

**Techniques for Evaluation of Occupational Safety and Health (OSH) Signage to Ensure
Better Comprehensibility**

*A thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy*

in Design

By

Sangeeta Bhanja Chaudhuri

Roll No. 186105007

Under the supervision of

Prof. Sougata Karmakar

&

Dr. Manoj Majhi

Department of Design



Indian Institute of Technology Guwahati

Guwahati- 781039, India

September 2023



Department of Design
Indian Institute of Technology Guwahati
Guwahati, Assam – 781 039, India

Declaration

I hereby declare that the research work presented in the thesis titled " Techniques for Evaluation of Occupational Safety and Health (OSH) Signage to Ensure Better Comprehensibility" has been conducted by me under the guidance of my supervisor(s)-Professor and HOD, Dr. Sougata Karmakar and Dr. Manoj Majhi, at the Department of Design, Indian Institute of Technology Guwahati, Assam, India. The content of the thesis (text, illustration, data, plots, diagrams, etc.) and the outcome of my research are original. Any relevant material from the open literature has been referred to and cited per established ethical norms and practices. All collaborations and critiques contributing to giving the thesis its final shape are duly acknowledged and credited. I have adhered to all academic honesty and integrity principles and have not misrepresented, fabricated, or falsified any idea/ data/ fact/ source in my submission. Finally, this is an original research work I conducted for my Ph.D. and has not been submitted elsewhere for a degree.

Sangeeta Bhanja Chaudhuri

Place: IIT Guwahati

Sangeeta Bhanja Chaudhuri

Date: 25/09/2023

Roll: 186105007
Department of Design
Indian Institute of Technology Guwahati
Guwahati 781039
Assam, India



Department of Design
Indian Institute of Technology Guwahati
Guwahati, Assam – 781 039, India

Certificate

This is to certify that the work contained in this thesis titled "Techniques for Evaluation of Occupational Safety and Health (OSH) Signage to Ensure Better Comprehensibility" submitted by Ms. Sangeeta Bhanja Chaudhuri to the Indian Institute of Technology Guwahati for the award of the degree of Doctor of Philosophy has been carried out under our supervision. This work has not been submitted elsewhere for the award of any other degree or diploma.

Place: IIT Guwahati
Date: 25/09/2023

Sougata Karmakar.

Prof. Sougata Karmakar

(Supervisor)

HOD, Department of Design
Indian Institute of Technology
Guwahati
Guwahati 781039
Assam, India

Manoj Majhi

Dr. Manoj Majhi

(Co-Supervisor)

Department of Design
Indian Institute of Technology
Guwahati
Guwahati 781039
Assam, India

Acknowledgment

First, I would like to thank God Almighty for giving me enough strength to finish my research work. I owe my deepest gratitude to my supervisors Prof. Sougata Karmakar and Dr. Manoj Majhi for their kind support throughout this journey. This journey of research would not have been possible without the support of our former HOD, Dr. D Udaya Kumar. I am obliged to my doctoral committee members: Prof. Debkumar Chakrabarti, Prof. Utpal Barua, Dr. Shakuntala Acharya (from the Dept. of Design), and Dr. Srinivasan Krishnaswamy (Dept. of EEE). They have periodically reviewed my work, provided significant insights and suggestions, and supported me in completing the research successfully. I would like to thank all faculties and Ergonomics laboratory members in the Department of Design for providing the necessary resources required during the research and supporting me throughout my research. I am also thankful to the volunteers who participated in the studies that are part of the thesis. Special thanks to all colleagues at IIT Guwahati for their voluntary participation in experiments. I am grateful to all the non-teaching staff of the department for their continuous support whenever required.

I will always be indebted to my parents, sister, husband, and brother-in-law (Mr. Swapan Kumar Bhanja Chaudhuri, Mrs. Gayatri Bhanja Chaudhuri, Nandita, Prosenjit, and Anindya), for their unwavering support, encouragement, and love without which I could not have endured the highs and lows of these years.

Sangeeta Bhanja Chaudhuri

Sangeeta Bhanja Chaudhuri

Abstract of the thesis

'Safety first' is the slogan of every organization to avoid any unwanted Occupational safety and health (OSH) hazards. The purpose of the OSH signs is to grab the attention of people rapidly to objects or situations affecting OSH and to gain a rapid understanding of a specific message. However, they are not always effective in communicating the intended message or necessary information to the employees or ordinary people. The result is that people misinterpret the signs, which may cause disaster and become hazards for them. Literature reveals a need for more study concerning the methods or techniques of sign comprehensibility, which is essential for the proper evaluation of the signs. Hence this necessitated the study of the techniques of comprehensibility of OSH signs.

Four techniques have been studied here one by one, and each technique has been found to perform well in finding comprehensibility levels of the signs. Seven sign pairs (seven with text and seven without text) have been used as a stimulus for all the techniques implemented here.

A rigorous study has been performed for all four techniques, and both advantages and disadvantages have been elicited and discussed in the thesis. Research is designed in such a way that the weakness of one technique is followed to design the consecutive, more standardized technique.

Initially, a subject-based technique was designed where the subjects scored comprehensibility of signs in response to the questions in the questionnaire. The subject-based technique suffers from ambiguous scores from the novices who are not experts in the scoring procedure. Hence, an expert-based technique has been designed where experts from industries score the textual responses of the subjects. This technique is standard; however, it may suffer from the biases of the experts while scoring the responses, which might be due to their thoughts, previous exposure, prolonged working hours, context, persuasion, etc. This biased result might produce inconsistent data.

To relieve the experts from this tedious scoring procedure and to generate a consistent result, a computational technique has been developed that is capable of generating scores without any ambiguity. The system can consistently decide whether a sign is comprehensible or incomprehensible. The model is validated by expert judgment and Pearson value >0.80 ($p\text{-value} < 0.05$) which is significant. This novel technique will act as a tool and assist the OSH experts in scoring the extensive surveys automatically within seconds.

Each technique designed generates two categories of signs- comprehensible and incomprehensible. However, these three techniques discussed may suffer from the personal biases of the subjects. It is impossible to eliminate the biases in their entirety, as subjects' views are essential in the case of human-centered evaluation and design. A fourth behavioral technique- eye tracking, has been designed and implemented, which helps collect objective data from the users. This eye-tracking technique provides unobtrusive data; however, it was not possible to eliminate the subjective data. The subjective score is collected in this technique so that a correlation study can be framed to find a correlation between subjective

and objective data. Forty-six (46) significant objective measures have been identified from the bivariate correlation between subjective and objective data.

All the data obtained from four techniques have been tested for descriptive statistics and data distribution. Finally, statistical tests have been performed to check the hypothesis proposed in the research. The comprehensibility scores generated in each technique have been used to check whether OSH-related textual information affects the sign's comprehensibility. Another test has been performed where the variances of the signs' comprehensibility are identified. A separate test has been conducted to check whether objective measures vary according to comprehensibility levels. Correlation (Pearson) plot and bivariate correlation have been used to see correlations between the objective measures and the comprehensibility of the signs. All the tests have been conducted using SPSS statistical software. Each technique in this work is unique in its way and standard. So any individual techniques and triangulation of the techniques can also be implemented and used based on the needs of the organizations.

