Statistics

Benchmark 3, All Users, Pass 1, Overall Impressions Ratings												
B3Ap1	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	4.15	0.52	3.08	5.00	> 0.1000	Yes	12.40%				
Wand	30	4.15	0.46	3.42	5.00	< 0.0001	No	11.17%				
86"Non	30	3.88	0.34	3.00	4.92	0.0047	No	8.72%				
W/S	30	3.75	0.41	2.52	5.00	0.0350	No	10.95%				
B3Ap1	Homogeneity of Variance					Roy's Greatest Root: F(3,27) 5.77 Pr > F 0.0035						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)		Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F			Pr > F					
Glove vs Wand			0.0900	0.7612	Yes	29	0.00	0.1610	Yes	Neither	Neither	
Glove vs 86"Non			4.8400	0.0317	No	29	7.10	0.0652	No	Var/Mean	Glove 6.57%	
Glove vs W/S			1.2000	0.2769	Yes	29	11.55	0.0146	No	Means	Glove 9.85%	
Wand vs 86"Non			13.2300	0.0006	No	29	14.46	0.3528	Yes	Var	Wand 6.50%	
Wand vs W/S			6.6100	0.0127	No	29	17.36	0.1048	Yes	Var	Wand 9.78%	
86"Non vs W/S			0.4000	0.5320	Yes	29	13.55	0.0112	No	Means	86"Non 3.50%	

**Benchmark 3 - All Users - Pass 2 - Part 1 Offsets**

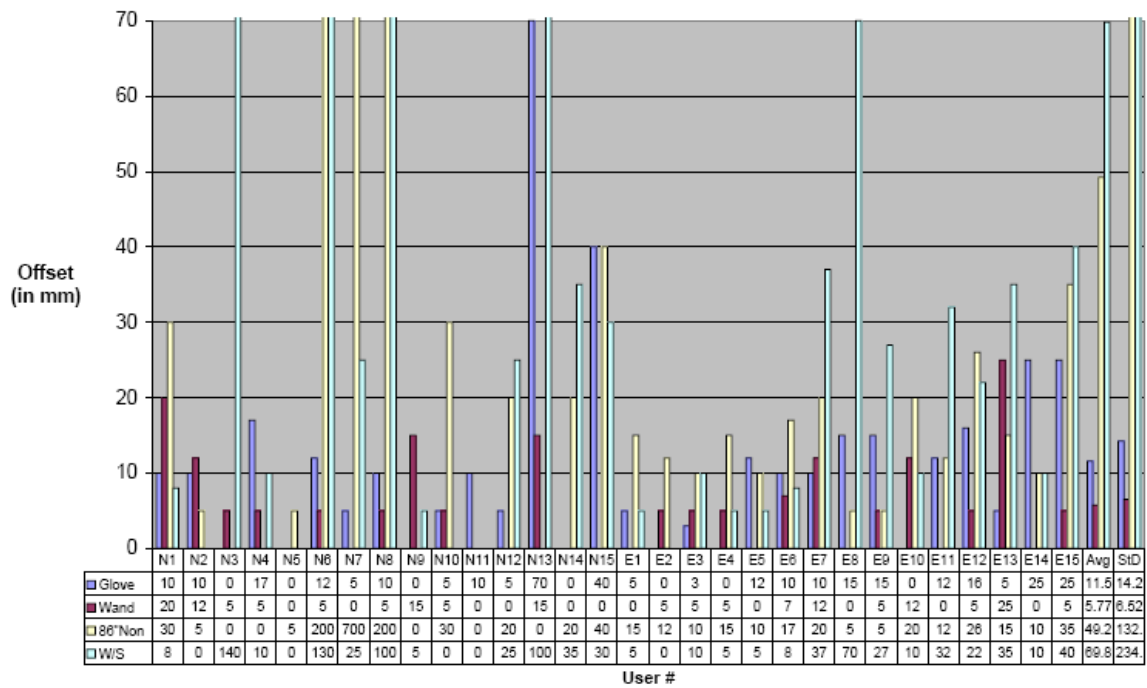


**Figure I- 9: B3Ap2-1off All Users Pass 2-Icon 1 Offsets**

**Table I- 9: B3Ap2-1off All Users Pass 2-Icon 1 Offsets Statistics**

Benchmark 3, All Users, Pass 2, Part 1 Offsets											
B3Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	17.17	21.29	0.00	95.00	< 0.0001	No	124.00%			
Wand	30	4.73	5.16	0.00	20.00	< 0.0001	No	109.00%			
86\"Non	30	48.50	132.00	0.00	700.00	< 0.0001	No	272.16%			
W/S	30	70.90	234.90	0.00	1300.00	< 0.0001	No	331.31%			
B3Ap2		Bartlett's Test				Levene's Test				Roy's Greatest Root:	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F	Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?
Glove vs Wand				3.1300	0.0822	Yes	29	8.60	0.2808	Yes	Neither
Glove vs 86\"Non				1.3700	0.2472	Yes	29	1.66	0.0408	No	Means
Glove vs W/S				1.1100	0.2968	Yes	29	1.57	0.4607	Yes	Neither
Wand vs 86\"Non				1.4400	0.2356	Yes	29	3.24	0.5560	Yes	Neither
Wand vs W/S				1.1300	0.2931	Yes	29	2.38	0.2638	Yes	Neither
86\"Non vs W/S				0.4900	0.4879	Yes	29	0.26	0.0177	No	Means
								F(3,27)		3.79	Pr > F
										0.0217	
										Statistically Better	
										Glove	
										64.60%	
										86\"Non	
										31.59%	

**Benchmark 3 - All Users - Pass 2 - Part 2 Offsets**



**Figure I- 10: B3Ap2-2off All Users Pass 2-Icon 2 Offsets**

**Table I- 10: B3Ap2-2off All Users Pass 2-Icon 2 Offsets Statistics**

Benchmark 3, All Users, Pass 2, Part 2 Offsets												
B3Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	11.57	14.26	0.00	70.00	0.0006	No	123.25%				
Wand	30	5.77	6.52	0.00	25.00	< 0.0001	No	113.07%				
86"Non	30	49.23	132.10	0.00	700.00	< 0.0001	No	268.33%				
W/S	30	69.80	234.80	0.00	1300.00	< 0.0001	No	336.39%				
B3Ap2	Homogeneity of Variance					Roy's Greatest Root: F(3,27) 3.79 Pr > F 0.0217						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			1.8900	0.1744	Yes	29	8.60	0.2808	Yes	Neither	Neither	
Glove vs 86"Non			2.1200	0.1509	Yes	29	1.66	0.0408	No	Means	Glove 325.50%	
Glove vs W/S			1.2200	0.2730	Yes	29	1.57	0.4607	Yes	Neither	Neither	
Wand vs 86"Non			2.1600	0.1468	Yes	29	3.24	0.5560	Yes	Neither	Neither	
Wand vs W/S			1.2400	0.2707	Yes	29	2.38	0.2638	Yes	Neither	Neither	
86"Non vs W/S			0.3100	0.5826	Yes	29	0.26	0.0177	No	Means	86"Non 41.78%	

### Benchmark 3 - All Users - Pass 2 - Total Offsets

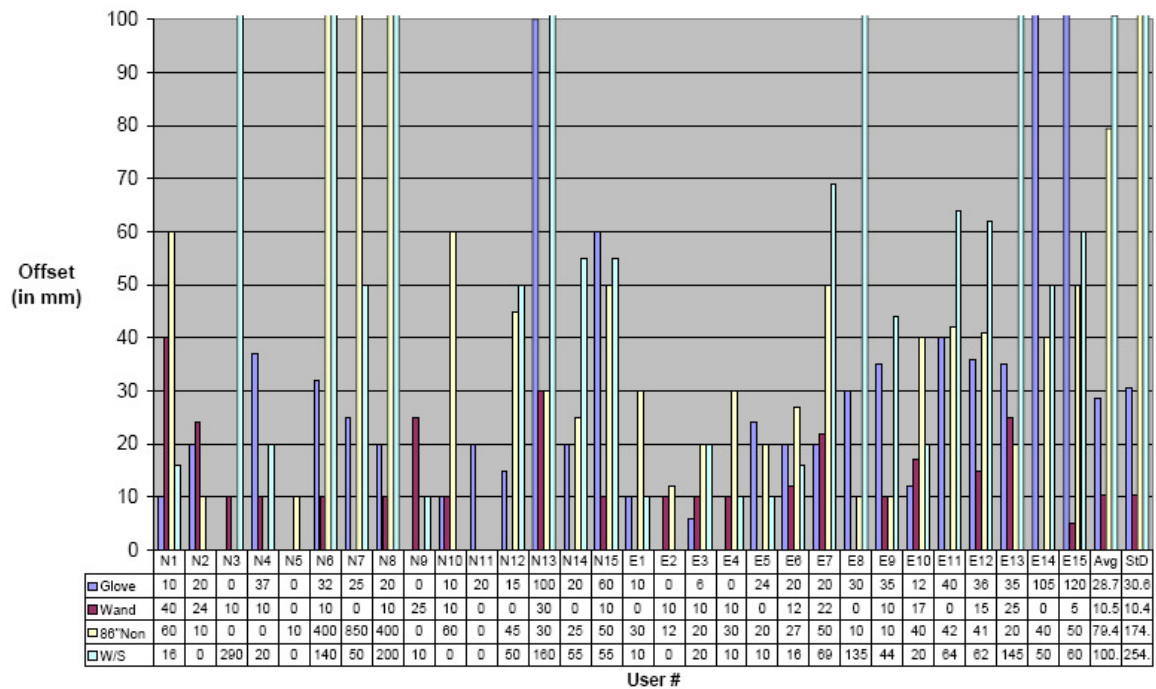
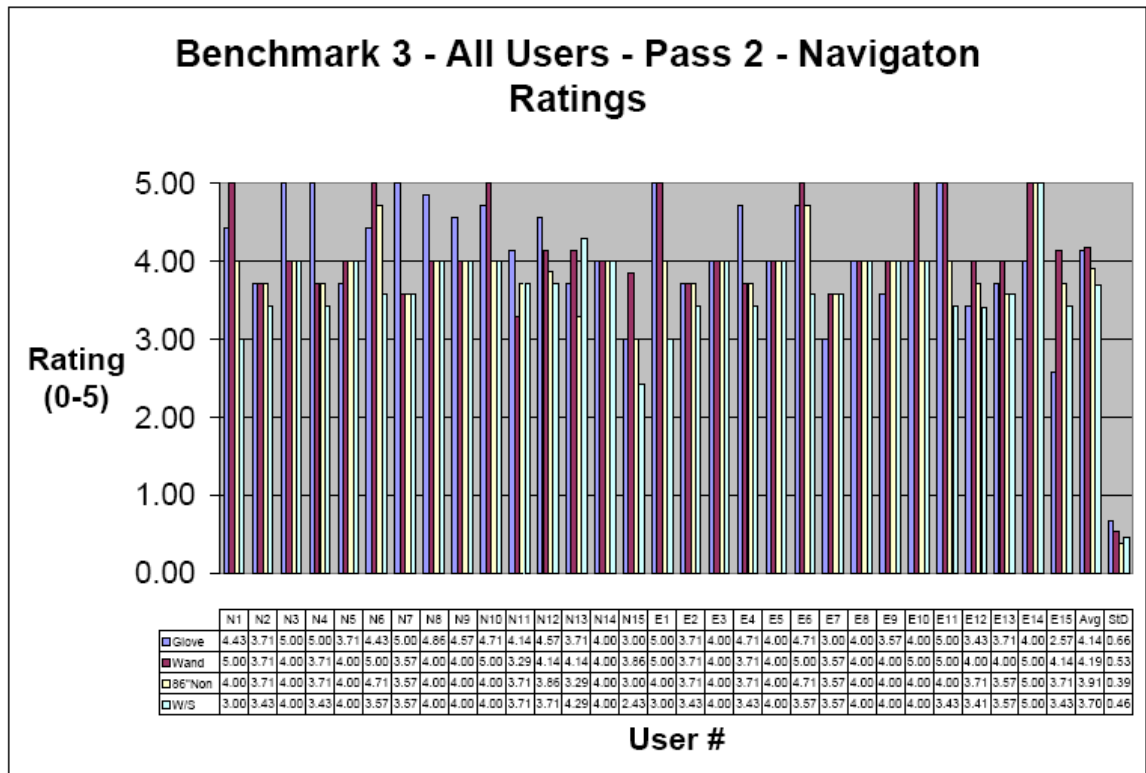


Figure I- 11: B3Ap2TotOff All Users Pass 2 Total Offsets

Table I- 11: B3Ap2TotOff All Users Pass 2 Total Offsets Statistics

Benchmark 3, All Users, Pass 2, Total Offsets												
B3Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	19.23	20.61	0.00	100.00	0.0073	No	107.18%				
Wand	30	9.50	10.53	0.00	40.00	0.0011	No	110.84%				
86"Non	30	73.90	176.10	0.00	850.00	< 0.0001	No	238.29%				
W/S	30	91.74	256.00	0.00	1400.00	< 0.0001	No	279.05%				
B3Ap2	Homogeneity of Variance					Roy's Greatest Root: F(3,27) 3.24 Pr > F 0.0605						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.5400	0.4688	Yes	29	1.26	0.2808	Yes	Neither	Neither	
Glove vs 86"Non			3.0900	0.0899	Yes	29	5.08	0.0408	No	Means	Glove 284.30%	
Glove vs W/S			0.0000	0.9482	Yes	29	0.58	0.4607	Yes	Neither	Neither	
Wand vs 86"Non			1.9500	0.1739	Yes	29	0.36	0.5560	Yes	Neither	Neither	
Wand vs W/S			0.9600	0.3359	Yes	29	1.36	0.2638	Yes	Neither	Neither	
86"Non vs W/S			1.9400	0.1747	Yes	29	7.22	0.0177	No	Means	86"Non 24.14%	

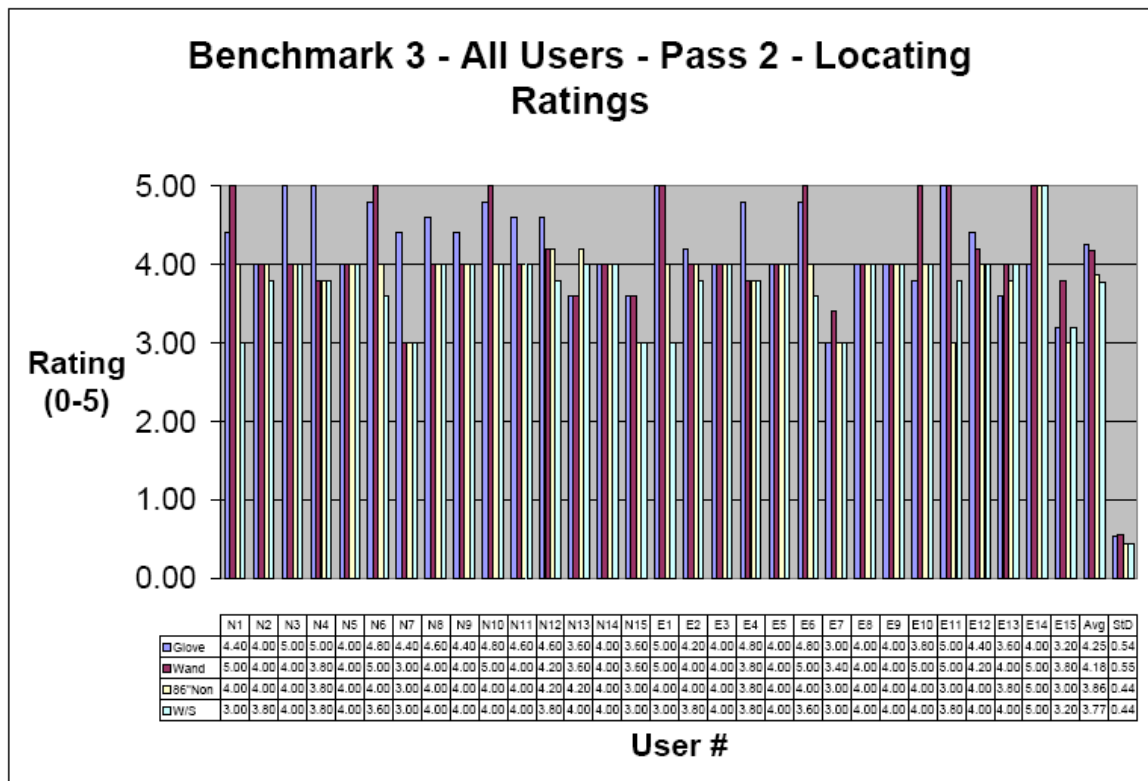


**Figure I- 12: B3Ap2Nav All Users Pass 2 Navigation Ratings**

**Table I- 12: B3Ap2Nav All Users Pass 2 Navigation Ratings Statistics**

Benchmark 3, All Users, Pass 2, Navigation Ratings											
B3Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.14	0.66	2.57	5.00	> 0.1000	Yes	15.94%			
Wand	30	4.19	0.53	3.29	5.00	< 0.0001	No	12.74%			
86\"Non	30	3.91	0.39	3.00	5.00	< 0.0001	No	10.00%			
W/S	30	3.70	0.46	2.43	5.00	0.0065	No	12.53%			
B3Ap2		Homogeneity of Variance				Roy's Greatest Root				F(3,27) 4.46 Pr > F 0.0252	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				2.0300	0.1596	Yes	29	4.95	0.0431	No	Means
Glove vs 86\"Non				6.2300	0.0154	No	29	5.54	0.0338	No	Var/Mean
Glove vs W/S				3.0900	0.0840	Yes	29	1.68	0.2165	Yes	Neither
Wand vs 86\"Non				2.9200	0.0929	No	29	1.81	0.1955	Yes	Var
Wand vs W/S				0.5500	0.4622	Yes	29	10.09	0.0067	No	Means
86\"Non vs W/S				0.4300	0.5161	Yes	29	11.27	0.0047	No	Means
											Statistically Better
											Wand 1.03%
											Glove 5.65%
											Wand 6.62%
											Wand 11.63%
											86\"Non 5.37%