It is significant to note that a user-centered design approach has been followed throughout the research work, and the research generated 'knowledge' and 'design' both at the end. A novel evaluation framework is generated, which will assist the organizations in evaluating the signs based on their severity, complexity of the situation, environment, experts available, etc. A comparative study has been done between the techniques, and the models (techniques) have been validated. The research is a contribution in terms of developing the framework for comprehensibility evaluation following four different techniques. The comparative study will assist the organizations in choosing the evaluation techniques based on various parameters discussed in the study. The novelty of the study lies in the contribution to the 'design process', 'research through design', and development of an assistive tool for experts, which is an intervention in the domain of OSH. The assistive tool will indirectly improve the design life cycle in terms of time and quality. The eye-tracking study helped identify important eye features that can represent subjective data.

Table of Contents

Declaration	ii
Certificate	iii
Acknowledgment	iv
Abstract of the thesis	v
Table of Contents	vii
List of Figures.....	xi
List of tables.....	xiv
Abbreviations used in the report.....	xvii
Chapter 1: Introduction.....	2
1.1. Introduction to Occupational Safety and Health (OSH) signs.....	2
1.2. Motivation and justification for studying comprehensibility evaluation of a sign.....	3
1.2.1. Theoretical constructs	3
1.2.1.1. Model of Semiotic Theory	5
1.2.2. Conceptual Framework.....	6
1.3. State-of-the-art review of the techniques for comprehensibility evaluation.....	8
1.3.1. Evaluation of comprehensibility utilizing the subject-based and expert-based techniques	9
1.3.2. Evaluation of comprehensibility utilizing computational evaluation technique.....	17
1.3.2.1. Deep Learning and Natural Language Processing (NLP) tasks.....	18
1.3.3. Evaluation of comprehensibility utilizing an eye-tracking evaluation technique	19
1.4. Summary of literature review and research gap.....	22
1.5. Research questions (RQ)	26
1.6. Problem statement	26
1.7. Aim	27
1.8. Objectives.....	28
1.9. Hypothesis.....	28
1.10. Expected outcome	29
1.11. A Framework of the Thesis.....	30
1.11.1. Brief of chapters.....	33
Chapter 2: Evaluation of Comprehensibility of OSH Signs by Subject-based and Expert-based Techniques.....	36
2.1. Introduction.....	36
2.2. Methodology	37

2.2.1.	Stimuli.....	37
2.2.2.	Participants.....	39
2.2.3.	Research Instrument and Research Design.....	39
2.3.	Results and Discussions	42
2.3.1.	Subject-based technique.....	42
2.3.1.1.	Descriptive Statistics.....	42
2.3.1.2.	Test for Normality.....	45
2.3.1.3.	Analysis of variance (ANOVA).....	46
2.3.1.4.	OSH sign with OSH-related textual information affects the comprehensibility of the messages intended to convey (subject-based technique)	52
2.3.1.5.	Comprehensibility percentage based on subject-based technique	56
2.3.2.	Expert-based technique	57
2.3.2.1.	Interrater reliability	57
2.3.2.2.	Descriptive Statistics for the final score.....	60
2.3.2.3.	Test for Normality.....	63
2.3.2.4.	Analysis of variance (ANOVA).....	64
2.3.2.5.	OSH sign with OSH-related textual information affects the comprehensibility of the messages intended to convey (expert-based).....	71
2.3.2.6.	Comprehensibility percentage based on expert-based technique.....	74
2.4.	Conclusion	75
Chapter 3: Evaluation of Comprehensibility of OSH Signs by Computational Technique		77
3.1.	Introduction.....	77
3.2.	Methodology	79
3.2.1.	Stimuli.....	79
3.2.2.	Participants.....	80
3.2.3.	Research Instrument and Research Design.....	80
3.2.3.1.	Phase-wise description of the Model	82
3.3.	Results and Discussions	89
3.3.1.	Results for the computational model (Automated score generation).....	89
3.3.2.	Descriptive Statistics.....	93
3.3.3.	Test for Normality.....	96
3.3.4.	Analysis of variance (ANOVA).....	96
3.3.5.	OSH sign with OSH-related textual information affects the understanding or comprehensibility of the messages intended to convey (computationally computed data)	104
3.3.6.	Comprehensibility percentage based on computationally generated data	107

3.3.7.	Tuning of Threshold and CDM (NC_p)	108
3.3.8.	Model Flexibility.....	109
3.3.9.	Robustness of the Model.....	109
3.4.	Conclusion	110
Chapter 4: Evaluation of Comprehensibility of OSH Signs by Eye-tracking Technique		113
4.1.	Introduction.....	113
4.2.	Methodology and Experimental Setup.....	114
4.2.1.	Stimuli.....	114
4.2.2.	Participants.....	115
4.2.3.	Research Instrument and Research Design	116
4.2.3.1.	Method1 (subject-based study)	116
4.2.3.2.	Method 2 (eye-tracking study).....	117
4.2.3.3.	Statistical tests using data from Method 1 and Method 2	122
4.3.	Results and Discussions	123
4.3.1.	Test for Normality.....	123
4.3.2.	Comprehensibility percentage from subjective data and the result interpretation for objective data	130
4.3.2.1.	Result Analysis using Correlation Plot	131
4.3.2.2.	Result Analysis using Bivariate Correlation	139
4.3.3.	Analysis of variance (ANOVA).....	145
4.3.4.	OSH sign with OSH-related textual information affects the understanding or comprehensibility of the messages intended to convey (eye-tracking technique)	148
4.3.5.	ANOVA of objective data for the comprehensible and incomprehensible signs.....	153
4.4.	Conclusion	155
Chapter 5: Comparative study between the techniques- its merits and demerits.....		158
5.1.	Introduction.....	158
5.2.	Methodology	158
5.3.	Results and Discussions	159
5.3.1.	Correlation (Pearson) plot of the techniques.....	159
5.3.2.	Bivariate correlation of the techniques	162
5.3.3.	Comparison of techniques considering different parameters	168
5.3.3.1.	Comparison of DAN-based computational model with related computational models	172
5.3.4.	Validation of techniques	173
5.4.	Conclusion	174

Chapter 6: Proof of concept	176
6.1. Introduction	176
6.2. Sign design/ redesign life cycle	176
6.2.1. Initial conceptualization	177
6.2.2. Design of variants	177
6.2.3. Sign Design Evaluation Phase	179
6.2.3.1. Participants	179
6.2.3.2. Stimuli	179
6.2.3.3. Research Design	179
6.2.3.4. Result Analysis and Discussion	182
6.3. Conclusion	183
Chapter 7: Discussion: Key findings, contribution of the thesis, limitations, and future scope	185
7.1. Discussion	185
7.1.1. Key findings of the present thesis	186
7.1.1.1. Key Findings from Chapter 1	186
7.1.1.2. Key findings from Chapter 2	186
7.1.1.3. Key findings from Chapter 3	188
7.1.1.4. Key findings from Chapter 4	189
7.1.1.5. Key findings from Chapter 5	190
7.1.1.6. Key Findings from Chapter 6	192
7.1.2. Fulfillment of the objectives	192
7.1.3. Hypothesis testing	196
7.1.4. Implications of the findings of the present research	199
7.1.5. Key Contributions of the present research	203
7.1.5.1. <i>Contribution to knowledgebase</i>	203
7.1.5.2. <i>Contribution toward methodological perspective (design research methods)</i>	204
7.1.5.3. <i>Contribution towards design community and design management</i>	205
7.1.5.4. <i>Contribution to society and organizations</i>	205
7.1.6. Recommendations from the present thesis	205
7.1.6.1. Criteria/ conditions for the selection of a particular technique	206
7.1.7. Limitations and future scope of the present research	207
7.2. Conclusion	208
8. References	212
Appendices	227