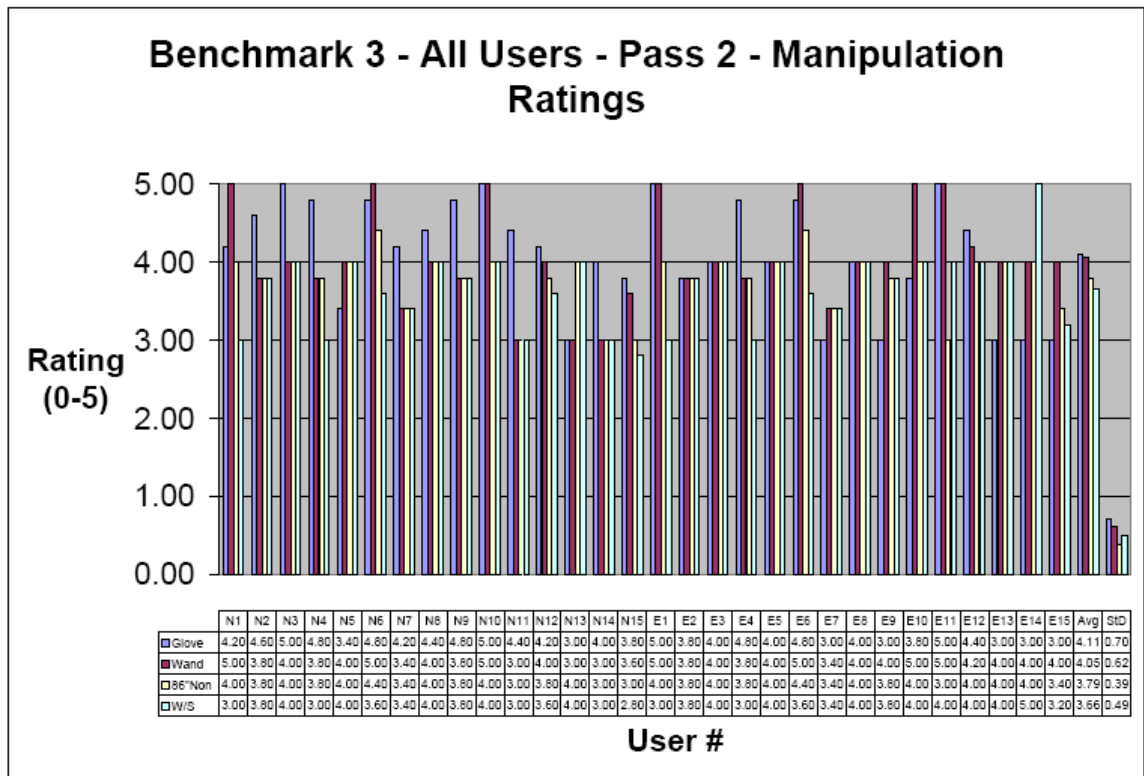




**Figure I- 13: B3Ap2Loc All Users Pass 2 Locating Ratings**

**Table I- 13: B3Ap2Loc All Users Pass 2 Locating Ratings Statistics**

Benchmark 3, All Users, Pass 2, Locating Ratings												
B3Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	4.09	4.42	3.00	5.00	> 0.1000	Yes	108.25%				
Wand	30	4.01	4.35	3.00	5.00	< 0.0001	No	108.58%				
86"Non	30	3.74	4.00	3.00	5.00	< 0.0001	No	106.79%				
W/S	30	3.64	3.91	3.00	5.00	< 0.0001	No	107.54%				
B3Ap2	Homogeneity of Variance					Roy's Greatest Root: F(3,27) 3.58 Pr > F 0.0446						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.0200	0.8918	Yes	29	3.11	0.0431	No	Means	Glove 1.88%	
Glove vs 86"Non			1.2100	0.2765	Yes	29	3.96	0.0338	No	Means	Glove 8.40%	
Glove vs W/S			1.2200	0.2745	Yes	29	4.07	0.2165	Yes	Neither	Neither	
Wand vs 86"Non			1.5000	0.2235	Yes	29	2.95	0.1955	Yes	Neither	Neither	
Wand vs W/S			1.5200	0.2227	Yes	29	10.95	0.0067	No	Means	Wand 9.28%	
86"Non vs W/S			0.0000	0.9858	Yes	29	12.54	0.0047	No	Means	86"Non 2.83%	



**Figure I- 14: B3Ap2Mov All Users Pass 2 Manipulation Ratings**

**Table I- 14: B3Ap2Mov All Users Pass 2 Manipulation Ratings Statistics**

Benchmark 3, All Users, Pass 2, Manipulation Ratings												
B3Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	3.89	4.32	3.00	5.00	> 0.1000	Yes	111.19%				
Wand	30	3.86	4.25	3.00	5.00	< 0.0001	No	109.92%				
86"Non	30	3.67	3.91	3.00	4.40	< 0.0001	No	106.52%				
W/S	30	3.51	3.81	2.80	5.00	0.0015	No	108.67%				
B3Ap2	Homogeneity of Variance					Roy's Greatest Root: F(3,27) 3.25 Pr > F 0.0372						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.8600	0.3367	Yes	29	0.16	0.0431	No	Means	Glove	0.69%
Glove vs 86"Non			12.7900	0.0007	No	29	5.24	0.0338	No	Var/Means	Glove	5.71%
Glove vs W/S			5.3200	0.0247	No	29	6.68	0.2165	Yes	Var	Glove	9.80%
Wand vs 86"Non			6.2600	0.0152	No	29	7.34	0.1955	Yes	Var	Wand	5.05%
Wand vs W/S			1.7400	0.1928	Yes	29	8.93	0.0067	No	Means	Wand	9.17%
86"Non vs W/S			1.4200	0.2377	Yes	29	2.25	0.0047	No	Means	86"Non	4.34%

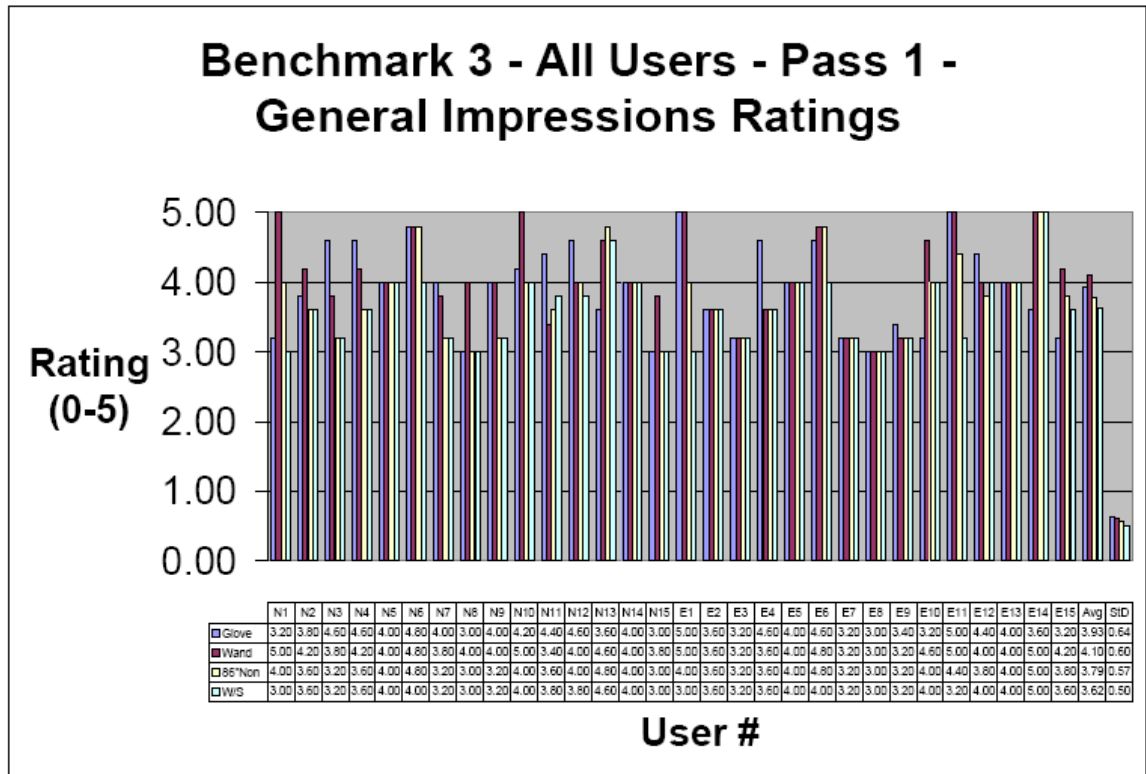
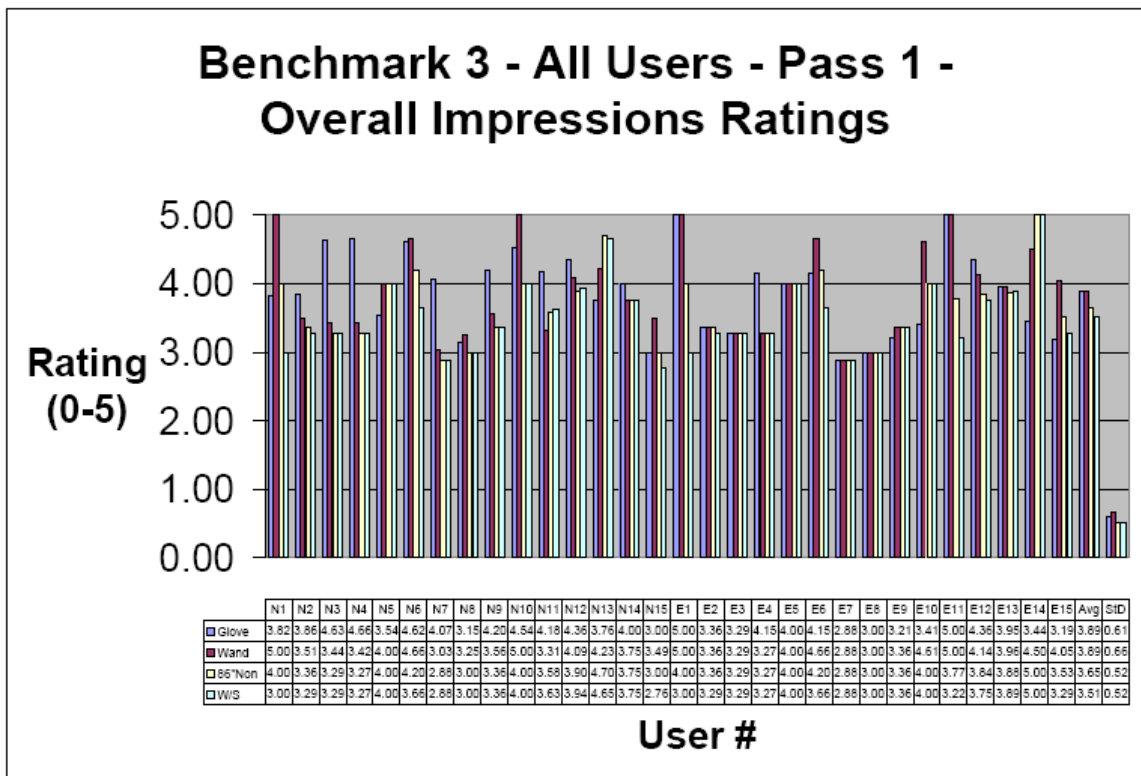


Figure I- 15: B3Ap2Gen All User Pass 2 General Impressions Ratings

Table I- 15: B3Ap2Gen All User Pass 2 General Impressions Ratings  
Statistics

Benchmark 3, All Users, Pass 2, General Impressions Ratings											
B3Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.20	0.57	2.80	5.00	0.0504	Yes	13.53%			
Wand	30	4.25	0.47	3.40	5.00	< 0.0001	No	10.97%			
86"Non	30	4.05	0.33	3.00	5.00	0.0006	No	8.18%			
W/S	30	3.84	0.43	3.00	5.00	< 0.0001	No	11.23%			
B3Ap2	Homogeneity of Variance					Roy's Greatest Root: F(3,27) 4.23 Pr > F 0.0142					
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > Chisq	F Value	Pr > F							
Glove vs Wand			1.4200	0.2384	Yes	29	0.22	0.0431	No	Means	Wand 1.11%
Glove vs 86"Non			5.5400	0.0220	No	29	2.21	0.0338	No	Var/Means	Glove 3.64%
Glove vs W/S			2.0000	0.1628	Yes	29	7.60	0.2165	Yes	Neither	Glove 8.57%
Wand vs 86"Non			2.8400	0.0973	Yes	29	9.60	0.1955	Yes	Neither	Neither
Wand vs W/S			0.1900	0.6638	Yes	29	12.62	0.0067	No	Means	Wand 9.58%
86"Non vs W/S			0.0500	0.3097	Yes	29	13.09	0.0047	No	Means	86"Non 5.11%

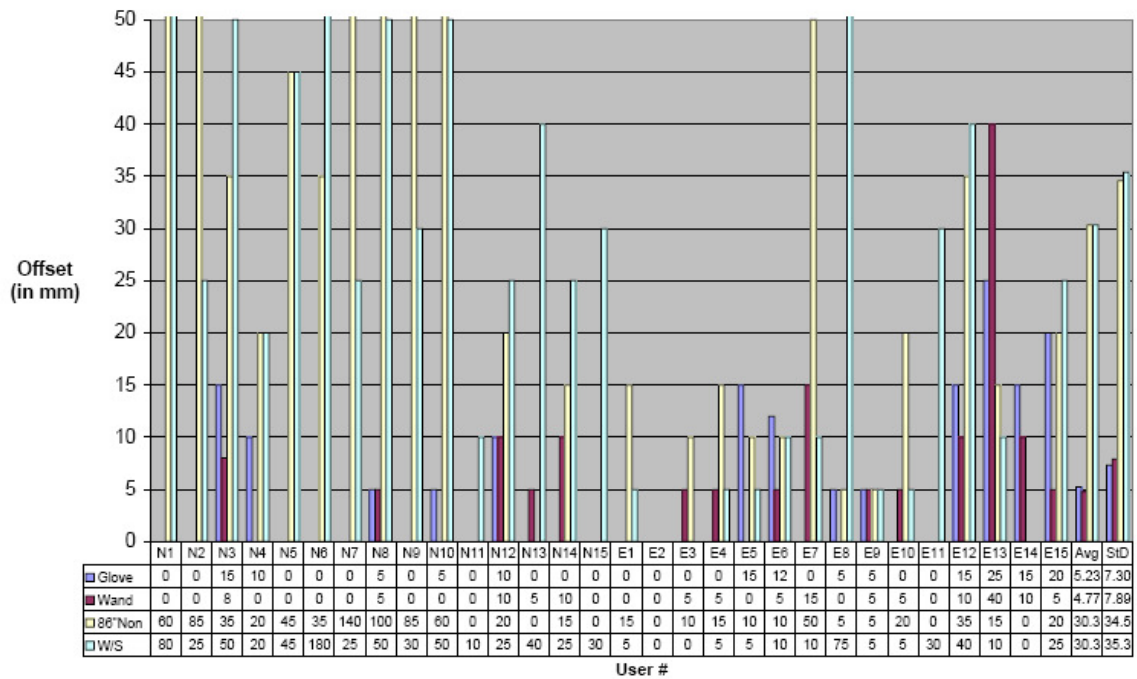


**Figure I- 16: B3Ap2Ovr All Users Pass 2 Overall Impressions Ratings**

**Table I- 16: B3Ap2Ovr All Users Pass 2 Overall Impressions Ratings Statistics**

Benchmark 3, All Users, Pass 2, Overall Impressions Ratings											
B3Ap2	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	3.95	0.60	2.89	5.00	> 0.1000	Yes	15.14%			
Wand	30	3.96	0.58	3.11	5.00	0.0153	No	14.75%			
86*Non	30	3.72	0.46	3.00	4.98	0.0386	No	12.45%			
W/S	30	3.58	0.48	2.68	5.00	0.0166	No	13.36%			
B3Ap2		Homogeneity of Variance				Roy's Greatest Root:				F(3,27) 3.90 Pr > F 0.0195	
		Bartlett's Test	Levene's Test	Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				0.0300	0.8596	Yes	29	0.00	0.9561	Yes	Neither
Glove vs 86*Non				2.6800	0.1072	Yes	29	3.91	0.0576	No	Means
Glove vs W/S				1.7400	0.1929	Yes	29	7.88	0.0088	No	Means
Wand vs 86*Non				2.0400	0.1587	Yes	29	10.98	0.0025	No	Means
Wand vs W/S				1.2900	0.2601	Yes	29	11.69	0.0019	No	Means
86*Non vs W/S				0.0200	0.8874	Yes	29	8.71	0.0062	No	Means

**Benchmark 3 - All Users - Pass 3 - Part 1 Offsets**

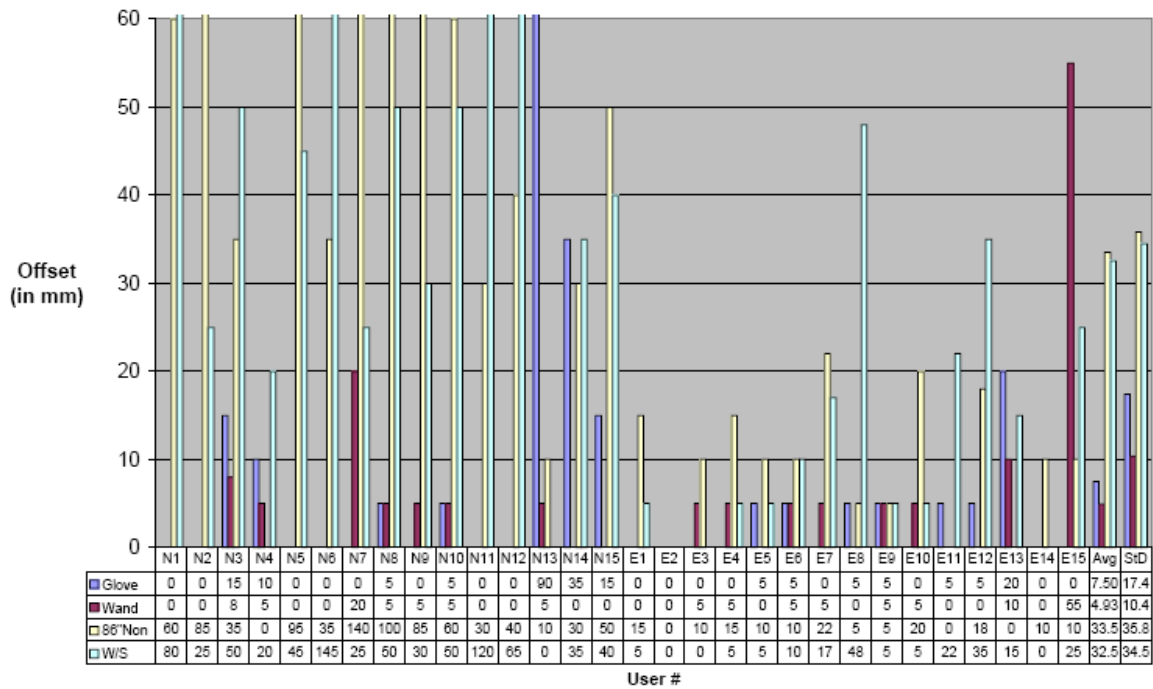


**Figure I- 17: B3Ap3-1off All Users Pass 3-Icon 1 Offsets**

**Table I- 17: B3Ap3-1off All Users Pass 3-Icon 1 Offsets Statistics**

Benchmark 3, All Users, Pass 3, Part 1 Offsets											
B3Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	5.23	7.31	0.00	25.00	< 0.0001	No	139.59%			
Wand	30	4.77	7.89	0.00	40.00	< 0.0001	No	165.60%			
86*Non	30	30.33	34.59	0.00	140.00	< 0.0001	No	114.05%			
W/S	30	30.33	35.38	0.00	180.00	0.0026	No	116.65%			
B3Ap3		Homogeneity of Variance				Roy's Greatest Root:				F(3,27) 6.72 Pr > F 0.0016	
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?
		Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand				0.0400	0.8418	Yes	29	0.14	0.0239	No	Means
Glove vs 86*Non				6.8700	0.0112	No	29	14.03	0.0019	No	Var/Means
Glove vs W/S				2.4600	0.1220	Yes	29	13.68	0.0022	No	Means
Wand vs 86*Non				6.7100	0.0121	No	29	14.69	0.9759	Yes	Var
Wand vs W/S				2.4200	0.1252	Yes	29	13.64	0.1367	Yes	Neither
86*Non vs W/S				0.0000	0.9502	Yes	29	0.00	0.0942	No	Means
										Statistically Better	
										Wand 8.91%	
										Glove 82.75%	
										Wand 84.28%	
										Neither	
										Neither	

**Benchmark 3 - All Users - Pass 3 - Part 2 Offsets**

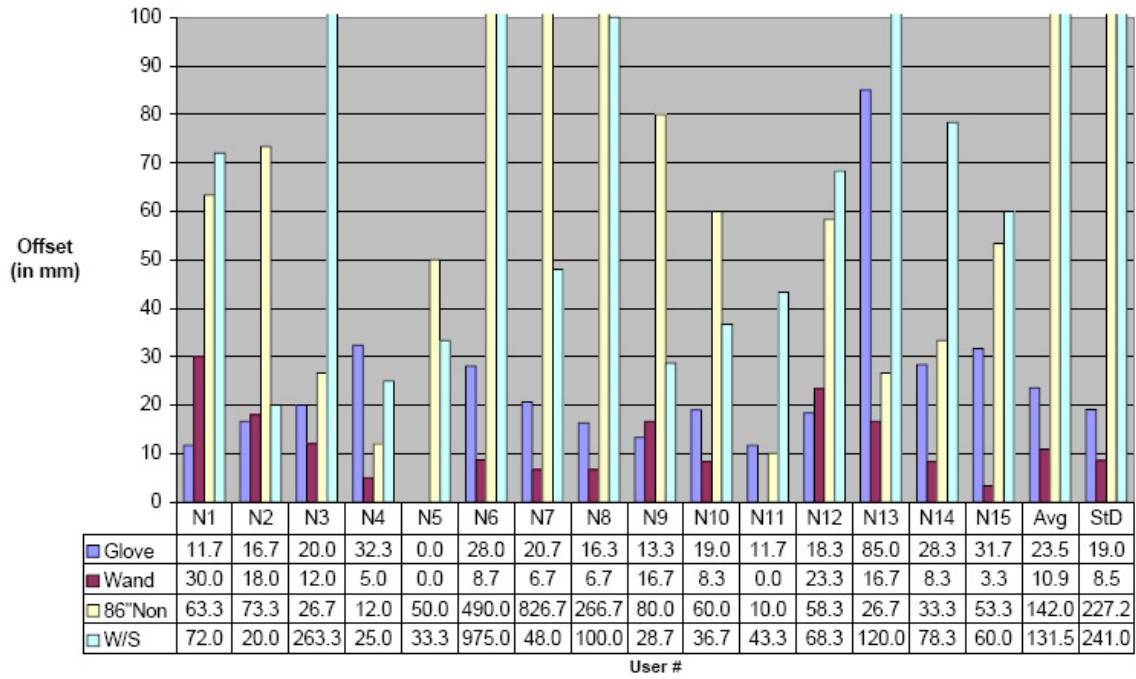


**Figure I- 18: B3Ap3-2off All Users Pass 3-Icon 2 Offsets**

**Table I- 18: B3Ap3-2off All Users Pass 3-Icon 2 Offsets Statistics**

Benchmark 3, All Users, Pass 3, Part 2 Offsets												
B3Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	7.50	17.41	0.00	90.00	< 0.0001	No	232.13%				
Wand	30	4.93	10.40	0.00	55.00	< 0.0001	No	210.83%				
86*Non	30	33.50	35.85	0.00	140.00	0.0061	No	107.01%				
W/S	30	32.57	34.50	0.00	145.00	0.0219	No	105.93%				
B3Ap3	Homogeneity of Variance				Roy's Greatest Root:				F(3,27)	6.72	Pr > F	0.0016
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.7400	0.3935	Yes	29	0.14	0.7148	Yes	Neither	Neither	
Glove vs 86*Non			7.0700	0.0101	No	29	14.03	0.0008	No	Var/Mean	Glove	346.67%
Glove vs W/S			2.9800	0.0898	Yes	29	13.68	0.0009	No	Means	Glove	334.27%
Wand vs 86*Non			7.2900	0.0071	No	29	14.69	0.0006	No	Var/Mean	Wand	579.10%
Wand vs W/S			3.2900	0.0747	Yes	29	13.64	0.0009	No	Means	Wand	84.85%
86*Non vs W/S			0.0300	0.8529	Yes	29	0.00	1.0000	Yes	Neither	Neither	

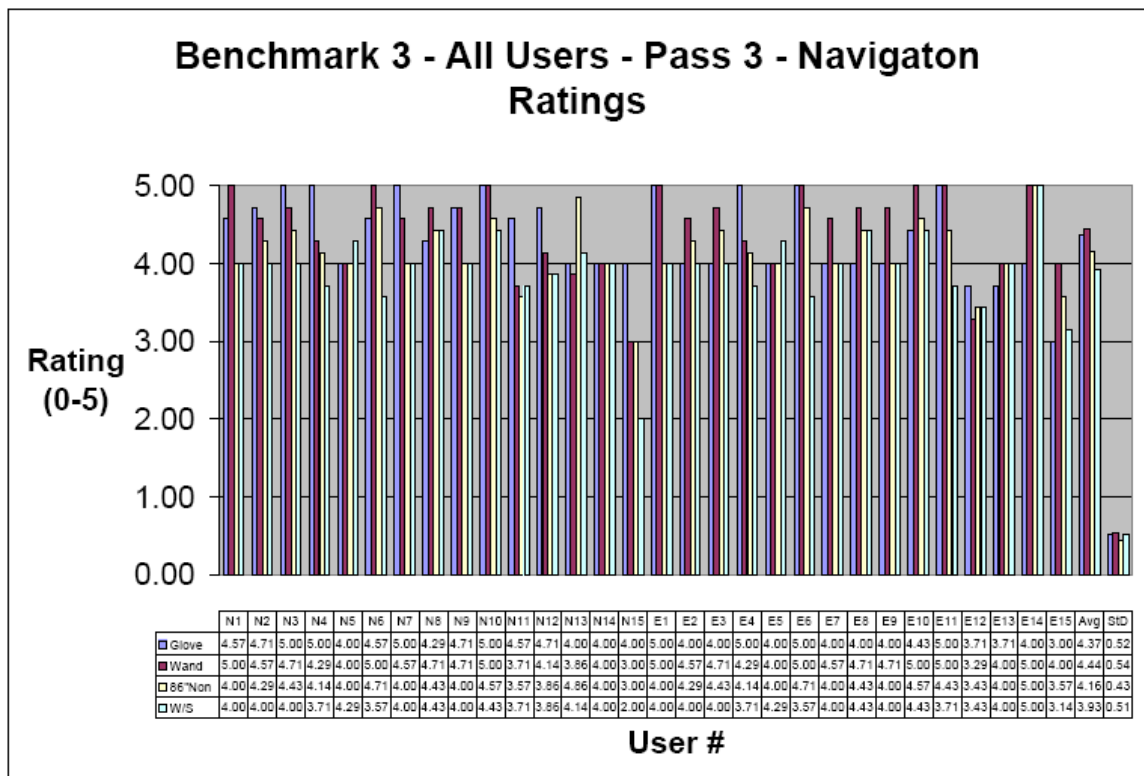
**Benchmark 3 - Novice Users - 3 Pass Avg. - Total Offsets**



**Figure I- 19: B3Ap3TotOff All Users Pass 3 Total Offsets**

**Table I- 19: B3Ap3TotOff All Users Pass 3 Total Offsets Statistics**

Benchmark 3, All Users, Pass 3, Total Offsets												
B3Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	6.73	17.88	0.00	90.00	< 0.0001	No	265.56%				
Wand	30	7.03	10.90	0.00	55.00	< 0.0001	No	154.98%				
86"Non	30	59.17	71.02	0.00	280.00	< 0.0001	No	120.03%				
W/S	30	58.73	66.99	0.00	325.00	0.0071	No	114.06%				
B3Ap3	Homogeneity of Variance					Roy's Greatest Root: F(3,27) 10.9 Pr > F 0.001						
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			1.0400	0.3175	Yes	29	6.41	0.0239	No	Means	Glove	4.27%
Glove vs 86"Non			1.1700	0.2892	Yes	29	14.59	0.0019	No	Means	Glove	778.81%
Glove vs W/S			1.4200	0.2427	Yes	29	13.67	0.0022	No	Means	Glove	772.27%
Wand vs 86"Non			0.1100	0.7432	Yes	29	0.00	0.9759	Yes	Neither		Neither
Wand vs W/S			0.9200	0.3445	Yes	29	1.84	0.1367	Yes	Neither		Neither
86"Non vs W/S			0.6700	0.4189	Yes	29	3.22	0.0942	No	Means	W/S	0.75%

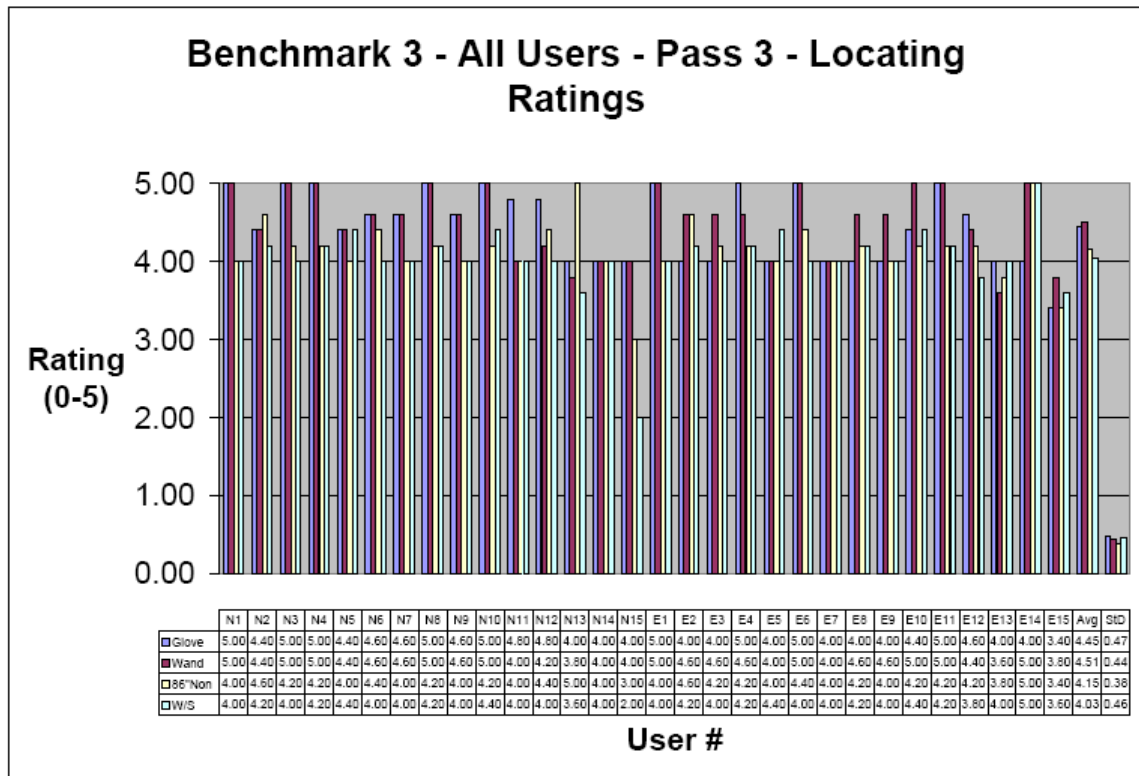


**Figure I- 20: B3Ap3Nav All Users Pass 3 Navigation Ratings**

**Table I- 20: B3Ap3Nav All Users Pass 3 Navigation Ratings Statistics**

Benchmark 3, All Users, Pass 3, Navigation Ratings											
B3Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.37	0.52	3.00	5.00	0.0005	No	11.96%			
Wand	30	4.44	0.54	3.00	5.00	0.0041	No	12.12%			
86\"Non	30	4.16	0.43	3.00	5.00	0.0085	No	10.42%			
W/S	30	3.93	0.51	2.00	5.00	0.0006	No	12.98%			
B3Ap3		Bartlett's Test		Levene's Test		Roy's Greatest Root		F(3,27)	3.97	Pr > F	0.0352
		Chi-Sq	Pr > ChiSq	F Value	Pr > F	Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?
Glove vs Wand				0.0300	0.8720	Yes	29	3.00	0.2434	Yes	Neither
Glove vs 86\"Non				1.0200	0.3164	Yes	29	11.62	0.1262	Yes	Neither
Glove vs W/S				0.0100	0.9359	Yes	29	2.43	0.3027	Yes	Neither
Wand vs 86\"Non				1.1100	0.2963	Yes	29	5.22	0.0713	No	Means
Wand vs W/S				0.0300	0.8546	Yes	29	0.06	0.0940	No	Means
86\"Non vs W/S				0.2600	0.6181	Yes	29	1.20	0.0386	No	Means
											Statistically Better
											Wand 6.22%
											86\"Non 11.47%
											5.60%

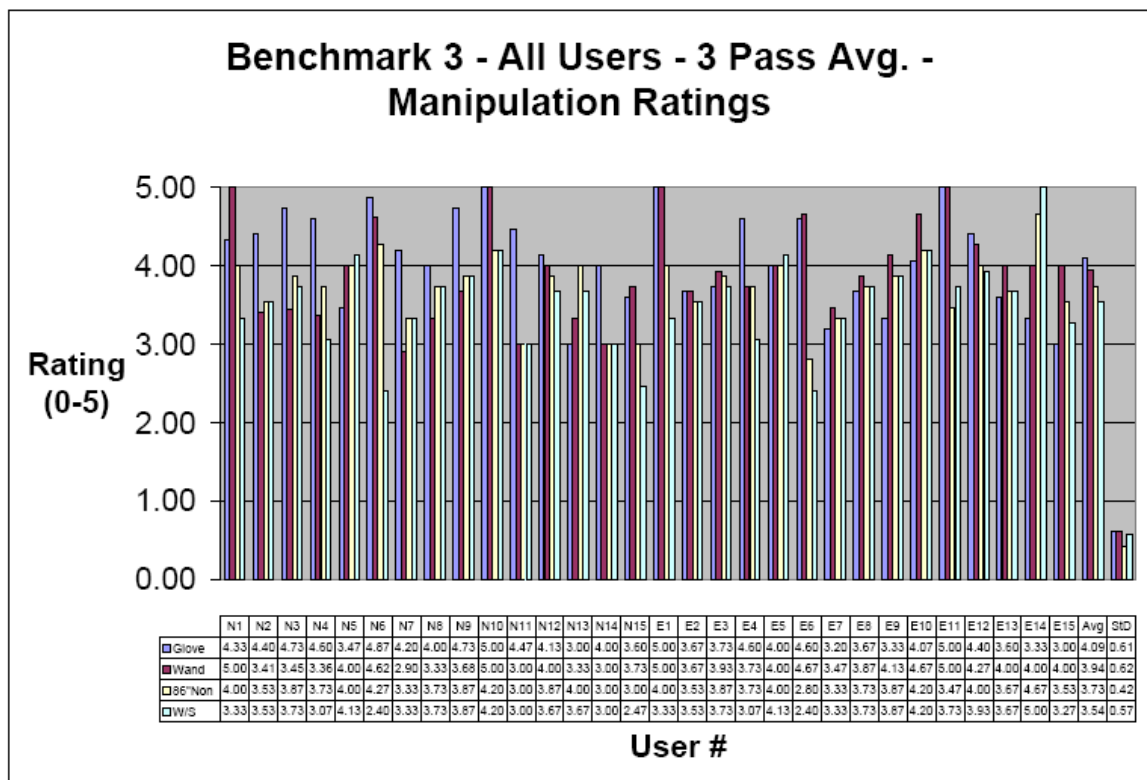




**Figure I- 21: B3Ap3Loc All Users Pass 3 Locating Ratings**

**Table I- 21: B3Ap3Loc All Users Pass 3 Locating Ratings Statistics**

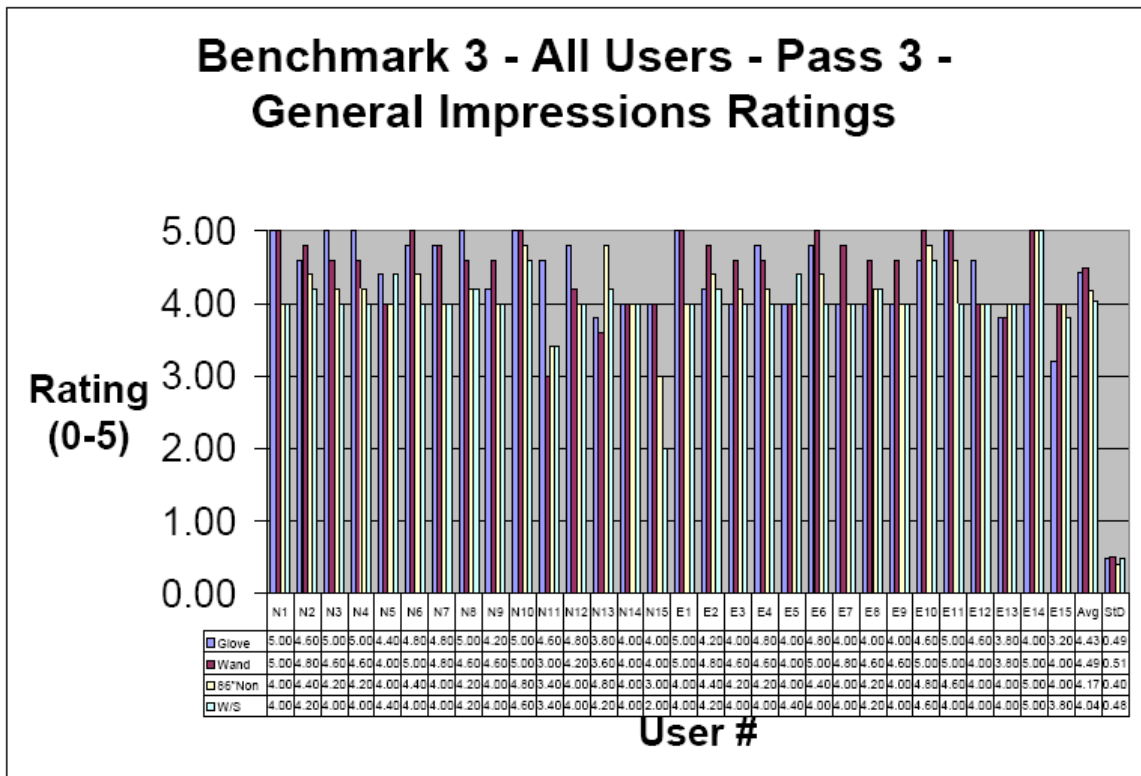
Benchmark 3, All Users, Pass 3, Locating Ratings											
B3Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.45	0.47	3.40	5.00	0.0002	No	10.55%			
Wand	30	4.51	0.44	3.00	5.00	0.0046	No	9.85%			
86\"Non	30	4.15	0.38	3.00	5.00	< 0.0001	No	9.27%			
W/S	30	4.03	0.46	2.00	5.00	< 0.0001	No	11.51%			
B3Ap3		Bartlett's Test		Levene's Test		Equal Var ?		Roy's Greatest Root:		F(3,27) 7.99	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
Glove vs Wand				0.1900	0.6636	Yes	29	6.92	0.0198	No	Means
Glove vs 86\"Non				1.1300	0.2927	Yes	29	19.03	0.0006	No	Means
Glove vs W/S				0.0000	0.9727	Yes	29	10.02	0.0069	No	Means
Wand vs 86\"Non				0.5500	0.4618	Yes	29	3.19	0.0958	No	Means
Wand vs W/S				0.0100	0.9056	Yes	29	0.00	0.9559	Yes	Neither
86\"Non vs W/S				0.1900	0.6649	Yes	29	1.40	0.2508	Yes	Neither