Appendix A: Eye tracking Set up (SMI ETG2 Wireless Analysis Pro)	227
Appendix B: Subject-based data	229
Appendix C: Expert-based data.....	236
Appendix D: Computational data.....	243
Appendix E: Subject-based data for the eye-tracking experiment.....	250
Appendix F: Eye-tracking data for the eye-tracking experiment.....	253
Appendix G: Bivariate correlation results (Pearson’s r, Kendall’s tau_b, and Spearman’s rho) for the subject-based data versus eye-tracking data for the significant signs or the significant eye measures in the eye-tracking experiment.....	260
Bivariate correlation results (Parametric- Pearson correlation) for the subject-based data versus eye-tracking data for the significant signs in the eye-tracking experiment	260
Appendix H: Example of Questionnaire Sample for Chapters 2, 3 and 4	274
Appendix I: Stimuli for eye-tracking experiment in Chapter 4	282
Appendix J: Sign variant-Ranking Questionnaire for Chapter 6.....	283
Appendix K: Glossary	285
List of Publications	290
Curriculum vitae	291

List of Figures

Chapter 1

Figure 1. 1. Pierce Triadic model for evaluation of comprehensibility for OSH sign (ideal case)	5
Figure 1. 2: Unlimited semiosis of OSH sign (non-ideal case)	5
Figure 1. 3: Model for Semiotic Theory	6
Figure 1.4. A Communication-Human Information Processing (C–HIP) model (Wogalter, 2018a) adapted for OSH sign evaluation	8
Figure 1. 5. Example of a test page for the comprehension test of a sign (ISO standard).....	11
Figure 1. 6. Research Gap.....	25
Figure 1. 7. Thesis workflow	31

Chapter 2s

Figure 2. 1. Stimuli	38
Figure 2. 2 . Comprehensibility evaluation model of OSH sign (subject-based technique)....	41
Figure 2. 3. Comprehensibility evaluation model of OSH sign (expert-based technique).....	42
Figure 2. 4. Post hoc analysis (subject-based technique).....	47
Figure 2. 5. Post hoc analysis result showing significantly different pairs in sequence-1 (subject-based technique).....	47

Figure 2. 6. Post hoc analysis result showing significantly different pairs in sequence-2 (subject-based technique).....	48
Figure 2. 7. Post hoc analysis result showing significantly different pairs in sequence-3 (subject-based technique).....	49
Figure 2. 8. Post hoc analysis result showing significantly different pairs in sequence-4 (subject-based technique).....	50
Figure 2. 9. Post hoc analysis result showing significantly different pairs in sequence-5 (subject-based technique).....	51
Figure 2. 10. Post hoc analysis result showing significantly different pairs in sequence-6 (subject-based technique).....	52
Figure 2. 11. Post hoc analysis (expert-based technique).....	65
Figure 2. 12. Post hoc analysis result showing significantly different pairs in sequence-1 (expert-based technique).....	65
Figure 2. 13. Post hoc analysis result showing significantly different pairs in sequence-2 (expert-based technique).....	66
Figure 2. 14. Post hoc analysis result showing significantly different pairs in sequence-3 (expert-based technique).....	67
Figure 2. 15. Post hoc analysis result showing significantly different pairs in sequence-4 (expert-based technique).....	68
Figure 2. 16. Post hoc analysis result showing significantly different pairs in sequence-5 (expert-based technique).....	69
Figure 2. 17. Post hoc analysis result showing significantly different pairs in sequence-6 (expert-based technique).....	70

Chapter 3

Figure 3. 1. Overview of DAN-MMH model.....	79
Figure 3. 2. Comprehensibility evaluation model of OSH sign (computational technique)....	81
Figure 3. 3. Phase-wise diagram of the computational model.....	83
Figure 3. 4. DAN-based DNN	86
Figure 3. 5. Score generation by the computational DAN-based MMH model	89
Figure 3. 6. Heat map showing semantic textual similarity generated by the computational technique.....	90
Figure 3. 7. Heat map generated for the first 72 participants (72/137) for Sign 1 (DNWSH) 91	
Figure 3. 8. Heat map generated for the remaining 65 participants (65/137) for Sign 1 (DNWSH)	92
Figure 3. 9. Post hoc analysis (computational technique)	98
Figure 3. 10. Post hoc analysis result showing significantly different pairs in sequence-1 (computational technique)	98
Figure 3. 11. Post hoc analysis result showing significantly different pairs in sequence-2 (computational technique)	99
Figure 3.12. Post hoc analysis result showing significantly different pairs in sequence-3 (computational technique)	100
Figure 3. 13. Post hoc analysis result showing significantly different pairs in sequence-4 (computational technique)	101

Figure 3. 14. Post hoc analysis result showing significantly different pairs in sequence-5 (computational technique)	102
Figure 3. 15. Post hoc analysis result showing significantly different pairs in sequence-6 (computational technique)	103
Figure 3. 16. Result reflecting upper- and lower-case flexibility of the model	109
Figure 3. 17. Pearson correlation coefficient for the automated model (computational technique).....	110

Chapter 4

Figure 4. 1. Experimental setup: Test for eye defect using Snell's chart	115
Figure 4. 2. Ishihara Test for color blindness	116
Figure 4. 3. Participant in front of a big screen wearing an eye-tracking glass.....	117
Figure 4. 4. Chin and hand rest in an eye-tracking set-up	118
Figure 4. 5. Viewing distance of the stimuli (eye-tracking setup).....	118
Figure 4. 6. One point calibration in an eye-tracking experiment	118
Figure 4. 7. Eye tracking set-up box	119
Figure 4. 8. Data extraction for signs without text	120
Figure 4. 9. Data extraction for signs with text.....	121
Figure 4. 10: Comprehensibility evaluation model of OSH sign (eye-tracking technique) ..	122
Figure 4. 11. Plot showing a relationship between Avg Pupil Diameter Right and comprehensibility percentage	132
Figure 4. 12. Plot showing a relationship between Avg Pupil Diameter Left and comprehensibility percentage	132
Figure 4. 13. Plot showing a relationship between visual intake count and comprehensibility percentage	132
Figure 4. 14. Plot showing a relationship between visual intake duration and comprehensibility percentage	133
Figure 4. 15. Plot showing a relationship between blink count and comprehensibility percentage	133
Figure 4. 16. Plot showing a relationship between blink frequency and comprehensibility percentage	134
Figure 4. 17. Plot showing a relationship between blink duration total and comprehensibility percentage	134
Figure 4. 18. Plot showing a relationship between saccade count and comprehensibility percentage	135
Figure 4. 19. Plot showing a relationship between saccade amplitude average and comprehensibility percentage	135
Figure 4. 20. Plot showing a relationship between saccade duration total and comprehensibility percentage	135
Figure 4. 21. Plot showing a relationship between saccade latency average and comprehensibility percentage	136
Figure 4. 22. Plot showing a relationship between scan path length and comprehensibility percentage	136