**Figure I- 22: B3Ap3Mov All Users Pass 3 Manipulation Ratings**

**Table I- 22: B3Ap3Mov All Users Pass 3 Manipulation Ratings Statistics**

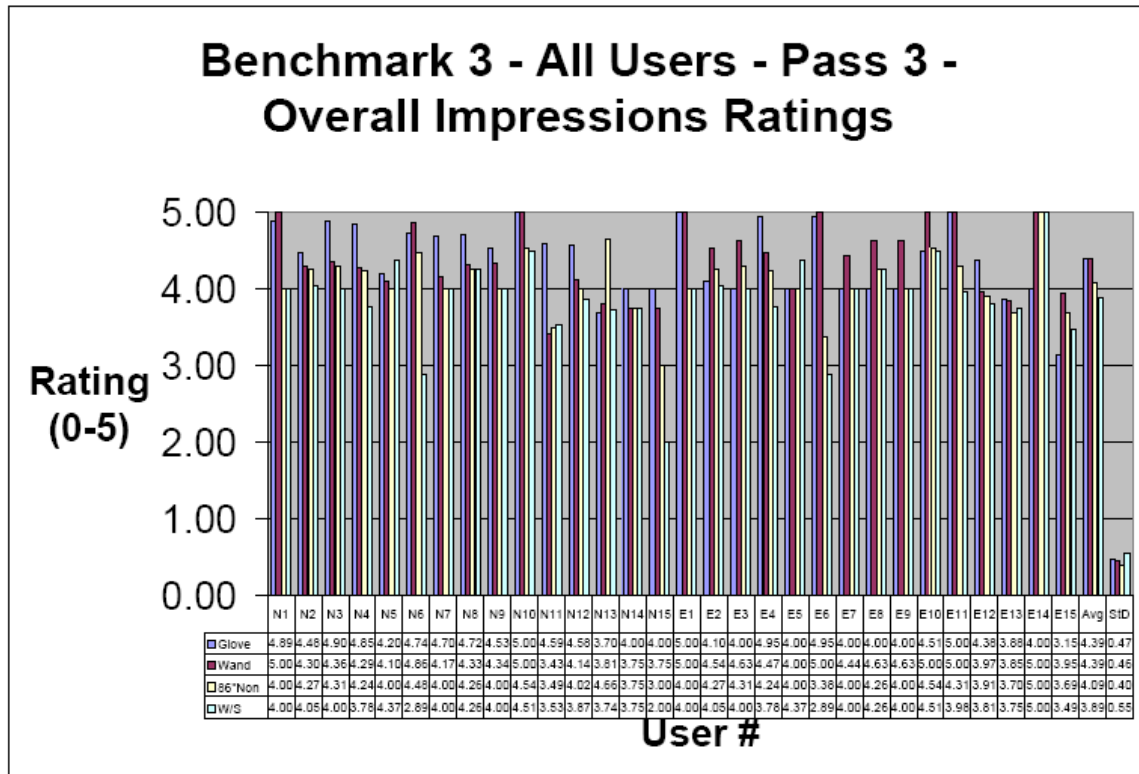
Benchmark 3, All Users, Pass 3, Manipulation Ratings											
B3Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.32	0.53	3.00	5.00	0.0019	No	12.38%			
Wand	30	4.13	0.73	2.71	5.00	0.0408	No	17.58%			
86\"Non	30	3.99	0.48	3.00	5.00	0.0010	No	12.14%			
W/S	30	3.80	0.61	2.00	5.00	0.0035	No	16.12%			
				Roy's Greatest Root:		F(3,27)	5.44	Pr > F	0.0046		
B3Ap3	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			3.6400	0.0612	Yes	29	1.82	0.0198	No	Means	Glove 4.42%
Glove vs 86\"Non			0.8600	0.3568	Yes	29	8.61	0.0006	No	Means	Glove 7.71%
Glove vs W/S			2.9200	0.0928	Yes	29	16.84	0.0069	No	Means	Glove 12.04%
Wand vs 86\"Non			0.2100	0.6481	Yes	29	1.27	0.0958	No	Means	Wand 3.44%
Wand vs W/S			1.6900	0.1989	Yes	29	5.71	0.0959	Yes	Neither	Neither
86\"Non vs W/S			0.4800	0.4912	Yes	29	6.02	0.2508	Yes	Neither	Neither



**Figure I- 23: B3Ap3Gen All users Pass 3 General Impressions Ratings**

**Table I- 23: B3Ap3Gen All users Pass 3 General Impressions Ratings  
Statistics**

Benchmark 3, All Users, Pass 3, General Impressions Ratings												
B3Ap3	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	CV
Glove	30	4.43		0.49		3.20		5.00		0.0159	No	11.06%
Wand	30	4.49		0.51		3.00		5.00		< 0.0001	No	11.39%
86"Non	30	4.17		0.40		3.00		5.00		< 0.0001	No	9.56%
W/S	30	4.04		0.48		2.00		5.00		< 0.0001	No	11.80%
B3Ap3						Roy's Greatest Root: F(3,27) 6.30 Pr > F 0.0022						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.0500	0.8204	Yes	29	0.31	0.5841	Yes	Neither	Neither
Glove vs 86"Non				1.1500	0.2887	Yes	29	5.92	0.0218	No	Means	Glove 5.87%
Glove vs W/S				0.0100	0.9334	Yes	29	11.11	0.0024	No	Means	Glove 8.87%
Wand vs 86"Non				1.1600	0.0858	Yes	29	14.30	0.0007	No	Means	Wand 7.00%
Wand vs W/S				0.0400	0.8371	Yes	29	19.78	0.0001	No	Means	Wand 9.96%
86"Non vs W/S				0.1900	0.6633	Yes	29	7.07	0.0126	No	Means	86"Non 3.19%

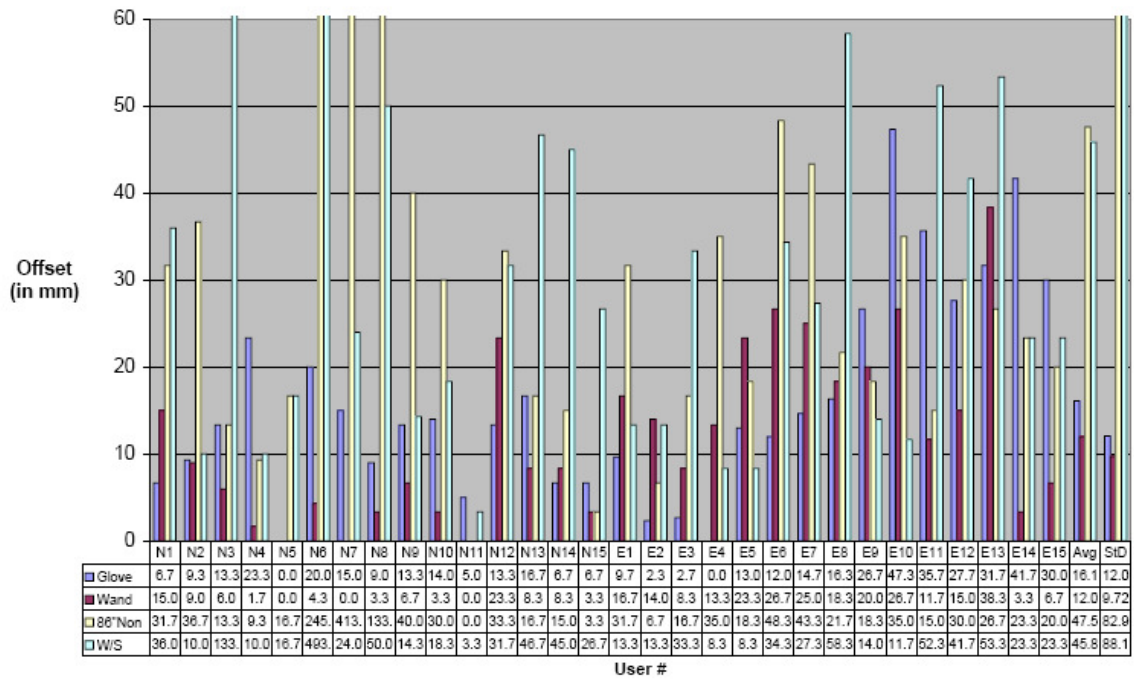


**Figure I- 24: B3Ap3Ovr All Users Pass 3 Overall Impressions Ratings**

**Table I- 24: B3Ap3Ovr All Users Pass 3 Overall Impressions Ratings  
Statistics**

Benchmark 3, All Users, Pass 3, Overall Impressions Ratings											
B3Ap3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	4.31	0.56	3.06	5.00	> 0.1000	Yes	12.96%			
Wand	30	4.11	0.52	3.36	5.00	< 0.0001	No	12.73%			
86*Non	30	3.86	0.32	3.00	4.74	0.0080	No	8.32%			
W/S	30	3.72	0.41	2.87	5.00	0.0778	Yes	10.95%			
						Roy's Greatest Root:		F(3,27)	3.97	Pr > F	0.0183
B3Ap3	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			0.2900	0.6296	Yes	29	0.07	0.8033	Yes	Neither	Neither
Glove vs 86*Non			9.1700	0.0037	No	29	6.13	0.0194	No	Var/Mean	Glove 10.37%
Glove vs W/S		2.7950	0.0946		Yes	29	9.46	0.0046	No	Means	Glove 13.64%
Wand vs 86*Non			6.6900	0.0122	No	29	9.59	0.0043	No	Var/Mean	Wand 5.96%
Wand vs W/S			1.6900	0.1993	Yes	29	11.60	0.0019	No	Means	Wand 9.40%
86*Non vs W/S			0.7500	0.3902	Yes	29	7.79	0.0092	No	Means	86*Non 3.65%

**Benchmark 3 - All Users - 3 Pass Avg. - Part 1 Offsets**

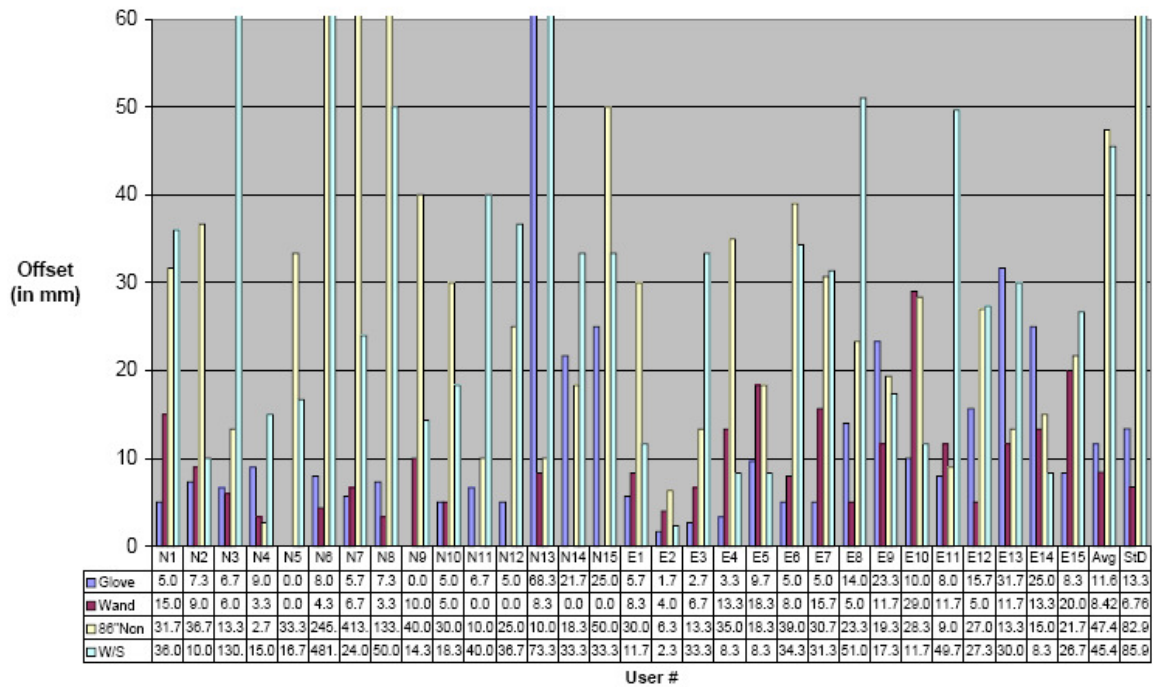


**Figure I- 25: B3A3pA-1Off All Users 3 Pass Avg. Part 1 Offsets**

**Table I- 25: B3A3pA-1Off All Users 3 Pass Avg. Part 1 Offsets Statistics**

Benchmark 3, All Users, 3 Pass Avg., Part 1 Offsets											
B3A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV			
Glove	30	16.12	12.08	0.00	47.33	0.0125	No	74.94%			
Wand	30	12.00	9.72	0.00	38.33	0.0654	Yes	80.99%			
86*Non	30	47.59	82.97	0.00	413.30	< 0.0001	No	174.34%			
W/S	30	45.86	88.10	3.33	493.30	< 0.0001	No	192.11%			
B3A3pA	Homogeneity of Variance					Roy's Greatest Root:			F(3,27) 2.38 Pr > F 0.0921		
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
	Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand			1.2100	0.2767	Yes	29	2.93	0.0976	No	Means	Wand 25.56%
Glove vs 86*Non			2.0300	0.1594	Yes	29	4.23	0.0488	No	Means	Glove 195.22%
Glove vs W/S			1.2200	0.2730	Yes	29	3.43	0.0744	No	Means	Glove 64.85%
Wand vs 86*Non			2.0600	0.1563	Yes	29	5.15	0.0308	No	Means	Wand 74.78%
Wand vs W/S			1.2400	0.2698	Yes	29	4.26	0.0482	No	Means	Wand 73.83%
86*Non vs W/S			0.0100	0.9166	Yes	29	0.01	0.9184	Yes	Neither	Neither

**Benchmark 3 - All Users - 3 Pass Avg. - Part 2 Offsets**

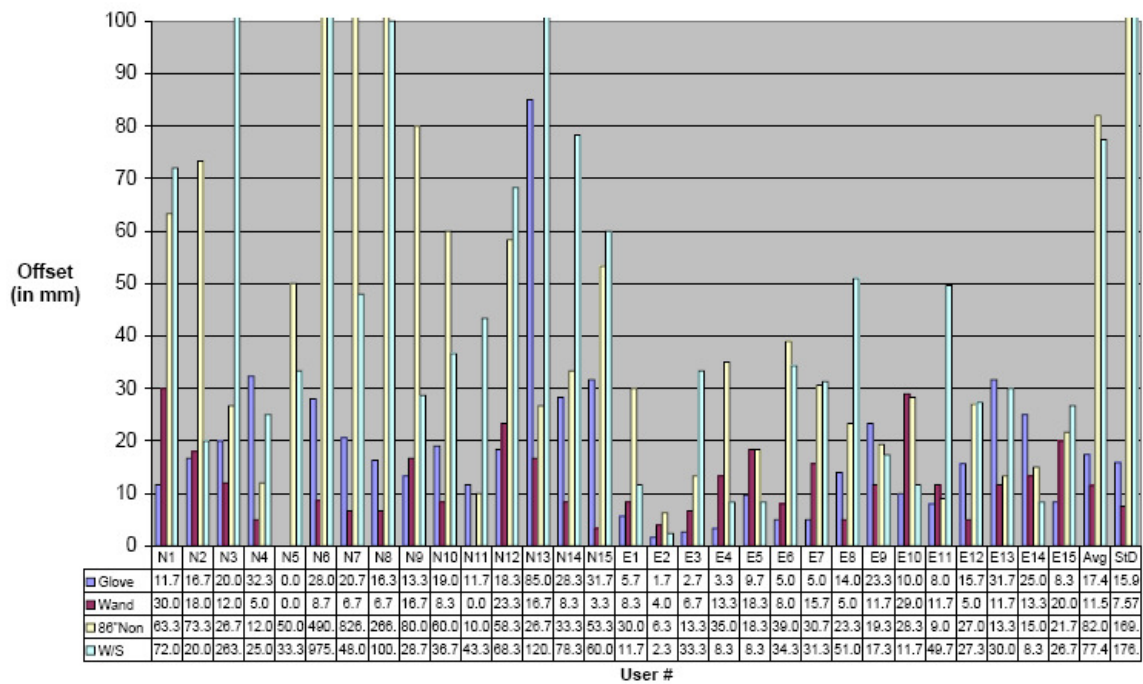


**Figure I- 26: B3A3pA-2Off All Users 3 Pass Avg. Part 2 Offsets**

**Table I- 26: B3A3pA-2Off All Users 3 Pass Avg. Part 2 Offsets Statistics**

Benchmark 3, All Users, 3 Pass Avg., Part 2 Offsets												
B3A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	11.66	13.36	0.00	68.30	< 0.0001	No	114.58%				
Wand	30	8.42	6.76	0.00	29.00	> 0.1000	Yes	80.27%				
86"Non	30	47.40	82.95	2.70	413.30	0.0061	No	175.00%				
W/S	30	45.47	85.99	2.30	481.70	0.0219	No	189.11%				
B3A3pA	Homogeneity of Variance					Roy's Greatest Root:			F(3,27)	2.38	Pr > F	0.0921
	Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand			0.7700	0.3836	Yes	29	2.93	0.0976	No	Means	Wand	27.79%
Glove vs 86"Non			2.0900	0.1540	Yes	29	4.23	0.0488	No	Means	Glove	306.52%
Glove vs W/S			1.2500	0.2689	Yes	29	3.43	0.0744	No	Means	Glove	289.97%
Wand vs 86"Non			2.1000	0.1522	Yes	29	5.15	0.0308	No	Means	Wand	462.95%
Wand vs W/S			1.2600	0.2671	Yes	29	4.26	0.0482	No	Means	Wand	81.48%
86"Non vs W/S			0.0100	0.9375	Yes	29	0.01	0.9184	Yes	Neither	Neither	

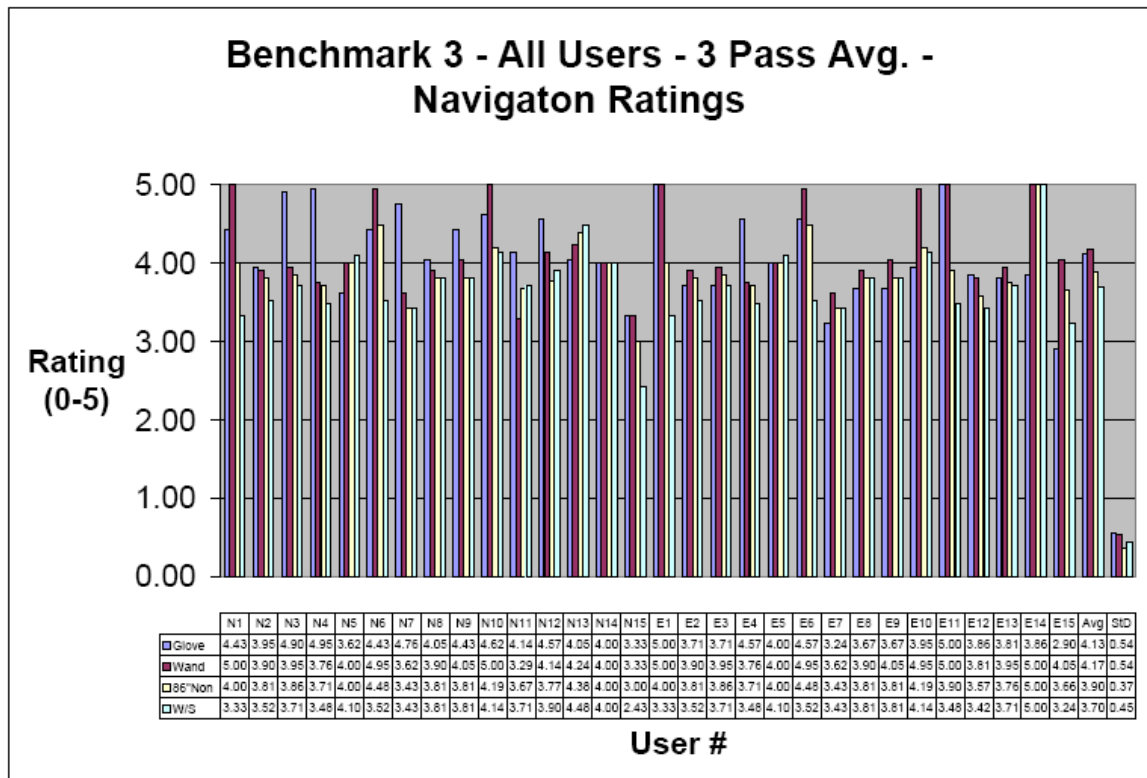
**Benchmark 3 - All Users - 3 Pass Avg. - Total Offsets**



**Figure I- 27: B3p3A-totOff All Users 3 Pass Avg. Total Offsets**

**Table I- 27: B3p3A-totOff All Users 3 Pass Avg. Total Offsets Statistics**

Benchmark 3, All Users, 3 Pass Avg., Total Offsets										
B3A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	30	20.43	14.86	0.00	85.00	0.0188	No	72.74%		
Wand	30	15.03	9.03	0.00	23.33	< 0.0001	No	60.05%		
86*Non	30	84.65	168.50	10.00	826.70	< 0.0001	No	199.05%		
W/S	30	80.57	175.70	2.33	975.00	< 0.0001	No	218.07%		
B3A3pA	Homogeneity of Variance				Roy's Greatest Root:				F(3,27)	Pr > F
	Bartlett's Test		Levene's Test		F Value		Pr > F	Equal Var ?	df	
	Chi-Sq	Pr > ChiSq	F Value	Pr > F						
Glove vs Wand			0.9700	0.3322	Yes	29	2.05	0.1737	Yes	Neither
Glove vs 86*Non			2.4000	0.1320	Yes	29	6.94	0.0196	No	Means
Glove vs W/S			2.5200	0.1234	Yes	29	0.88	0.3653	Yes	Neither
Wand vs 86*Non			3.7200	0.0640	Yes	29	0.33	0.5744	Yes	Neither
Wand vs W/S			3.1900	0.0849	Yes	29	2.44	0.1409	Yes	Neither
86*Non vs W/S			0.4100	0.5281	Yes	29	8.59	0.0110	No	Means

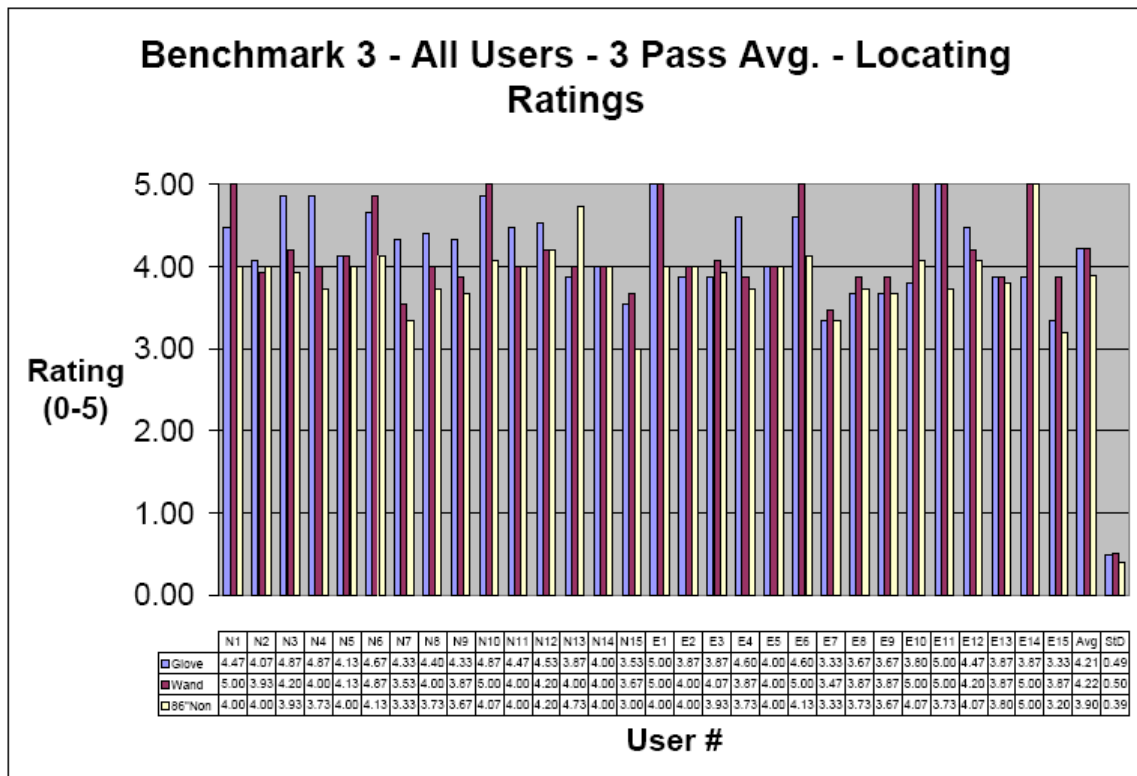


**Figure I- 28: B3A3pAnav All Users 3 Pass Avg. Navigation Ratings**

**Table I- 28: B3A3pAnav All Users 3 Pass Avg. Navigation Ratings Statistics**

Benchmark 3, All Users, 3 Pass Avg., Navigation Ratings												
B3A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	4.13	0.54	2.91	5.00	> 0.1000	Yes	13.21%				
Wand	30	4.17	0.54	3.29	5.00	< 0.0001	No	12.91%				
86"Non	30	3.90	0.37	3.00	5.00	0.0072	No	9.48%				
W/S	30	3.70	0.45	2.43	5.00	> 0.1000	Yes	12.06%				
B3A3pA		Homogeneity of Variance				Roy's Greatest Root: F(3,27) 14.13				Pr > F 0.0003		
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.0100	0.9314	Yes	29	5.29	0.0374	No	Means	Wand 1.03%
Glove vs 86"Non				3.8500	0.0545	Yes	29	6.73	0.0213	No	Means	Glove 5.53%
Glove vs W/S		0.9843	0.3211			Yes	29	11.65	0.0042	No	Means	Glove 10.23%
Wand vs 86"Non				0.9200	0.3402	Yes	29	13.82	0.0023	No	Means	Wand 6.50%
Wand vs W/S				4.1500	0.0461	No	29	5.49	0.0344	No	Var/Means	Wand 11.16%
86"Non vs W/S				0.4600	0.5011	Yes	29	16.49	0.0012	No	Means	86"Non 4.98%

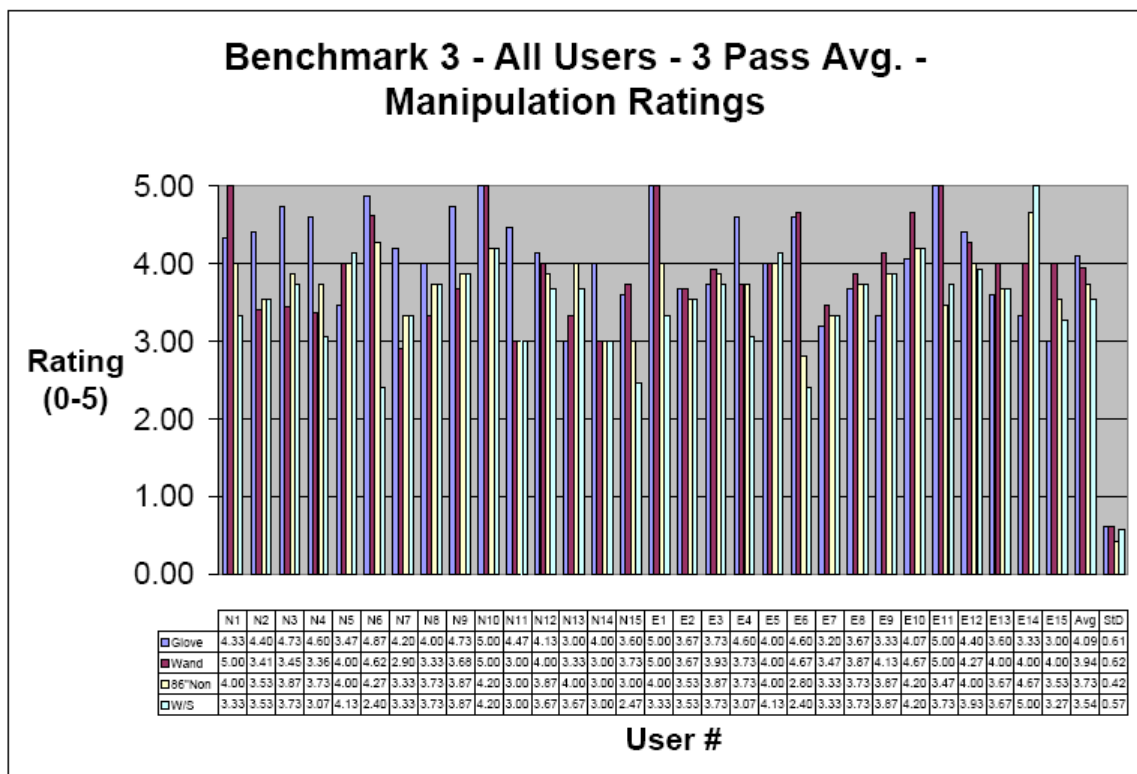




**Figure I- 29: B3A3pAloc All Users 3 Pass Avg. Locating Ratings**

**Table I- 29: B3A3pAloc All Users 3 Pass Avg. Locating Ratings**

Benchmark 3, All Users, 3 Pass Avg., Locating Ratings										
B3A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV		
Glove	30	4.06	0.54	3.33	5.00	> 0.1000	Yes	13.42%		
Wand	30	4.06	0.54	3.47	5.00	< 0.0001	No	13.25%		
86*Non	30	3.78	0.37	3.00	5.00	0.0207	No	9.78%		
W/S	30	3.66	0.45	2.53	5.00	0.0085	No	12.21%		
B3A3pA		Homogeneity of Variance				Roy's Greatest Root:				Pr > F
		Bartlett's Test	Levene's Test	Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	
		Chi-Sq	F Value	Pr > F						Statistically Better
Glove vs Wand			0.0300	0.8664	Yes	29	5.17	0.0393	No	Means
Glove vs 86*Non			1.4300	0.2372	Yes	29	8.69	0.0106	No	Means
Glove vs W/S			0.6300	0.4299	Yes	29	10.97	0.0051	No	Means
Wand vs 86*Non			1.7100	0.1937	Yes	29	5.98	0.0298	No	Means
Wand vs W/S			0.8200	0.3692	Yes	29	13.97	0.0022	No	Means
86*Non vs W/S			0.0400	0.8511	Yes	29	15.61	0.0014	No	Means

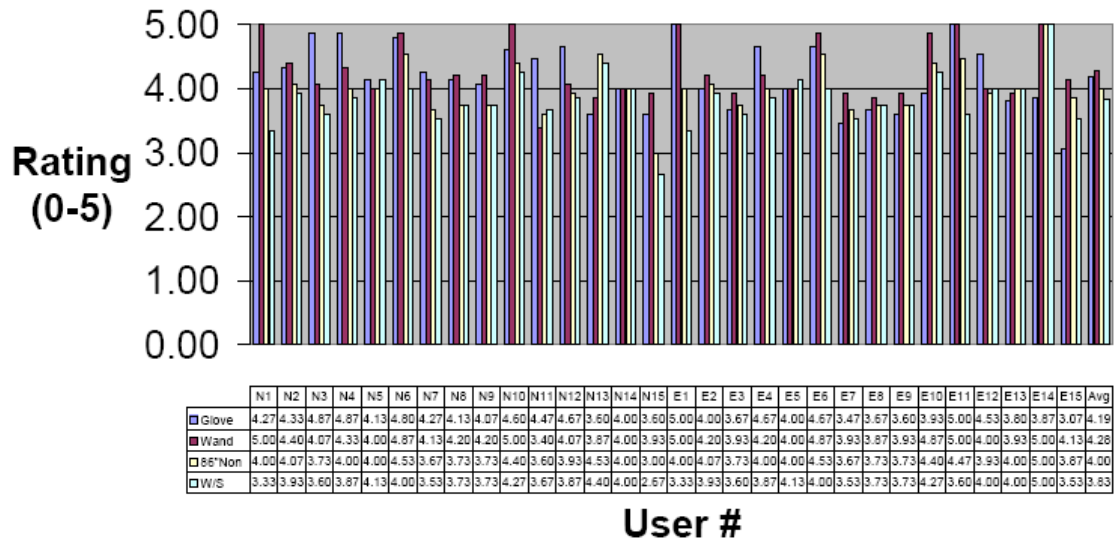


**Figure I- 30: B3A3pAmov All Users 3 Pass Avg. Manipulation Ratings**

**Table I- 30: B3A3pAmov All Users 3 Pass Avg. Manipulation Ratings Statistics**

Benchmark 3, All Users, 3 Pass Avg., Manipulation Ratings												
B3A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	4.09	0.61	3.00	5.00	> 0.1000	Yes	14.91%				
Wand	30	3.94	0.62	2.90	5.00	0.0437	No	15.69%				
86"Non	30	3.73	0.42	2.80	4.67	> 0.1000	Yes	11.28%				
W/S	30	3.66	0.51	2.47	5.00	> 0.1000	Yes	13.84%				
B3A3pA		Homogeneity of Variance				Roy's Greatest Root: F(3,27) 3.60 Pr > F 0.0263						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.0100	0.9207	Yes	29	1.37	0.2508	Yes	Neither	Neither
Glove vs 86"Non		3.8636	0.0493			No	29	7.34	0.0112	No	Var/Mean	Glove 8.90%
Glove vs W/S		0.1542	0.6945			Yes	29	10.88	0.0026	No	Means	Glove 10.63%
Wand vs 86"Non				5.0400	0.0286	No	29	3.93	0.0569	No	Var/Mean	Wand 5.43%
Wand vs W/S				0.2500	0.6710	Yes	29	7.95	0.0086	No	Means	Wand 7.23%
86"Non vs W/S		2.5070	0.1133			Yes	29	6.24	0.0184	No	Means	86"Non 1.91%

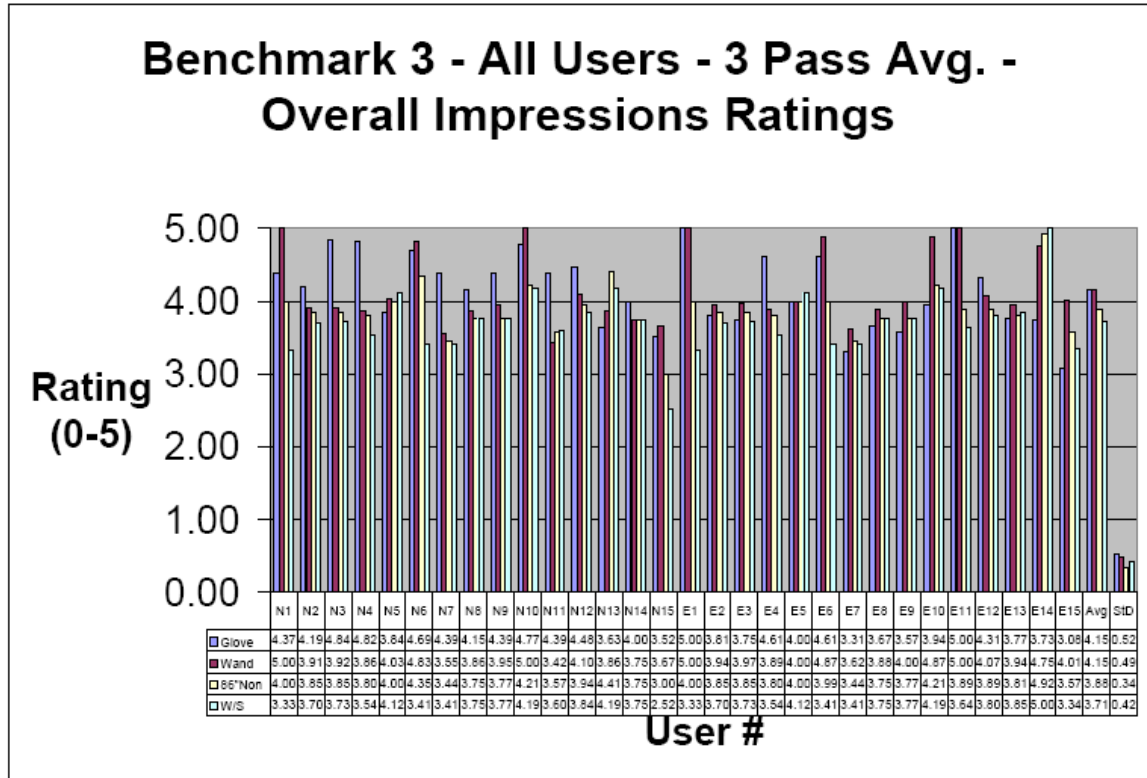
## Benchmark 3 - All Users - 3 Pass Avg. - General Impressions Ratings