Figure 4. 23. Abstract view of analysis for event statistics – correlation of objective measures with subjective scores (eye tracking technique)	141
Figure 4. 24. Abstract view of Analysis for event statistics – correlation of objective measures with subjective scores continued (eye-tracking technique)	142
Figure 4. 25. Abstract view of correlation and significance analysis for AOI statistics (eye-tracking technique).....	143
Figure 4. 26. Post hoc analysis (subjective data- eye-tracking technique)	147
Figure 4. 27. Post hoc analysis result showing significantly different pairs in sequence-1 (subjective data- eye-tracking technique)	147
Figure 4. 28. Post hoc analysis result showing significantly different pairs in sequence-2 (subjective data- eye-tracking technique)	148
Figure 4. 29. Hypothesis test summary for the significant objective measures-Kruskal-Wallis test (eye-tracking technique).....	155

Chapter 5

Figure 5. 1. Correlation plot of subject-based technique versus expert-based technique for COSACOEOT.....	160
Figure 5. 2. Correlation plot of expert-based technique versus computational technique for COSACOEOT.....	161
Figure 5. 3. Correlation plot of subject-based technique versus computational technique for COSACOEOT.....	161

Chapter 6

Figure 6. 1. Conceptualization for a redesign of COSACOEOT (sign)	177
Figure 6. 2. Design of variants of COSACOEOT (Phase 1 and Phase 2)	178
Figure 6. 3. Redesigned variants of Sign 5b (Sign: Cough or sneeze in a tissue- COSIT) ...	178
Figure 6. 4. UCD-based sign design/ redesign cycle	181
Figure 6. 5. Variant 5b.d with maximum comprehensibility	182
Figure 6. 6. Final Variant 6 after slight modification in the variant 5b.d	183

Chapter 7

Figure 7. 1. Implication from the perspective of the Design process and Human-centered approach.....	202
--	-----

List of tables

Chapter 1

Table 1. 1: Geometric shapes and safety colors for safety sign.....	2
Table 1. 2. Literature showing various studies and methods related to the evaluation of Sign	15
Table 1. 3. Research questions, objectives, and hypotheses organized in various chapters and publications.	32

Chapter 2

Table 2. 1. Descriptive statistics of subject-based technique (signs without text)	43
Table 2. 2. Descriptive statistics of subject-based technique (signs with text).....	44
Table 2. 3. Test for normality (subject-based technique)	45
Table 2. 4. Rank of the signs generated by Friedman's ANOVA (subject-based technique) ..	46
Table 2. 5. Test statistics of Friedman's ANOVA (subject-based technique)	46
Table 2. 6. Descriptive statistics for the subject-based comprehensibility score in the Wilcoxon Signed Rank test.....	53
Table 2. 7. Wilcoxon Signed Rank test ranks for the sign pairs: with and without text (subject-based technique).....	54
Table 2. 8. Test statistics for Wilcoxon Signed Rank test for the sign pairs: with and without text (subject-based technique).....	55
Table 2. 9. Effect Size for Wilcoxon Signed Rank test for the sign pairs: with and without text (subject-based technique).....	55
Table 2. 10. Comprehensibility percentage (subject-based technique)	56
Table 2. 11. Reliability statistics of two experts	57
Table 2. 12. Descriptive statistics- Mean and Standard deviation for the scores assigned by two experts (expert-based technique)	58
Table 2. 13. Inter-Expert Correlation Matrix.....	59
Table 2. 14. Intraclass correlation coefficient (ICC) (expert-based technique).....	60
Table 2. 15. Descriptive statistics of expert-based technique (signs without text).....	61
Table 2. 16. Descriptive statistics of expert-based technique (signs with text).....	62
Table 2. 17. Test for normality (expert-based technique).....	63
Table 2. 18. Rank of the signs generated by Friedman's ANOVA (expert-based technique) .	64
Table 2. 19. Test statistics of Friedman's ANOVA (expert-based technique).....	64
Table 2. 20. Descriptive statistics for the expert-based comprehensibility score in the Wilcoxon Signed Rank test.....	71
Table 2. 21. Wilcoxon Signed Rank test ranks for the sign pairs: with and without text (expert-based technique).....	72
Table 2. 22. Test statistics for Wilcoxon Signed Rank test for the sign pairs: with and without text (expert-based technique).....	73
Table 2. 23. Effect Size for Wilcoxon Signed Rank test for the sign pairs: with and without text (expert-based technique).....	73
Table 2. 24. Comprehensibility percentage (expert-based technique).....	74

Chapter 3

Table 3. 1. Descriptive statistics of computational technique (signs without text)	94
Table 3. 2. Descriptive statistics of computational technique (signs with text)	95
Table 3. 3. Test for normality (computational technique)	96
Table 3. 4. Rank of the signs generated by Friedman's ANOVA (computational technique) .	97
Table 3. 5. Test statistics of Friedman's ANOVA (computational technique)	97
Table 3. 6. Descriptive statistics for the computational-based comprehensibility score in the Wilcoxon Signed Rank test.....	104

Table 3. 7. Wilcoxon Signed Rank test ranks for the sign pairs: with and without text (computational technique)	105
Table 3. 8. Test statistics for Wilcoxon Signed Rank test for the sign pairs: with and without text (computational technique)	106
Table 3. 9. Effect Size for Wilcoxon Signed Rank test for the sign pairs: with and without text (computational technique)	106
Table 3. 10. Comprehensibility percentage (computational technique)	107
Table 3. 11. Comprehensibility decision (CDM) based on NC_p	108

Chapter 4

Table 4. 1. Test for normality for subjective variable- comprehensibility scores of 14 signs (eye-tracking technique)	123
Table 4. 2. Normality test result for the eye-tracking variables (N=46) for the 7 signs.....	125
Table 4. 3. Normality test result for the eye-tracking variables (N=46) for the remaining 7 signs	126
Table 4. 4. Normality test result for the eye-tracking variables (N=46) considering AOI for 7 signs	128
Table 4. 5. Normality test result for the eye-tracking variables (N=46) considering AOI for the remaining 7 signs	129
Table 4. 6. Comprehensibility percentage based on subjective data (eye-tracking technique)	130
Table 4. 7. Trend analysis of eye tracking data	138
Table 4. 8. Rank of the signs generated by Friedman's ANOVA (subjective data- eye-tracking technique).....	146
Table 4. 9. Test statistics of Friedman's ANOVA (subjective data- eye-tracking technique)	146
Table 4. 10. Descriptive statistics for the subjective comprehensibility score in the Wilcoxon Signed Rank test (eye-tracking technique)	149
Table 4. 11. Wilcoxon Signed Rank test ranks for the sign pairs: without and with text (subjective data- eye-tracking technique)	150
Table 4. 12. Test statistics for Wilcoxon Signed Rank test for the sign pairs: without and with text (eye-tracking technique)	151
Table 4. 13. Effect Size for Wilcoxon Signed Rank test for the sign pairs: without and with text (eye-tracking technique)	151
Table 4. 14. Descriptive statistics for the significant objective measures-Kruskal-Wallis test (eye-tracking technique)	153
Table 4. 15. Rank for the significant objective measures-Kruskal-Wallis test (eye-tracking technique).....	153
Table 4. 16. Test statistics for the significant objective measures-Kruskal-Wallis test (eye-tracking technique).....	154
Table 4. 17. Effect Size for Kruskal-Wallis test for the variation of objective measures across categories	154