**Figure I- 31: B3A3pAgen All Users 3 Pass Avg. General Impressions Ratings**

**Table I- 31: B3A3pAgen All Users 3 Pass Avg. General Impressions Ratings**

Benchmark 3, All Users, 3 Pass Avg., General Impressions Ratings													
B3A3pA	# Users	Mean		St. Dev.		Low		High		P Value	Normal?	CV	
Glove	30	4.19		0.51		3.07		5.00		> 0.1000	Yes	12.15%	
Wand	30	4.28		0.45		3.40		5.00		0.0002	No	10.48%	
86"Non	30	4.00		0.38		3.00		4.67		0.0029	No	9.55%	
W/S	30	3.83		0.41		2.67		5.00		> 0.1000	Yes	10.57%	
B3A3pA		Homogeneity of Variance					Roy's Greatest Root: F(3,27) 6.44 Pr > F 0.002						
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better	
		Chi-Sq	Pr > ChiSq	F Value	Pr > F								
Glove vs Wand				0.7600	0.3870	Yes	29	0.97	0.3331	Yes	Neither	Neither	
Glove vs 86"Non		2.3013	0.1293			Yes	29	3.36	0.0769	No	Means	Glove 4.42%	
Glove vs W/S		1.4513	0.2283			Yes	29	9.01	0.0055	No	Means	Glove 8.45%	
Wand vs 86"Non				0.8000	0.3758	Yes	29	19.02	0.0001	No	Means	Wand 6.45%	
Wand vs W/S				0.2400	0.6238	Yes	29	20.58	<0.0001	Yes	Neither	Neither	
86"Non vs W/S		0.1007	0.7510			Yes	29	13.64	0.0009	No	Means	86"Non 4.22%	



**Figure I- 32: B3A3pAvgOvr All User 3 Pass Avg. Overall Impressions Ratings**

**Table I- 32: B3A3pAvgOvr All User 3 Pass Avg. Overall Impressions Ratings Statistics**

Benchmark 3, All Users, 3 Pass Avg., Overall Impressions Ratings												
B3A3pA	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV				
Glove	30	4.08	0.55	3.04	5.00	> 0.1000	Yes	13.43%				
Wand	30	4.07	0.52	3.38	5.00	0.0012	No	12.88%				
86*Non	30	3.82	0.36	3.00	4.88	> 0.1000	Yes	9.47%				
W/S	30	3.68	0.42	2.69	5.00	> 0.1000	Yes	11.38%				
B3A3pA		Homogeneity of Variance				Roy's Greatest Root:			F(3,27)	4.52	Pr > F	0.0108
		Bartlett's Test		Levene's Test		Equal Var ?	df	F(1,29)	Pr > F	Equal Means?	Significant?	Statistically Better
		Chi-Sq	Pr > ChiSq	F Value	Pr > F							
Glove vs Wand				0.1000	0.7519	Yes	29	0.01	0.9411	Yes	Neither	Neither
Glove vs 86*Non		4.7869	0.0087			No	29	5.70	0.0237	No	Var/Mean	Glove 6.40%
Glove vs W/S		2.0420	0.1530			Yes	29	9.68	0.0042	No	Means	Glove 9.83%
Wand vs 86*Non				4.1500	0.0461	No	29	11.90	0.0017	No	Var/Mean	Wand 6.21%
Wand vs W/S				1.3300	0.2529	Yes	29	13.51	0.0010	No	Means	Wand 9.65%
86*Non vs W/S		0.6080	0.4356			Yes	29	10.65	0.0028	No	Means	86*Non 3.66%

## **Appendix J**

### **Benchmark 1 Statistically Significant Environment Comparisons**

Tables J-1 (Novice), J-2 (Experienced), and J-3 (All) present a compilation of the statistically significant findings of the overall Benchmark 1 testing. The tabulation includes the significance of both the objective (elapsed time) measures and the subjective user ratings by category. Overall the tables show that 285 of the 432 Benchmark 1 elapsed time and ratings comparisons produce statistically significant results (65.97%).

Table J-1 shows that 92 of the 144 novice user Benchmark 1 elapsed times and ratings comparisons produced statistically significant results (63.89%).

**Table J- 1 Benchmark 1 Novice Users Statistically Significant Comparisons**

Benchmark 1 - Novice Users				
Statistically Significant Environment Comparisons Results Summary				
Elapsed Times 45.83%	Pass 1 50%	Pass 2 33.33%	Pass 3 66.67%	3 Pass Avg. 33.33%
Glove vs Wand	Wand 14.04%	Neither	Wand 23.50%	Wand 23.76%
Glove vs 86" Non	86" Non 5.65%	Neither	Neither	Neither
Glove vs W/S	Neither	Neither	Neither	Neither
Wand vs 86" Non	Neither	Wand 23.92%	Wand 13.15%	Neither
Wand vs W/S	Wand 14.84%	Wand 37.70%	Wand 14.25%	Wand 21.10%
86" Non vs W/S	Neither	Neither	86" Non 0.98%	Neither
Navigation Ratings 50%	Pass 1 33.33%	Pass 2 50%	Pass 3 50%	3 Pass Avg. 66.67%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86" Non	Neither	Neither	Glove 9.92%	Glove 7.84%
Glove vs W/S	Glove 15.11%	Glove 15.93%	Glove 16.87%	Glove 15.87%
Wand vs 86" Non	Neither	Neither	Neither	Neither
Wand vs W/S	Neither	Wand 13.68%	Wand 9.68%	Wand 11.04%
86" Non vs W/S	86" Non 7.67%	86" Non 8.39%	Neither	86" Non 7.44%
Locating Ratings 54.17%	Pass 1 66.67%	Pass 2 66.67%	Pass 3 50%	3 Pass Avg. 33.33%
Glove vs Wand	Wand 13.71%	Glove 7.43%	Neither	Neither
Glove vs 86" Non	Glove 5.85%	Glove 8.41%	Glove 10.96%	Glove 8.43%
Glove vs W/S	Glove 8.08%	Glove 13.36%	Glove 13.22%	Glove 10.44%
Wand vs 86" Non	Neither	Neither	Neither	Neither
Wand vs W/S	Wand 22.89%	Wand 5.52%	Wand 6.79%	Neither
86" Non vs W/S	Neither	Neither	Neither	Neither
Manipulation Ratings 70.83%	Pass 1 66.67%	Pass 2 50%	Pass 3 83.33%	3 Pass Avg. 83.33%
Glove vs Wand	Neither	Glove 11.42%	Glove 12.00%	Glove 8.91%
Glove vs 86" Non	Glove 8.08%	Glove 14.57%	Glove 19.56%	Glove 14.20%
Glove vs W/S	Glove 15.29%	Glove 16.66%	Glove 24.44%	Glove 18.85%
Wand vs 86" Non	Neither	Neither	Wand 6.75%	Wand 4.86%
Wand vs W/S	Wand 11.76%	Neither	Wand 11.11%	Wand 9.13%
86" Non vs W/S	86" Non 6.68%	Neither	Neither	Neither
General Impressions 75%	Pass 1 66.67%	Pass 2 83.33%	Pass 3 66.67%	3 Pass Avg. 83.33%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86" Non	Glove 11.84%	Glove 12.56%	Glove 12.31%	Glove 12.28%
Glove vs W/S	Glove 19.32%	Glove 20.73%	Glove 22.18%	Glove 20.79%
Wand vs 86" Non	Wand 5.23%	Wand 5.09%	Neither	Wand 6.09%
Wand vs W/S	Wand 12.27%	Wand 12.71%	Wand 17.24%	Wand 14.13%
86" Non vs W/S	Neither	86" Non 7.25%	86" Non 8.79%	86" Non 7.58%
Overall Impressions 87.5%	Pass 1 66.67%	Pass 2 83.33%	Pass 3 100%	3 Pass Avg. 100%
Glove vs Wand	Neither	Glove 6.54%	Glove 7.89%	Glove 7.55%
Glove vs 86" Non	Glove 8.23%	Glove 10.82%	Glove 13.69%	Glove 10.83%
Glove vs W/S	Glove 14.79%	Glove 17.80%	Glove 21.78%	Glove 17.76%
Wand vs 86" Non	Neither	Neither	Wand 5.37%	Wand 2.86%
Wand vs W/S	Wand 9.40%	Wand 10.57%	Wand 12.87%	Wand 9.50%
86" Non vs W/S	86" Non 6.06%	86" Non 6.30%	86" Non 7.11%	86" Non 6.45%
92 Statistically Significant Results of 144 Comparisons (63.89%)				

Table J-2 shows that the experienced user group test environment comparisons tests provide a somewhat lower percentage (63.19%) of statistically significant results.

**Table J- 2 Benchmark 1 Experienced Users Statistically Significant Comparisons**

Benchmark 1 - Experienced Users				
Statistically Significant Environment Comparisons Results Summary				
Elapsed Times 45.83%	Pass 1 33.33%	Pass 2 50%	Pass 3 50%	3 Pass Avg. 50%
Glove vs Wand	Neither	Wand 34.06%	Wand 57.18%	Wand 31.49%
Glove vs 86"Non	Glove 18.18%	Neither	Neither	Neither
Glove vs W/S	Neither	Neither	Neither	Neither
Wand vs 86"Non	Wand 32.21%	Wand 50.02%	Wand 54.25%	Wand 44.09%
Wand vs W/S	Neither	Wand 34.10%	Wand 54.58%	Wand 37.50%
86"Non vs W/S	Neither	Neither	Neither	Neither
Navigation Ratings 58.33%	Pass 1 66.67%	Pass 2 66.67%	Pass 3 50%	3 Pass Avg. 50%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Glove 8.32%	Glove 12.43%	Glove 9.11%	Glove 9.92%
Glove vs W/S	Glove 2.84%	Glove 11.29%	Neither	Neither
Wand vs 86"Non	Wand 12.45%	Wand 11.64%	Wand 12.90%	Wand 12.33%
Wand vs W/S	Neither	Wand 10.50%	Wand 9.51%	Wand 8.91%
86"Non vs W/S	W/S 5.32%	Neither	Neither	Neither
Locating Ratings 66.67%	Pass 1 66.67%	Pass 2 66.67%	Pass 3 66.67%	3 Pass Avg. 66.67%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Glove 18.32%	Glove 15.22%	Glove 14.38%	Glove 13.40%
Glove vs W/S	Glove 15.78%	Glove 15.63%	Glove 10.96%	Glove 11.58%
Wand vs 86"Non	Wand 10.99%	Wand 9.43%	Wand 10.97%	Wand 10.43%
Wand vs W/S	Wand 8.60%	Wand 9.82%	Wand 7.65%	Wand 8.66%
86"Non vs W/S	Neither	Neither	Neither	Neither
Manipulation Ratings 79.17%	Pass 1 66.67%	Pass 2 83.33%	Pass 3 83.33%	3 Pass Avg. 83.33%
Glove vs Wand	Glove 12.18%	Neither	Neither	Glove 8.84%
Glove vs 86"Non	Glove 21.24%	Glove 16.51%	Glove 31.09%	Glove 24.67%
Glove vs W/S	Glove 13.85%	Glove 10.20%	Glove 20.65%	Glove 17.65%
Wand vs 86"Non	Wand 8.08%	Wand 14.92%	Wand 24.00%	Wand 14.55%
Wand vs W/S	Neither	Wand 8.69%	Wand 14.13%	Wand 8.10%
86"Non vs W/S	Neither	W/S 5.73%	W/S 8.65%	Neither
General Impressions 70.83%	Pass 1 66.67%	Pass 2 66.67%	Pass 3 83.33%	3 Pass Avg. 83.33%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Glove 20.30%	Glove 16.54%	Glove 20.73%	Glove 18.80%
Glove vs W/S	Glove 13.88%	Glove 14.03%	Glove 13.81%	Glove 11.60%
Wand vs 86"Non	Wand 18.69%	Wand 13.23%	Wand 16.07%	Wand 11.71%
Wand vs W/S	Wand 12.36%	Wand 10.79%	Wand 9.42%	Wand 6.74%
86"Non vs W/S	Neither	Neither	W/S 6.08%	Neither
Overall Impressions 58.33%	Pass 1 50%	Pass 2 66.67%	Pass 3 66.67%	3 Pass Avg. 50%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Glove 15.48%	Glove 17.71%	Glove 17.99%	Glove 16.55%
Glove vs W/S	Neither	Glove 12.61%	Glove 12.79%	Neither
Wand vs 86"Non	Wand 12.58%	Wand 13.50%	Wand 14.20%	Wand 13.27%
Wand vs W/S	Wand 7.22%	Wand 8.58%	Wand 9.17%	Wand 8.26%
86"Non vs W/S	Neither	Neither	Neither	Neither
91 Statistically Significant Results of 144 Comparisons (63.19%)				

Similarly, Table J-3 shows that the all users Benchmark 1 elapsed times and ratings comparisons produced statistically significant results in 102 of the 144 elapsed times and ratings environment comparisons (70.83%).

**Table J- 3 Benchmark 1 All Users Statistically Significant Comparisons**

Benchmark 1 - All Users				
Statistically Significant Environment Comparisons Results Summary				
Elapsed Times 50%	Pass 1 50%	Pass 2 50%	Pass 3 50%	3 Pass Avg. 50%
Glove vs Wand	Neither	Wand 36.40%	Wand 41.98%	Wand 27.92%
Glove vs 86" Non	86" Non 4.65%	Neither	Neither	Neither
Glove vs W/S	Neither	Neither	Neither	Neither
Wand vs 86" Non	Wand 35.67%	Wand 38.84%	Wand 35.67%	Wand 30.06%
Wand vs W/S	Wand 36.09%	Wand 35.68%	Wand 36.09%	Wand 29.89%
86" Non vs W/S	Neither	Neither	Neither	Neither
<b>Navigation Ratings 70.83%</b>	<b>Pass 1 66.67%</b>	<b>Pass 2 83.33%</b>	<b>Pass 3 66.67%</b>	<b>3 Pass Avg. 66.67%</b>
Glove vs Wand	Wand 0.52%	Neither	Neither	Neither
Glove vs 86" Non	Glove 7.56%	Glove 9.61%	86" Non 9.52%	Glove 8.85%
Glove vs W/S	Neither	W/S 13.61%	W/S 11.23%	Glove 11.13%
Wand vs 86" Non	Wand 7.00%	Wand 8.14%	Wand 7.91%	Wand 7.68%
Wand vs W/S	Wand 8.27%	Wand 12.08%	Wand 9.59%	Wand 9.94%
86" Non vs W/S	Neither	W/S 3.65%	Neither	Neither
<b>Locating Ratings 70.83%</b>	<b>Pass 1 66.67%</b>	<b>Pass 2 83.33%</b>	<b>Pass 3 66.67%</b>	<b>3 Pass Avg. 66.67%</b>
Glove vs Wand	Neither	Wand 6.33%	Neither	Neither
Glove vs 86" Non	Glove 11.89%	Glove 11.70%	86" Non 12.65%	Glove 10.85%
Glove vs W/S	W/S 11.89%	W/S 14.49%	W/S 12.08%	Glove 10.99%
Wand vs 86" Non	Wand 7.79%	Wand 5.05%	Wand 7.77%	Wand 6.84%
Wand vs W/S	Wand 7.79%	Wand 7.67%	Wand 7.22%	Wand 6.98%
86" Non vs W/S	Neither	Neither	Neither	Neither
<b>Manipulation Ratings 83.33%</b>	<b>Pass 1 83.33%</b>	<b>Pass 2 83.33%</b>	<b>Pass 3 83.33%</b>	<b>3 Pass Avg. 83.33%</b>
Glove vs Wand	Wand 7.55%	Wand 6.28%	Wand 8.78%	Wand 8.89%
Glove vs 86" Non	Neither 13.91%	Glove 15.51%	86" Non 24.72%	Glove 19.17%
Glove vs W/S	W/S 14.57%	W/S 13.42%	W/S 22.53%	Glove 17.61%
Wand vs 86" Non	Wand 5.91%	Wand 8.69%	Wand 14.65%	Wand 9.44%
Wand vs W/S	Wand 6.52%	Wand 6.72%	Wand 12.64%	Wand 8.01%
86" Non vs W/S	Neither	Neither	Neither	Neither
<b>General Impressions 70.83%</b>	<b>Pass 1 66.67%</b>	<b>Pass 2 66.67%</b>	<b>Pass 3 66.67%</b>	<b>3 Pass Avg. 83.33%</b>
Glove vs Wand	Neither	Neither	Neither	Wand 5.20%
Glove vs 86" Non	Glove 15.89%	Glove 14.47%	Glove 16.30%	Glove 14.40%
Glove vs W/S	Glove 16.53%	Glove 17.36%	Glove 17.92%	Glove 16.20%
Wand vs 86" Non	Wand 10.47%	Wand 8.99%	Wand 11.71%	Wand 8.74%
Wand vs W/S	Wand 11.07%	Wand 11.74%	Wand 13.27%	Wand 10.46%
86" Non vs W/S	Neither	Neither	Neither	Neither
<b>Overall Impressions 79.17%</b>	<b>Pass 1 66.67%</b>	<b>Pass 2 83.33%</b>	<b>Pass 3 83.33%</b>	<b>3 Pass Avg. 83.33%</b>
Glove vs Wand	Neither	Glove 5.13%	Glove 5.63%	Wand 5.25%
Glove vs 86" Non	Glove 11.67%	Glove 14.05%	Glove 15.68%	Glove 13.43%
Glove vs W/S	Glove 12.36%	Glove 15.18%	Glove 17.26%	Glove 14.58%
Wand vs 86" Non	Wand 7.65%	Wand 8.49%	Wand 9.52%	Wand 7.77%
Wand vs W/S	Wand 8.30%	Wand 9.56%	Wand 11.01%	Wand 8.87%
86" Non vs W/S	Neither	Neither	Neither	Neither
102 Statistically Significant Results of 144 Comparisons (70.83%)				



## **Appendix K**

### **Benchmark 2 Statistically Significant Environment Comparisons**

Tables K-1 (Novice), K-2 (Experienced), and K-3 (All) present a compilation of the statistically significant findings of the overall Benchmark 2 testing. The tabulation includes the significance of both the objective (elapsed time) measures and the subjective user ratings by category. Overall the tables show that 221 of the 432 Benchmark 2 elapsed time and ratings comparisons produce statistically significant results (51.2%).

Table K-1 shows that 78 of the 144 novice user Benchmark 2 elapsed times and ratings comparisons produced statistically significant results (54.17%).