Chapter 5

Table 5. 1. Descriptive statistics of all the signs with three different techniques (subject-based, expert-based, and computational technique)	164
Table 5. 2. Bivariate correlation of the seven signs with three different techniques (subject-based, expert-based, and computational technique)	166
Table 5. 3. Bivariate correlation of the remaining seven signs with three different techniques (subject-based, expert-based, and computational technique)	167
Table 5. 4. Comparison of the techniques using various parameters.....	170
Table 5. 5. Comparative study between the proposed MMH model- based on DAN and the other similar kinds of models.....	172

Chapter 6

Table 6. 1. Result of the ranking technique applied for sign 5 variants (Sign: Cough or Sneeze into a tissue).....	182
--	-----

Abbreviations used in the reports

ANSI	American National Standards Institute
AOI	Area of Interest
C_Score	Comprehensibility Score
CDM	Comprehensibility Decision Making used synonymously with <i>NC_P</i> .
CM	Communication Media
COSACOEOT	Cough or sneeze at the crease of the elbow or in a tissue
COSACOEOT_T	Cough or sneeze at the crease of the elbow or in a tissue (with text)
DAN	Deep Averaging network
DNN	Deep Neural Network
DNT	Do not touch
DNT_T	Do not touch (with text)
DNWSH	Do not Walk or stand here
DNWSH_T	Do not Walk or stand here (with text)
EHS	Environment, Health and Safety
e- Training	Electronic Training

IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
M	Mean of the sample data
MCQ	Multiple choice questions
NAFUP	No access for unauthorized persons
NAFUP_T	No access for unauthorized persons (with text)
NC_p	Non-comprehensibility in terms of percentage- a decision-making factor used synonymously with <i>CDM</i>
OH	Occupational Health
OSH	Occupational Health and Safety
OSH-AT	Occupational Safety and Health- Awareness and Training
SD	Standard Deviation
SE	Standard Error
SI	Standard Interpretation (<i>SI</i>)
Sim	Similarity
SMI	Senso Motoric Instruments
SUS	System Usability Scale
UBC	Use barrier cream
UBC_T	Use barrier cream (with text)
WAFS	Wear a face shield
WAFS_T	Wear a face shield (with text)
WRP	Wear respiratory protection
WRP_T	Wear respiratory protection (with text)



Introduction

Chapter 1: Introduction





Abstract







Comprehension of an OSH sign is a predominant factor and comes in priority over any other factors in the domain of Occupational safety and health (OSH). Signs are hardly understood or misinterpreted by people, which compromises health and safety at work. This chapter starts with a brief discussion on OSH signs and then discusses the motivation behind studying the comprehensibility evaluation of signs. The theoretical constructs behind the study of a sign are discussed. The conceptual framework behind the comprehensibility evaluation of the OSH sign is narrated, followed by a detailed state-of-the-art literature review. The literature discusses the state-of-the-art evaluation techniques practiced. There are discussions about the research gap from which the problem statement has been framed. The study reveals a need for more focus on the techniques/ methods of comprehensibility evaluation that must be dealt with. The study aims to facilitate the evaluation of OSH sign design to ensure better comprehensibility. Three main hypotheses have been framed here. The expected outcome of the work is defined, and the chapter concludes with the framework of the thesis with a briefing of the chapters.

1.1. Introduction to Occupational Safety and Health (OSH) signs

OSH Signs disseminates a general safety message. This is achieved by a combination of color and geometric shape which, by the addition of a graphical symbol, gives a particular safety message related to Health and safety in the workplace (I. S. O. 3864-1, 2011). The colors attributed to different OSH signs have unique characteristics that carry a safety connotation (I. S. O. 3864-1, 2011). The general meaning attributed to geometric shapes and safety colors is tabulated in Table 1. 1.

Table 1. 1: Geometric shapes and safety colors for safety sign

Geometric shape	Meaning	Safety color	Contrast color to safety color	Graphical symbol color	Examples of use
 Circle with a diagonal bar	Prohibition	Red	White	Black	 No smoking
 Circle	Mandatory action	Blue	White	White	 Wear protective gloves

 Equilateral triangle with radiused outer corners	Warning	Yellow	Black	Black	 Slippery surface
 Square	Safe condition	Green	White	White	 First aid
 Square	Fire equipment	Red	White	White	 Fire emergency telephone

The color white includes the color for phosphorescent material under daylight conditions with properties as defined in ISO 3864-4.

1.2. Motivation and justification for studying comprehensibility evaluation of a sign

The purpose of the OSH signs is to grab rapid attention to objects and situations affecting safety and health at work and to gain a rapid understanding of a specific message (I. S. O. 3864-1, 2011). However, in real situations, such signs are ambiguous, poorly understood, or misinterpreted by people resulting in disaster or compromising safety and health at work (Chaudhuri et al., 2021; Duarte & Rebelo, 2005; Montagne, 2013; Wogalter & Laughery, 1996). Hence there is a need to study and evaluate such sign(s) for safety at work. Though numerous studies have been conducted on various aspects of safety signs, there are seldom studies solely related to the techniques or methods of comprehensibility evaluation.

Safety signs come with many features, which include color, position, graphics, and flashing (Ding et al., 2022), shape, content (Chen et al., 2018), visual content (pictograms) (Dowse, 2021), position of text/ signal word (Tam et al., 2003); cognitive features like familiarity, concreteness, simplicity, meaningfulness, and semantic closeness (A. H. S. Chan & Ng, 2010); utility in respect to its context of use (Montagne, 2013), etc. Many such features and various user factors (Ng & Chan, 2015) have been studied; however, the study for techniques related to the evaluation of comprehensibility is inadequately addressed (Arcia et al., 2019). The reason behind selecting comprehensibility over other factors is that shreds of evidence exhibit that the OSH signs are less understood and misinterpreted by the employees, and understanding the sign is more important than any other factor (Dewar, 1988; Montagne, 2013). Besides, understandability has been considered one of the most important criteria for determining the effectiveness of a symbol (Lerner & Collins, 1980; Maulina et al., 2022). The sign intends to communicate the message of the situation (hazard) appropriately. This can be ensured through a standard evaluation method/ technique that confirms the comprehensibility of the existing and re-designed sign(s). Literature supports the fact that there is a dearth of published work in understanding the methods/ techniques of comprehensibility or understanding of signs (Zashchirinskaia, 2020). The following sub-section discusses the

motivation behind studying the techniques of comprehensibility evaluation of OSH signs. The comprehensibility of an OSH sign refers to understanding the meaning of the sign and understanding the message the sign intends to communicate. The Peirce model and Saussure's Model form the basis of the interpretation of the sign. Both the models are slightly discussed here considering the visual pictogram that must generate the correct meaning. Besides, the models are not studied here and out of the scope of this study. Designing various comprehensibility evaluation techniques represents the various ways to determine whether a sign is understandable/ comprehensible.

1.2.1. Theoretical constructs

“Every message is made of signs; correspondingly, the science of signs termed semiotic deals with those general principles which underlie the structure of all signs whatever, and with the character of their utilization within messages, as well as with the specifics of the various sign systems, and of the diverse messages using those different kinds of signs.”(Jakobson, 2010).

According to Charles Sanders Peirce, *“A sign is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the interpretant of the first sign. The sign stands for something, its object. It stands for that object, not in all respects, but in reference to a sort of idea, which I have sometimes called the ground.”* This study is based on the Semiotics model (Etymology: from "Semeion", a Greek term which means "sign"). Semiotics is the area of research that studies signs, symbols, and communication through the processes of signification. In simple words, it's the “science of signs” and depicts how meaning is constructed and understood. Semiotics has two founders the Swiss linguist Ferdinand de Saussure (1857–1913) and the North American philosopher Charles Sanders Peirce (1839–1914) (Lagopoulos & Boklund-Lagopoulou, 2021).

According to Peirce, there are three semiotic elements- a sign (representamen), an interpretant, and the object. The representamen is the form that the sign takes. Some theorists call it a “sign vehicle”. An interpretant is not an interpreter, but rather it refers to the sense made of the sign; as being produced in the mind of the receiver. An object is something beyond the sign to which it refers; it's the subject matter of a sign and an interpretant (a referent) (Crow, 2017; Zeman, 1977). *In simple words, a sign represents an object/ fact to the interpretant.* The present study is motivated by Peirce's triadic model of the sign where the sign is interpreted by target users keeping the object constant (context is Industry). Any other context is beyond the scope of the thesis. If the interpretation is correct, the sign is comprehensible otherwise the sign is incomprehensible. The interpretant is the receiver of the communication, i.e. the participants that produce meaning (mental concept) in his/her act of interpretation. The sign (representamen) is the pictogram of the message “DO NOT WALK OR STAND HERE”. The sign might also indicate something else. Thus the sign may indicate “DO NOT KEEP SHOES”, “DO NOT WEAR SHOES”, or “OPEN SHOES”. *“Pictograms tend to be more abstract, attempting to convey messages by analogy or symbolic representation”* (Montagne, 2013). The interpretant might thus extend (or abuse) the sign to make it mean something else. This is the “unlimited semiosis” of Peirce (Waugh, 2006)

which happens in the case when the participants are unable to comprehend the sign and make unusual and unlimited guesses. The triadic model for the ideal case of comprehension of OSH signs is represented in Figure 1. 1. The case when there is unlimited semiosis is represented in Figure 1. 2.

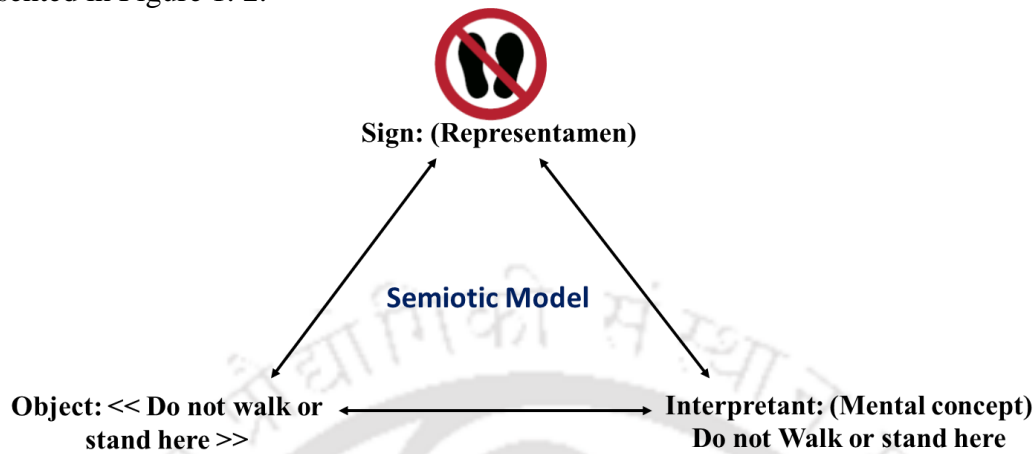


Figure 1. 1. Pierce Triadic model for evaluation of comprehensibility for OSH sign (ideal case)

1.2.1.1. Model of Semiotic Theory

According to Berger and Luckmann’s proposition (1966) in the grand theory of social constructionism, the semiotics model is made up of (i) three main concepts, namely sign, context, and meaning (ii) a construct (semiosis), and (iii) the proposition that depicts that signs that convey accurate context can trigger appropriate responses from the target audience (Berger & Luckmann, 2016) (Figure 1. 3).

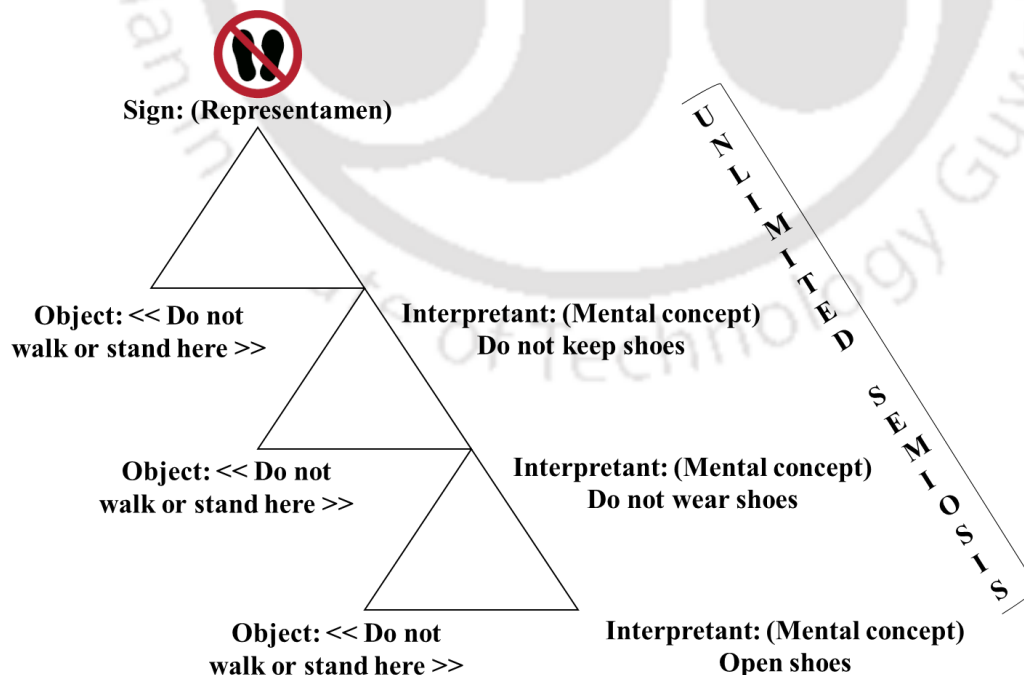


Figure 1. 2: Unlimited semiosis of OSH sign (non-ideal case)