**Table K- 1 Benchmark 2 Novice Users Statistically Significant Comparisons**

Benchmark 2 - Novice Users				
Statistically Significant Environment Comparisons Results Summary				
Elapsed Times 66.67%	Pass 1 66.67%	Pass 2 66.67%	Pass 3 50%	3 Pass Avg. 83.33%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Glove 17.66%	Glove 23.23%	Glove 31.48%	Glove 24.44%
Glove vs W/S	Glove 21.57%	Glove 17.62%	Glove 25.29%	Glove 21.09%
Wand vs 86"Non	Wand 10.58%	Wand 23.23%	Wand 31.48%	Wand 24.44%
Wand vs W/S	Wand 14.82%	Wand 17.12%	Neither	Wand 21.36%
86"Non vs W/S	Neither	Neither	Neither	W/S 4.25%
Navigation Ratings 58.33%	Pass 1 33.33%	Pass 2 66.67%	Pass 3 83.33%	3 Pass Avg. 50%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Neither	Glove 10.13%	Glove 12.97%	Neither
Glove vs W/S	Neither	Glove 9.31%	Glove 15.80%	Glove 10.90%
Wand vs 86"Non	Wand 3.46%	Wand 6.56%	Wand 12.10%	Wand 8.97%
Wand vs W/S	Wand 5.90%	Wand 6.56%	Wand 12.10%	Wand 8.97%
86"Non vs W/S	Neither	Neither	86"Non 2.51%	Neither
Locating Ratings 58.33%	Pass 1 33.33%	Pass 2 66.67%	Pass 3 66.67%	3 Pass Avg. 66.67%
Glove vs Wand	Neither	Neither	Glove 3.80%	Neither
Glove vs 86"Non	Neither	Glove 8.67%	Glove 13.50%	Glove 8.50%
Glove vs W/S	Neither	Glove 7.55%	Glove 13.09%	Glove 9.38%
Wand vs 86"Non	Wand 5.79%	Wand 5.55%	Neither	Wand 6.87%
Wand vs W/S	Wand 9.91%	Wand 4.46%	Wand 8.95%	Wand 7.73%
86"Non vs W/S	Neither	Neither	Neither	Neither
Manipulation Ratings 25%	Pass 1 0%	Pass 2 16.67%	Pass 3 66.67%	3 Pass Avg. 16.67%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Neither	Glove 9.41%	Glove 12.50%	Glove 9.24%
Glove vs W/S	Neither	Neither	Glove 13.68%	Neither
Wand vs 86"Non	Neither	Neither	Wand 7.29%	Neither
Wand vs W/S	Neither	Neither	Wand 8.42%	Neither
86"Non vs W/S	Neither	Neither	Neither	Neither
General Impressions 50%	Pass 1 50%	Pass 2 33.33%	Pass 3 50%	3 Pass Avg. 66.67%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Neither	Neither	Neither	Glove 9.40%
Glove vs W/S	Glove 6.95%	Neither	Glove 12.88%	Glove 9.28%
Wand vs 86"Non	Wand 10.08%	Wand 8.25%	Wand 11.02%	Wand 9.73%
Wand vs W/S	Wand 10.08%	Wand 6.79%	Wand 12.18%	Wand 9.62%
86"Non vs W/S	Neither	Neither	Neither	Neither
Overall Impressions 66.67%	Pass 1 66.67%	Pass 2 66.67%	Pass 3 66.67%	3 Pass Avg. 66.67%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Glove 9.00%	Glove 9.88%	Glove 9.57%	Glove 9.44%
Glove vs W/S	Glove 9.14%	Glove 10.22%	Glove 9.69%	Glove 9.67%
Wand vs 86"Non	Wand 7.20%	Wand 9.36%	Wand 8.77%	Wand 8.43%
Wand vs W/S	Wand 7.34%	Wand 9.70%	Wand 8.88%	Wand 8.65%
86"Non vs W/S	Neither	Neither	Neither	Neither
78 Statistically Significant Results of 144 Comparisons (54.17%)				

Table K-2 shows that 64 of the 144 experienced user Benchmark 2 elapsed times and ratings comparisons produced statistically significant results (44.44%).

**Table K- 2 Benchmark 2 Experienced Users Statistically Significant Comparisons**

<b>Benchmark 2 - Experienced Users</b>				
<b>Statistically Significant Environment Comparisons Results Summary</b>				
<b>Elapsed Times</b> 54.17%	<b>Pass 1</b> 50%	<b>Pass 2</b> 50%	<b>Pass 3</b> 50%	<b>3 Pass Avg.</b> 66.67%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Neither	Glove 14.77%	Neither	Glove 18.28%
Glove vs W/S	Glove 13.47%	Glove 10.47%	Glove 20.79%	Glove 20.79%
Wand vs 86"Non	Wand 10.42%	Wand 13.24%	Wand 15.11%	Wand 15.11%
Wand vs W/S	Wand 16.38%	Neither	Wand 17.72%	Wand 17.72%
86"Non vs W/S	Neither	Neither	Neither	Neither
<b>Navigation Ratings</b> 45.83%	<b>Pass 1</b> 16.67%	<b>Pass 2</b> 66.67%	<b>Pass 3</b> 66.67%	<b>3 Pass Avg.</b> 33.33%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Neither	Glove 2.84%	Glove 7.66%	Glove 3.88%
Glove vs W/S	Glove 10.39%	Glove 2.58%	Glove 7.69%	Glove 4.31%
Wand vs 86"Non	Neither	Wand 4.89%	Wand 6.65%	Neither
Wand vs W/S	Neither	Wand 4.62%	Wand 6.68%	Neither
86"Non vs W/S	Neither	Neither	Neither	Neither
<b>Locating Ratings</b> 45.83%	<b>Pass 1</b> 50%	<b>Pass 2</b> 33.33%	<b>Pass 3</b> 66.67%	<b>3 Pass Avg.</b> 33.33%
Glove vs Wand	Wand 0.34%	Glove 1.68%	Neither	Neither
Glove vs 86"Non	Neither	Neither	Glove 6.91%	Neither
Glove vs W/S	Glove 2.11%	Glove 3.04%	Glove 7.97%	Glove 4.46%
Wand vs 86"Non	Neither	Neither	Wand 4.93%	Neither
Wand vs W/S	Neither	Neither	Wand 5.98%	Neither
86"Non vs W/S	86"Non 4.92%	Neither	Neither	86"Non 2.27%
<b>Manipulation Ratings</b> 37.5%	<b>Pass 1</b> 0%	<b>Pass 2</b> 33.33%	<b>Pass 3</b> 50%	<b>3 Pass Avg.</b> 66.67%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Neither	Neither	Neither	Glove 6.96%
Glove vs W/S	Neither	Glove 4.76%	Glove 14.22%	Glove 7.74%
Wand vs 86"Non	Neither	Neither	Wand 13.88%	Wand 4.77%
Wand vs W/S	Neither	Neither	Wand 12.82%	Wand 5.54%
86"Non vs W/S	Neither	86"Non 2.93%	Neither	Neither
<b>General Impressions</b> 25%	<b>Pass 1</b> 0%	<b>Pass 2</b> 0%	<b>Pass 3</b> 66.67%	<b>3 Pass Avg.</b> 33.33%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Neither	Neither	Glove 10.42%	Glove 3.61%
Glove vs W/S	Neither	Neither	Glove 9.32%	Glove 3.61%
Wand vs 86"Non	Neither	Neither	Neither	Neither
Wand vs W/S	Neither	Neither	Wand 8.97%	Neither
86"Non vs W/S	Neither	Neither	W/S 1.01%	Neither
<b>Overall Impressions</b> 58.33%	<b>Pass 1</b> 50%	<b>Pass 2</b> 16.67%	<b>Pass 3</b> 66.67%	<b>3 Pass Avg.</b> 100%
Glove vs Wand	Neither	Neither	Neither	Glove 0.65%
Glove vs 86"Non	Neither	Neither	Glove 9.99%	Glove 4.12%
Glove vs W/S	Glove 1.85%	Glove 2.87%	Glove 9.83%	Glove 4.96%
Wand vs 86"Non	Neither	Neither	Wand 8.83%	Wand 3.45%
Wand vs W/S	Wand 1.50%	Neither	Wand 8.67%	Wand 4.28%
86"Non vs W/S	86"Non 1.83%	Neither	Neither	86"Non 0.80%
<b>64 Statistically Significant Results of 144 Comparisons (44.44%)</b>				

Similarly, Table K-3 shows that the all users Benchmark 2 elapsed times and ratings comparisons produced statistically significant results in 79 of the 144 elapsed times and ratings environment comparisons (54.86%).

**Table K- 3 Benchmark 2 All Users Statistically Significant Comparisons**

<b>Benchmark 2 - All Users</b>				
<b>Statistically Significant Environment Comparisons Results Summary</b>				
<b>Elapsed Times</b> 66.67%	<b>Pass 1</b> 66.67%	<b>Pass 2</b> 66.67%	<b>Pass 3</b> 50%	<b>3 Pass Avg.</b> 83.33%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Glove 12.71%	Glove 20.30%	Glove 25.81%	Glove 26.81%
Glove vs W/S	Glove 17.67%	Glove 13.84%	Neither	Glove 23.22%
Wand vs 86"Non	Wand 10.47%	Wand 19.84%	Wand 28.99%	Wand 28.99%
Wand vs W/S	Wand 15.56%	Wand 13.35%	Neither	Wand 28.52%
86"Non vs W/S	Neither	Neither	W/S 3.37%	W/S 3.37%
<b>Navigation Ratings</b> 50%	<b>Pass 1</b> 0%	<b>Pass 2</b> 66.67%	<b>Pass 3</b> 66.67%	<b>3 Pass Avg.</b> 66.67%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Neither	Glove 6.42%	Glove 10.29%	Glove 6.22%
Glove vs W/S	Neither	Glove 5.90%	Glove 11.66%	Glove 7.17%
Wand vs 86"Non	Neither	Wand 6.11%	Wand 8.01%	Wand 5.42%
Wand vs W/S	Neither	Wand 5.59%	Wand 9.34%	Wand 6.36%
86"Non vs W/S	Neither	Neither	Neither	Neither
<b>Locating Ratings</b> 66.67%	<b>Pass 1</b> 66.67%	<b>Pass 2</b> 33.33%	<b>Pass 3</b> 83.33%	<b>3 Pass Avg.</b> 83.33%
Glove vs Wand	Wand 1.18%	Neither	Glove 2.83%	Neither
Glove vs 86"Non	Neither	Glove 5.29%	Glove 10.12%	Glove 5.30%
Glove vs W/S	Glove 4.91%	Glove 5.29%	Glove 10.48%	Glove 6.93%
Wand vs 86"Non	Neither	Neither	Wand 7.08%	Wand 3.91%
Wand vs W/S	Wand 6.15%	Neither	Wand 7.44%	Wand 5.52%
86"Non vs W/S	86"Non 4.38%	Neither	Neither	86"Non 1.55%
<b>Manipulation Ratings</b> 37.5%	<b>Pass 1</b> 0%	<b>Pass 2</b> 33.33%	<b>Pass 3</b> 33.33%	<b>3 Pass Avg.</b> 66.67%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Neither	Glove 5.64%	Neither	Glove 8.13%
Glove vs W/S	Neither	Glove 5.64%	Neither	Glove 7.67%
Wand vs 86"Non	Neither	Neither	Wand 10.60%	Wand 5.09%
Wand vs W/S	Neither	Neither	Wand 10.80%	Wand 4.64%
86"Non vs W/S	Neither	Neither	Neither	W/S 0.43%
<b>General Impressions</b> 50%	<b>Pass 1</b> 0%	<b>Pass 2</b> 50%	<b>Pass 3</b> 83.33%	<b>3 Pass Avg.</b> 66.67%
Glove vs Wand	Neither	Neither	Glove 0.46%	Neither
Glove vs 86"Non	Neither	Glove 4.41%	Glove 11.05%	Glove 6.49%
Glove vs W/S	Neither	Neither	Glove 11.05%	Glove 6.38%
Wand vs 86"Non	Neither	Wand 5.79%	Wand 10.54%	Wand 6.78%
Wand vs W/S	Neither	Wand 5.07%	Wand 10.54%	Wand 6.67%
86"Non vs W/S	Neither	Neither	Neither	Neither
<b>Overall Impressions</b> 58.33%	<b>Pass 1</b> 50%	<b>Pass 2</b> 66.67%	<b>Pass 3</b> 50%	<b>3 Pass Avg.</b> 66.67%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Glove 4.53%	Glove 5.93%	Glove 9.80%	Glove 6.77%
Glove vs W/S	Glove 5.55%	Glove 6.53%	Glove 9.77%	Glove 7.30%
Wand vs 86"Non	Neither	Wand 5.82%	Wand 8.80%	Wand 5.94%
Wand vs W/S	Wand 4.05%	Wand 6.43%	Neither	Wand 6.47%
86"Non vs W/S	Neither	Neither	Neither	Neither
<b>79 Statistically Significant Results of 144 Comparisons (54.86%)</b>				

## **Appendix L**

### **Benchmark 3 Statistically Significant Environment Comparisons**

Tables L-1 (Novice), L-2 (Experienced), and L-3 (All) present a compilation of the statistically significant findings of the overall Benchmark 3 testing. The tabulation includes the significance of both the objective (distance offsets) measures and the subjective user ratings by category. Overall the tables show that 247 of the 432 Benchmark 3 distance offsets and ratings comparisons produce statistically significant results (57.18%).

Table L-1 shows that 85 of the 144 novice user Benchmark 3 distance offsets and ratings comparisons produced statistically significant results (59.03%).

**Table L- 1 Benchmark 3 Novice Users Statistically Significant Comparisons**

Benchmark 3 - Novice Users				
Statistically Significant Environment Comparisons Results Summary				
Distance Offsets 29.17%	Pass 1 0%	Pass 2 0%	Pass 3 83.33%	3 Pass Avg. 33.33%
Glove vs Wand	Neither	Neither	Wand 58.64%	Neither
Glove vs 86"Non	Neither	Neither	Glove 85.85%	Glove 83.43%
Glove vs W/S	Neither	Neither	Neither	Neither
Wand vs 86"Non	Neither	Neither	Wand 94.15%	Neither
Wand vs W/S	Neither	Neither	Wand 93.79%	Neither
86"Non vs W/S	Neither	Neither	W/S 5.81%	W/S 7.39%
Navigation Ratings 58.33%	Pass 1 33.33%	Pass 2 50%	Pass 3 66.67%	3 Pass Avg. 83.33%
Glove vs Wand	Neither	Glove 5.30%	Neither	Glove 4.81%
Glove vs 86"Non	Glove 9.05%	Neither	Glove 9.20%	Glove 9.83%
Glove vs W/S	Neither	Glove 14.99%	Glove 14.66%	Glove 13.78%
Wand vs 86"Non	Wand 4.26%	Wand 6.28%	Wand 5.24%	Wand 5.27%
Wand vs W/S	Neither	Neither	Wand 10.94%	Wand 9.42%
86"Non vs W/S	Neither	Neither	Neither	Neither
Locating Ratings 54.17%	Pass 1 33.33%	Pass 2 33.33%	Pass 3 66.67%	3 Pass Avg. 83.33%
Glove vs Wand	Neither	Glove 5.30%	Neither	Glove 4.81%
Glove vs 86"Non	Glove 9.80%	Neither	Glove 9.20%	Glove 9.83%
Glove vs W/S	Neither	Neither	Glove 14.66%	Glove 13.78%
Wand vs 86"Non	Wand 5.47%	Wand 6.28%	Wand 5.24%	Wand 5.27%
Wand vs W/S	Neither	Neither	Wand 10.94%	Wand 9.42%
86"Non vs W/S	Neither	Neither	Neither	Neither
Manipulation Ratings 62.5%	Pass 1 50%	Pass 2 33.33%	Pass 3 83.33%	3 Pass Avg. 83.33%
Glove vs Wand	Neither	Glove 9.61%	Neither	Glove 12.16%
Glove vs 86"Non	Glove 15.12%	Neither	Glove 10.12%	Glove 12.82%
Glove vs W/S	Glove 25.07%	Neither	Glove 16.26%	Glove 19.52%
Wand vs 86"Non	Wand 3.02%	Wand 4.11%	86"Non 4.68%	Wand 0.75%
Wand vs W/S	Neither	Neither	Wand 2.26%	Neither
86"Non vs W/S	Neither	Neither	86"Non 6.83%	86"Non 7.69%
General Impressions 62.5%	Pass 1 50%	Pass 2 50%	Pass 3 83.33%	3 Pass Avg. 66.67%
Glove vs Wand	Neither	Neither	Neither	Glove 1.86%
Glove vs 86"Non	Neither	Glove 8.90%	Glove 11.02%	Glove 8.86%
Glove vs W/S	Glove 9.71%	Glove 12.28%	Glove 14.50%	Glove 12.27%
Wand vs 86"Non	Wand 10.54%	Neither	Wand 6.70%	Neither
Wand vs W/S	Wand 13.73%	Neither	Wand 10.35%	Neither
86"Non vs W/S	Neither	86"Non 3.71%	86"Non 3.91%	86"Non 3.74%
Overall Impressions 87.5%	Pass 1 66.67%	Pass 2 100%	Pass 3 83.33%	3 Pass Avg. 100%
Glove vs Wand	Neither	Glove 6.68%	Glove 6.27%	Glove 6.27%
Glove vs 86"Non	Glove 10.11%	Glove 11.24%	Glove 9.72%	Glove 10.12%
Glove vs W/S	Glove 13.09%	Glove 15.02%	Glove 16.44%	Glove 16.44%
Wand vs 86"Non	Wand 6.00%	Wand 4.89%	Neither	Wand 4.10%
Wand vs W/S	Wand 9.12%	Wand 8.94%	Wand 10.84%	Wand 10.84%
86"Non vs W/S	Neither	86"Non 4.26%	86"Non 7.44%	86"Non 7.03%
85 Statistically Significant Results of 144 Comparisons (59.03%)				

Table L-2 shows that but 60 of the 144 experienced user Benchmark 3 distance offsets and ratings comparisons produced statistically significant results (41.67%).

**Table L- 2 Benchmark 3 Experienced Users Statistically Significant Comparisons**

Benchmark 3 - Experienced Users				
Statistically Significant Environment Comparisons Results Summary				
Distance Offsets 41.67%	Pass 1 50%	Pass 2 33.33%	Pass 3 66.67%	3 Pass Avg. 33.33%
Glove vs Wand	Neither	Neither	Wand 31.11%	Neither
Glove vs 86"Non	Neither	Glove #####	Glove #####	Glove #####
Glove vs W/S	Neither	Neither	Glove #####	Neither
Wand vs 86"Non	Wand #####	Neither	Neither	Neither
Wand vs W/S	Wand 48.96%	Neither	Neither	Neither
86"Non vs W/S	Neither	86"Non 21.00%	W/S 2.56%	W/S 5.06%
Navigation Ratings 33.33%	Pass 1 0%	Pass 2 33.33%	Pass 3 66.67%	3 Pass Avg. 33.33%
Glove vs Wand	Neither	Neither	Wand 23.50%	Wand 23.76%
Glove vs 86"Non	Neither	Neither	Neither	Neither
Glove vs W/S	Neither	Neither	Neither	Neither
Wand vs 86"Non	Neither	Wand 23.92%	Wand 13.15%	Neither
Wand vs W/S	Neither	Wand 23.92%	Wand 13.15%	Wand 13.89%
86"Non vs W/S	Neither	Neither	86"Non 0.98%	Neither
Locating Ratings 37.5%	Pass 1 0%	Pass 2 33.33%	Pass 3 66.67%	3 Pass Avg. 50%
Glove vs Wand	Neither	Neither	Wand 23.50%	Wand 23.76%
Glove vs 86"Non	Neither	Neither	Neither	Neither
Glove vs W/S	Neither	Neither	Neither	Neither
Wand vs 86"Non	Neither	Wand 23.92%	Wand 13.15%	Wand 13.89%
Wand vs W/S	Neither	Wand 23.92%	Wand 13.15%	Wand 13.89%
86"Non vs W/S	Neither	Neither	86"Non 0.98%	Neither
Manipulation Ratings 33.33%	Pass 1 0%	Pass 2 33.33%	Pass 3 66.67%	3 Pass Avg. 33.33%
Glove vs Wand	Neither	Neither	Wand 23.50%	Wand 23.76%
Glove vs 86"Non	Neither	Neither	Neither	Neither
Glove vs W/S	Neither	Neither	Neither	Neither
Wand vs 86"Non	Neither	Wand 23.92%	Wand 13.15%	Neither
Wand vs W/S	Neither	Wand 23.92%	Wand 13.15%	Wand 13.89%
86"Non vs W/S	Neither	Neither	86"Non 0.98%	Neither
General Impressions 33.33%	Pass 1 0%	Pass 2 33.33%	Pass 3 66.67%	3 Pass Avg. 33.33%
Glove vs Wand	Neither	Neither	Wand 23.50%	Wand 23.76%
Glove vs 86"Non	Neither	Neither	Neither	Neither
Glove vs W/S	Neither	Neither	Neither	Neither
Wand vs 86"Non	Neither	Wand 23.92%	Wand 13.15%	Neither
Wand vs W/S	Neither	Wand 23.92%	Wand 13.15%	Wand 13.89%
86"Non vs W/S	Neither	Neither	86"Non 0.98%	Neither
Overall Impressions 33.33%	Pass 1 0%	Pass 2 33.33%	Pass 3 66.67%	3 Pass Avg. 33.33%
Glove vs Wand	Neither	Neither	Wand 23.50%	Wand 23.76%
Glove vs 86"Non	Neither	Neither	Neither	Neither
Glove vs W/S	Neither	Neither	Neither	Neither
Wand vs 86"Non	Neither	Wand 23.92%	Wand 13.15%	Neither
Wand vs W/S	Neither	Wand 23.92%	Wand 13.15%	Wand 13.89%
86"Non vs W/S	Neither	Neither	86"Non 0.98%	Neither
60 Statistically Significant Results of 144 Comparisons (41.67%)				

Similarly, Table K-3 shows that the all users Benchmark 3 distance offsets and ratings comparisons produced statistically significant results in 102 of the 144 elapsed times and ratings environment comparisons (70.83%).

**Table L- 3 Benchmark 3 All Users Statistically Significant Comparisons**

Benchmark 3 - All Users				
Statistically Significant Environment Comparisons Results Summary				
Distance Offsets 45.83%	Pass 1 16.67%	Pass 2 33.33%	Pass 3 66.67%	3 Pass Avg. 66.67%
Glove vs Wand	Neither	Neither	Glove 4.27%	Neither
Glove vs 86"Non	Neither	Glove #####	Glove #####	Glove #####
Glove vs W/S	Neither	Neither	Glove #####	Neither
Wand vs 86"Non	Wand #####	Neither	Neither	Wand #####
Wand vs W/S	Neither	Neither	Neither	Wand 81.35%
86"Non vs W/S	Neither	86"Non 24.14%	W/S 0.75%	W/S 5.06%
Navigation Ratings 70.83%	Pass 1 50%	Pass 2 83.33%	Pass 3 50%	3 Pass Avg. 100%
Glove vs Wand	Neither	Wand 1.03%	Neither	Wand 1.03%
Glove vs 86"Non	Neither	Glove 5.65%	Neither	Glove 5.53%
Glove vs W/S	Neither	Neither	Neither	Glove 10.23%
Wand vs 86"Non	Wand 4.90%	Wand 6.62%	Wand 6.22%	Wand 6.50%
Wand vs W/S	Wand 10.33%	Wand 11.63%	Wand 11.47%	Wand 11.16%
86"Non vs W/S	86"Non 5.72%	86"Non 5.37%	86"Non 5.60%	86"Non 4.98%
Locating Ratings 79.17%	Pass 1 66.67%	Pass 2 66.67%	Pass 3 83.33%	3 Pass Avg. 100%
Glove vs Wand	Neither	Glove 1.88%	Wand 1.33%	Wand 0.02%
Glove vs 86"Non	Glove 7.35%	Glove 8.40%	Glove 6.74%	Glove 7.00%
Glove vs W/S	Neither	Neither	Glove 9.43%	Glove 9.95%
Wand vs 86"Non	Wand 6.91%	Neither	Wand 7.98%	Wand 7.02%
Wand vs W/S	Wand 10.27%	Wand 9.28%	Wand 10.64%	Wand 9.97%
86"Non vs W/S	86"Non 3.61%	86"Non 2.83%	Neither	86"Non 3.18%
Manipulation Ratings 75%	Pass 1 50%	Pass 2 100%	Pass 3 66.67%	3 Pass Avg. 83.33%
Glove vs Wand	Neither	Glove 0.69%	Glove 4.42%	Neither
Glove vs 86"Non	Glove 8.16%	Glove 5.71%	Glove 7.71%	Glove 8.90%
Glove vs W/S	Glove 11.62%	Glove 9.80%	Glove 12.04%	Glove 10.83%
Wand vs 86"Non	Neither	Wand 5.05%	Wand 3.44%	Wand 5.43%
Wand vs W/S	Neither	Wand 9.17%	Neither	Wand 7.23%
86"Non vs W/S	86"Non 3.76%	86"Non 4.34%	Neither	86"Non 1.91%
General Impressions 70.83%	Pass 1 50%	Pass 2 83.33%	Pass 3 83.33%	3 Pass Avg. 66.67%
Glove vs Wand	Neither	Wand 1.11%	Neither	Neither
Glove vs 86"Non	Glove 3.57%	Glove 3.64%	Glove 5.87%	Glove 4.42%
Glove vs W/S	Glove 7.82%	Glove 8.57%	Glove 8.87%	Glove 8.45%
Wand vs 86"Non	Neither	Neither	Wand 7.00%	Wand 6.45%
Wand vs W/S	Neither	Wand 9.58%	Wand 9.96%	Neither
86"Non vs W/S	86"Non 4.41%	86"Non 5.11%	86"Non 3.19%	86"Non 4.22%
Overall Impressions 83.33%	Pass 1 83.33%	Pass 2 83.33%	Pass 3 83.33%	3 Pass Avg. 83.33%
Glove vs Wand	Neither	Neither	Neither	Neither
Glove vs 86"Non	Glove 6.57%	Glove 5.94%	Glove 10.37%	Glove 6.40%
Glove vs W/S	Glove 9.85%	Glove 9.59%	Glove 13.64%	Glove 9.83%
Wand vs 86"Non	Wand 6.50%	Wand 6.13%	Wand 5.96%	Wand 6.21%
Wand vs W/S	Wand 9.78%	Wand 9.77%	Wand 9.40%	Wand 9.65%
86"Non vs W/S	86"Non 3.50%	86"Non 3.67%	86"Non 3.65%	86"Non 3.66%
102 Statistically Significant Results of 144 Comparisons (70.83%)				

Overall the three benchmarks produced statistically significant comparison in 753 of 1296 cases (58.10%).



## **Appendix M**

### **Cronbach's Alpha Reliability Calculations Detail**

**Table M- 1 Cronbach's Coefficient Alpha - Benchmark 1 Elapsed Times**

Benchmark 1 Elapsed Times - Cronbach Coefficient Alpha				
	Raw		94.06%	
	Standardized		94.63%	
Cronbach Coefficient Alpha with Deleted Environment				
Deleted Environment	Raw Variables		Standardized Variables	
	Correlation with Total	Alpha	Correlation with Total	Alpha
Glove Pass1	0.243083	94.52%	0.239528	95.26%
Glove Pass2	0.521029	94.04%	0.528505	94.67%
Glove Pass3	0.586659	93.92%	0.606408	94.50%
Glove 3 Pass Avg.	0.610138	93.87%	0.613047	94.49%
Wand Pass1	0.486028	94.22%	0.524953	94.67%
Wand Pass2	0.856688	93.52%	0.856684	93.96%
Wand Pass3	0.722563	93.87%	0.727977	94.24%
Wand 3 Pass Avg.	0.794861	93.63%	0.810059	94.06%
86"Non Pass1	0.687629	93.86%	0.664131	94.38%
86"Non Pass2	0.84764	93.27%	0.826394	94.03%
86"Non Pass3	0.854051	93.26%	0.845009	93.99%
86"Non 3 Pass	0.86203	93.22%	0.833961	94.01%
W/S Pass1	0.724341	93.59%	0.731139	94.23%
W/S Pass2	0.825594	93.33%	0.831912	94.01%
W/S Pass3	0.789896	93.43%	0.786321	94.11%
W/S 3 Pass Avg.	0.860884	93.28%	0.861516	93.95%

**Table M- 2 Chronbach's Coefficient Alpha – Benchmark 1 Overall Impressions Ratings**

Benchmark 1 Overall Impressions - Cronbach Coefficient Alpha				
	Raw		48.20%	
	Standardized		92.05%	
Cronbach Coefficient Alpha with Deleted Environment				
Deleted Environment	Raw Variables		Standardized Variables	
	Correlation with Total	Alpha	Correlation with Total	Alpha
Glove Pass1	0.074331	48.05%	0.526484	91.83%
Glove Pass2	0.567743	44.71%	0.758973	91.16%
Glove Pass3	0.595601	43.58%	0.716966	91.28%
Glove 3 Pass Avg.	0.289919	46.46%	0.661766	91.44%
Wand Pass1	0.675905	41.60%	0.917809	90.68%
Wand Pass2	0.624863	41.48%	0.888486	90.77%
Wand Pass3	0.596493	41.64%	0.877167	90.80%
Wand 3 Pass Avg.	0.634937	41.61%	0.901454	90.73%
86"Non Pass1	0.469563	44.73%	0.657503	91.45%
86"Non Pass2	0.462928	44.78%	0.718824	91.27%
86"Non Pass3	0.247371	46.32%	0.624722	91.55%
86"Non 3 Pass	0.473166	44.79%	0.711588	91.29%
W/S Pass1	0.46664	44.50%	0.371156	92.27%
W/S Pass2	0.555386	43.26%	0.463244	92.01%
W/S Pass3	0.477261	44.76%	0.373464	92.26%
W/S 3 Pass Avg.	-0.139188	93.75%	-0.143902	93.62%

**Table M- 3 Chronbach's Coefficient Alpha – Benchmark 2 Elapsed Times**

Benchmark 2 Elapsed Times - Cronbach Coefficient Alpha				
	Raw		89.62%	
	Standardized		86.12%	
Cronbach Coefficient Alpha with Deleted Environment				
Deleted Environment	Raw Variables		Standardized Variables	
	Correlation with Total	Alpha	Correlation with Total	Alpha
Glove Pass1	0.437073	89.55%	0.395684	85.77%
Glove Pass2	0.67058	88.79%	0.769506	83.93%
Glove Pass3	0.276142	89.78%	0.39575	85.77%
Glove 3 Pass Avg.	0.72991	88.81%	0.783157	83.86%
Wand Pass1	0.861853	87.63%	0.774207	83.91%
Wand Pass2	0.767006	88.10%	0.628555	84.64%
Wand Pass3	0.819607	87.90%	0.71024	84.23%
Wand 3 Pass Avg.	0.862681	87.69%	0.742147	84.07%
86"Non Pass1	0.725958	88.47%	0.628305	84.64%
86"Non Pass2	0.871928	87.69%	0.819021	83.68%
86"Non Pass3	0.771664	88.32%	0.799375	83.78%
86"Non 3 Pass	0.906137	87.59%	0.834115	83.60%
W/S Pass1	-0.088547	90.75%	0.037531	87.39%
W/S Pass2	0.013944	90.39%	0.079492	87.21%
W/S Pass3	-0.016786	90.57%	0.062631	87.28%
W/S 3 Pass Avg.	-0.483495	90.35%	-0.405823	89.23%

**Table M- 4 Chronbach's Coefficient Alpha – Benchmark 2 Overall Impressions Ratings**

Benchmark 2 Overall Impressions - Cronbach Coefficient Alpha				
	Raw		34.00%	
	Standardized		95.74%	
Cronbach Coefficient Alpha with Deleted Environment				
Deleted Environment	Raw Variables		Standardized Variables	
	Correlation with Total	Alpha	Correlation with Total	Alpha
Glove Pass1	0.461823	28.25%	0.731035	95.50%
Glove Pass2	0.480913	27.76%	0.887302	95.20%
Glove Pass3	0.402497	29.96%	0.752394	95.46%
Glove 3 Pass Avg.	0.498998	28.43%	0.86011	95.25%
Wand Pass1	0.545869	28.07%	0.915483	95.15%
Wand Pass2	0.614421	27.32%	0.91424	95.15%
Wand Pass3	0.603859	28.92%	0.78582	95.39%
Wand 3 Pass Avg.	0.61252	28.01%	0.913109	95.15%
86"Non Pass1	0.576183	28.42%	0.887732	95.20%
86"Non Pass2	0.590352	27.79%	0.950846	95.08%
86"Non Pass3	0.61191	29.40%	0.929563	95.12%
86"Non 3 Pass	0.622814	28.46%	0.967789	95.05%
W/S Pass1	0.478661	28.81%	0.749105	95.46%
W/S Pass2	0.306533	31.41%	0.422079	96.06%
W/S Pass3	0.469663	29.58%	0.703437	95.55%
W/S 3 Pass Avg.	-0.304178	97.11%	-0.297973	97.27%

**Table M- 5 Chronbach's Coefficient Alpha – Benchmark 3 Distance Offsets**

Benchmark 3 Distance Offsets - Cronbach Coefficient Alpha				
	Raw		74.84%	
	Standardized		77.80%	
Cronbach Coefficient Alpha with Deleted Environment				
Deleted Environment	Raw Variables		Standardized Variables	
	Correlation with Total	Alpha	Correlation with Total	Alpha
Glove Pass1	-0.28889	75.75%	-0.126597	80.39%
Glove Pass2	0.003633	75.19%	0.369897	76.67%
Glove Pass3	0.852934	74.71%	0.735863	73.61%
Glove 3 Pass Avg.	-0.004127	75.19%	0.275553	77.42%
Wand Pass1	0.805255	66.37%	0.358398	76.76%
Wand Pass2	0.826932	66.02%	0.390947	76.50%
Wand Pass3	0.706263	70.85%	0.34495	76.87%
Wand 3 Pass Avg.	0.839837	65.61%	0.38885	76.52%
86"Non Pass1	0.193969	74.95%	0.493795	75.67%
86"Non Pass2	0.287516	74.06%	0.585082	74.91%
86"Non Pass3	0.112599	75.03%	0.141717	78.44%
86"Non 3 Pass	0.290407	74.11%	0.558478	75.13%
W/S Pass1	0.171367	74.87%	0.390688	76.51%
W/S Pass2	0.142479	74.96%	0.545837	75.24%
W/S Pass3	0.239222	74.81%	0.599441	74.79%
W/S 3 Pass Avg.	-0.077736	75.21%	-0.010784	79.57%

**Table M- 6 Chronbach's Coefficient Alpha – Benchmark 3 Overall Impressions Ratings**

Benchmark 3 Overall Impressions - Cronbach Coefficient Alpha				
	Raw		67.46%	
	Standardized		89.42%	
Cronbach Coefficient Alpha with Deleted Environment				
Deleted Environment	Raw Variables		Standardized Variables	
	Correlation with Total	Alpha	Correlation with Total	Alpha
Glove Pass1	0.470185	65.15%	0.598455	88.62%
Glove Pass2	0.380794	66.03%	0.631623	88.50%
Glove Pass3	0.321902	66.43%	0.614357	88.57%
Glove 3 Pass Avg.	0.456495	65.73%	0.699087	88.25%
Wand Pass1	0.750028	63.38%	0.70699	88.22%
Wand Pass2	0.801942	64.79%	0.926253	87.36%
Wand Pass3	0.649367	64.72%	0.624205	88.53%
Wand 3 Pass Avg.	0.880903	64.10%	0.89639	87.48%
86"Non Pass1	0.817604	62.83%	0.642681	88.46%
86"Non Pass2	0.824376	63.73%	0.67249	88.35%
86"Non Pass3	0.625032	63.85%	0.639366	88.47%
86"Non 3 Pass	0.876886	63.23%	0.766962	87.98%
W/S Pass1	-0.152518	68.65%	0.022831	90.66%
W/S Pass2	-0.157067	68.55%	0.051828	90.56%
W/S Pass3	-0.057898	68.18%	0.074473	90.48%
W/S 3 Pass Avg.	0.450319	89.47%	0.418238	89.29%

## **APPENDIX N**

### **Human Subjects Protection Certification**





## Human Participant Protections Education for Research

### Completion Certificate

This is to certify that

**Kurt Satter**

has completed the **Human Participants Protection Education for Research Teams** online course, sponsored by the National Institutes of Health (NIH), on 07/15/2005.

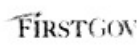
This course included the following:

- key historical events and current issues that impact guidelines and legislation on human participant protection in research.
- ethical principles and guidelines that should assist in resolving the ethical issues inherent in the conduct of research with human participants.
- the use of key ethical principles and federal regulations to protect human participants at various stages in the research process.
- a description of guidelines for the protection of special populations in research.
- a definition of informed consent and components necessary for a valid consent.
- a description of the role of the IRB in the research process.
- the roles, responsibilities, and interactions of federal agencies, institutions, and researchers in conducting research with human participants.

National Institutes of Health  
<http://www.nih.gov>

[Home](#) | [Contact Us](#) | [Policies](#) | [Accessibility](#) | [Site Help](#) | [Site Map](#)

A Service of the National Cancer Institute



## **Vita**

Kurt Michael Satter was born in New Orleans, LA and attended public schools in Orleans Parish, graduating from Alcee Fortier Senior High School in 1961. Upon completion of his Bachelor of Science degree in Mathematics from Louisiana Polytechnic Institute, Ruston, LA (now Louisiana Tech University) in 1967, he began a 33 year career in industry working first as a technical writer/editor and, after several University of Houston post-baccalaureate courses in computer science, progressed through the many levels and roles in software services to his retirement from industry as an Information Systems Manager in heavy manufacturing applications.

In 1998 he received a Master of Science degree in Software Engineering from Southern Methodist University, Dallas, TX. He followed that program with a second SMU M.S. degree in Systems Engineering awarded in 1999.

Upon completion of his first M.S. degree Mr. Satter began teaching first as an Adjunct Professor of Computer Science at the University of Dayton (OH) then as a member of the Adjunct Faculty of Management Information Systems Department at Wright State University, Dayton, OH.

Mr. Satter retired from industry in December, 2000 and upon returning to New Orleans, continued teaching as an Adjunct Professor of MIS at Tulane University. In the fall of 2001 he became a member of the full-time faculty of the University of New Orleans as an MIS Instructor in the Management Department of the College of Business.

In September of 2002, Kurt was one of the first to be accepted into a new Engineering Management PhD Program at UNO where he has been pursuing his doctorate under the supervision of the department's chairman, Prof. William Lannes, and his advisor, Dr. Alley C. Butler.