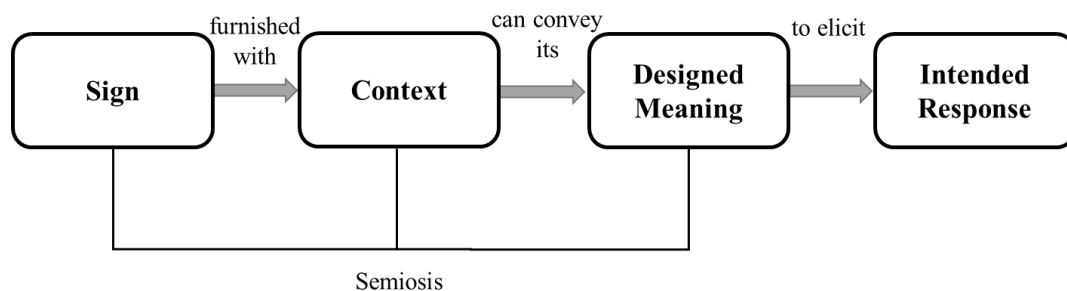


Figure 1. 3: Model for Semiotic Theory

A *sign* (visual pictogram) consists of both the form it takes in physical reality (a signifier) and how it is conceived or interpreted by its audience (the signified). A sign (visual pictogram) must have both; it is the integrated whole that results from the combination of the signifier and signified (what) (Saussure, 1983) (Lagopoulos & Boklund-Lagopoulou, 2021). The *context* in the context of semiotics mentions those aspects of an interaction that provide relevant and specific meaning to the particular exchange that is happening. This enables the receiver in this communication to make proper sense of the interaction and derive the intended meaning, which is meant for communication. Meaning is generated by the receiver when they make sense of the sign by connecting and interacting with their surrounding reality (environment). *Construct* (semiosis) narrates the relationship between the sign, the context, and the meaning. According to Peirce (1873), semiosis is defined as any process that involves signs for the referents to produce meaning. Within the context of the semiotics model above, semiosis describes the interrelation and interaction between the concepts of sign, context, and meaning. The *proposition* is that a sign with an accurate social and cultural context can communicate the sender's intended meaning more clearly, increasing the chances that the recipient will decipher the meaning and respond appropriately.

1.2.2. Conceptual Framework

The Communication-Human Information Processing (C-HIP) model (Lin et al., 2021; Wogalter, 2018a; Wogalter & Sojourner, 1999) serves as the foundation for this investigation. This C-HIP model has been partially applied here to study human information processing (comprehensibility evaluation stage) of the OSH signs (Figure 1.4). According to the model, once the information is delivered by the source (in this case, the source is the researcher who is experimenting) via a channel (OSH signs with text and without textual information), the recipient of the information must pay close attention for a sufficient period to completely process and understand/ comprehend the information. Once the target user comprehends the transferred message, the decision to engage in the target activity (e.g., ‘Do not walk or stand here’, ‘Use barrier cream’, or ‘Cough or sneeze at the crease of the elbow or in a tissue’ as prescribed) depends on how well the information fits the receiver's ideas and attitudes (or whether it is convincing enough to change them). Each of these stages has the potential to create a bottleneck that inhibits information from moving to the next stage (s). Particularly, information processing at later stages can affect whether and how information is processed at earlier stages. For instance, a person may choose to disregard a piece of specific information (in an earlier stage) if they feel the information is unnecessary or insignificant

(later stage). An attention switch occurs when a person shifts, modifies, or flips their focus from something else—like a warning—to something else. It is linked to the ideas of salience, noticeability, and attention-getting (conspicuousness or prominence). Salient qualities are connected with greater size, high contrast, distinct hue, relative distinctiveness, apparent movement, and other types of stimulus change, and they benefit from attention switching. For example, symbolic graphics of OSH signs and texts can aid with attention-switching. *Comprehensibility* is a salient concept that is part of the C–HIP model's comprehension stage when applied to OSH signs.

In the context of OSH sign research, *comprehensibility* refers to the extent to which the intended message of a sign is interpreted correctly by a group of target users (Lin et al., 2021). The comprehensibility score is the score/rating of how much the sign is understandable by each participant. The scores are coded as 0 or 1 ('0' means incomprehensible, and '1' means comprehensible). The comprehensibility percentage is calculated by the total number of participants assigned '1' (comprehensible sign), divided by the total number of participants in the study multiplied by 100. For example, if 15 % of the respondents interpreted the meaning of a given OSH sign correctly, the comprehensibility percentage of this sign would be 15 (15%). However, a less comprehensibility score/percentage suggests that the sign is ineffective at communicating its intended message to the target audience. The present research restricts the study to the comprehension stage. Later phases can be studied in the future. The C-HIP model's inventors claim that the model can be used as an investigative tool to assist in finding the cause(s) of a warning's ineffectiveness which is beyond the scope of this research and can be a future study.

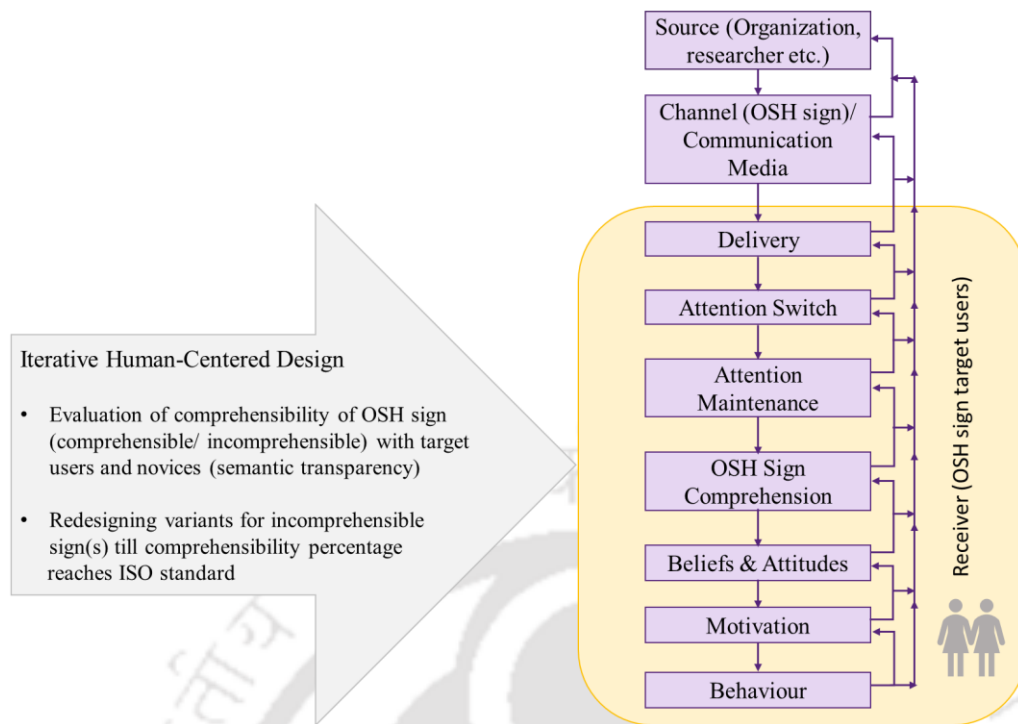


Figure 1.4. A Communication-Human Information Processing (C-HIP) model (Wogalter, 2018a) adapted for OSH sign evaluation

It is important to note that in the current thesis, the existing OSH signs from ISO have been adopted to study comprehensibility. Here, an attempt has been made to evaluate the comprehensibility of those signs by the industrial employees in terms of their understanding or making meaning out of it. For this purpose signs were presented with and without text which describe the purpose of the sign. In the whole experiment, the visual elements of the sign have been kept unaltered as per ISO standards. The researcher did not play with the visual elements (graphics, color, foreground-background ratio, size, physical or digital presentation, etc.) to check the variation of the comprehensibility by the intended audience. Therefore, it can be stated that the present research deals only with the semantics of the existing sign by the audience while the existing standard signs are presented with or without text.

1.3. State-of-the-art review of the techniques for comprehensibility evaluation

‘Comprehension’ or ‘Comprehensibility’ of a sign is defined as the ability of a target user to accurately interpret the meaning of a sign (Berrio et al., 2022). The International Organization for Standardization (ISO) is a worldwide federation responsible for preparing International Standards. According to ISO 3864, a sign is considered acceptable if 67% of the target population understands the sign's meaning correctly in a comprehension test (Berrio et al., 2022; ISO 3864-1, 2011). ISO method for testing comprehensibility (ISO, 2014a) deals with what proportion of participants can comprehend a graphical symbol correctly. The test measures how much a graphical sign communicates its intended message. The purpose is to ensure that the target audience readily understands signs. The initial plan was to develop

OSH sign(s) without any added text, which must be understood by a large audience. However, the supplementary text has to be added later with a sign to make it comprehensible/ understandable.

A recent study (Arcia et al., 2019) insisted on the need for rigorous evaluation methods/ techniques for sign comprehension tests. ISO 9186 technique/ method is the recognized standard for evaluating the comprehensibility of OSH signs and symbols; however, modification is required to improve and maximize its utility in terms of low literate audiences and complex OSH symbols (Arcia et al., 2019). This subsection tries to report various studies on the techniques/ methods of comprehensibility evaluation of OSH signs used to date. The literature review is segregated into three subsections based on data collection approaches. The first subsection, 1.3.1, discusses the literature on two techniques (subject-based and expert-based techniques). Subsection 1.3.2 discusses the literature on computational techniques. Both these subsections are based on subjective data collection. Subsection 1.3.3 discusses the literature on the studies of the eye-tracking technique. This technique is based on objective data collection. The literature discussion mainly focuses on the methods and techniques rather than the outcome of the study.

1.3.1. Evaluation of comprehensibility utilizing the subject-based and expert-based techniques

The study of comprehensibility/ understandability can be traced back to the mid-30s (Brainard et al., 1961; Janda et al., 1935; Lerner & Collins, 1980). Although the exact date of the first study for comprehensibility evaluation of OSH sign(s) is unknown, it can be regarded as the moment when the first attempt to standardize any sign occurred, and here is where the evaluation process began. The earliest attempt to standardize safety signs can be traced back to 1914 in the United States (Bureau & Hansen, 1914).

This thesis found an excellent way of classifying the literature for the techniques of comprehensibility evaluation based on the source of data and data collection techniques used for comprehension analysis. Both subject-based and expert-based techniques use subjective data from the participants/ users/ subjects. The intervention of the expert and their coding of data for evaluation in the expert-based technique makes them different from the subject-based technique.

Before classifying the evaluation based on different comprehensibility techniques and discussing their literature, it is essential to understand the nomenclature of the techniques.

A *Subject-based technique* is a self-reporting technique, meaning that the ratings/ scores provided by the subjects/users/ participants are directly used in the study to evaluate the comprehensibility level of a sign(s). The subjects assess their understandability or comprehensibility of a sign and give their ratings/ scores. No expert is involved here in coding the scores/ rates. An exhaustive literature survey reveals that there are very few studies where the direct scores of the subjects have been used for studying the comprehensibility of OSH signs. Studies with slight variations have been found in the literature.

Ahmadi and his groups (Ahmadi et al., 2021) studied the guessability and comprehensibility of the pictograms using the ISO 9186 technique. Next, they studied five cognitive features, which include meaningfulness. They used a numeric rating scale (0-100, where 0 means very unfamiliar and 100 means very familiar). The guessability and cognitive features studies used 351 subjects and 21 pharmaceutical pictograms. Their study reveals that most pictograms are not comprehensible to the participants, and the study concluded that not all signs are designed effectively. *Aljabry and the group* used multiple choice questions (MCQ) and their variation, like true or false, besides open-ended questions in testing dental patients' comprehensibility about information disseminated to them (Aljabry et al., 2023). The study found that augmented reality is more effective than a leaflet in knowledge retention in patients. MCQs have also been used for studying health-related signs (infographics) (Comello et al., 2016; Leiner et al., 2018) as they are easy to use, objective, and less external administration is needed. In a special case of MCQ, forced-choice questions, for example, yes/no (Brewer et al., 2012; Liu et al., 2009; Snyder et al., 2017), have been used for studying comprehensibility evaluation of health visualizations, but the study has been unsuccessful due to the little acquaintances of the target population with the format of MCQs. A study by *Krmizioglu and Yaman* (Kirmizioglu & Tuydes-Yaman, 2012) has highlighted a significant loophole in the comprehensibility study of traffic signs with MCQ. It is found that the signs studied are not only unrecognizable to the participants or misinterpreted by them, but they are also comprehended in an opposite way which is an even more critical issue. This necessitates the researcher to include the open-ended question in the study, which helped capture different comprehensibility levels.

Expert-based technique means that the experts provide the ratings/ scores. The subjects might provide short or long descriptive textual answers coded/ rated/ scored by the experts for studying the comprehensibility of signs. ISO method for testing comprehensibility can be classified into expert-based techniques. It provides an instruction sheet to the audience(s)/ subject(s) and instructs them to write the response below the question provided. The question that they ask is, "What do you think this symbol means?". The respondent(s) are asked to answer this question or write "Don't know" if they cannot comprehend the sign. Two experts who are not respondents are provided with the likely responses for each sign. The responses are categorized on a 5-point scale, indicating '1' as 'correct', '2a' as 'wrong', '2b' as 'wrong, and the response given is the opposite of the intended meaning', '3' as 'the response given is don't know', and '4' as 'no response is given'. For each variant of sign(s), the frequency followed by the percentage of responses in each category is counted. An example of a test page for a comprehension test according to the ISO standard is shown in Figure 1. 5.



This symbol appears in hotels, airports, stations, town centres, and in guidebooks.

You might find it on a signpost, door, or wall, or printed in a guidebook.

What do you think this symbol means?

What action should you take in response to this symbol?

Figure 1. 5. Example of a test page for the comprehension test of a sign (ISO standard)

The Study of the various literature reveals the use of various methods/ techniques of sign comprehension evaluation following and deviating from ISO-9186 standards and following the subject-based, expert-based, or both techniques. *Corbett and his group (2008)* (Corbett et al., 2008) studied fifteen (15) airline graphical symbols recommended by ANSI and safety cards with 785 participants. The written response was often lengthy and needed manual scoring throughout. Five experts established comprehension criteria (Corbett et al., 2008) before evaluating the signs' correctness. Participants were rated as type-1 if they addressed the specific question asked and type-0 if they did not address the specific question asked. Type-1 response was further categorized as follows:

- i. *Certain* = response was correct and complete.
- ii. *Likely* = response was mostly correct but missing key element(s).
- iii. *Arguably* = response contained words or ideas that indicated partial correctness but were unclear or ambiguous.
- iv. *Suspect* = response contained words or ideas that were related but misinterpreted.
- v. *Opposite* = response contained words or ideas that were related but contradictory to the correct response.

- vi. *Wrong* = response was wrong.
- vii. *None* = response was “don’t know,” and *blank* = no response was given.

The categorized responses were then transformed using a weighted algorithm to issue a comprehensibility score. The frequency for each comprehension category for each sign is divided by the number of participants. Blank responses were discarded from the scoring procedure. These calculated percentages are multiplied by the comprehension category weights as follows: ‘certain * 1.0’; ‘likely * 0.75’; ‘arguable * 0.50’; ‘suspect * 0.25’; ‘opposite * -1.0’; and ‘wrong and none * 0.0’, and they are summed to achieve the pictogram comprehension score, i.e., the percentage of total comprehension for any particular pictorial/pictogram. This approach is largely a recapitulation of ISO 9186's (2001) methodology for comprehension tests, with the exception that ISO 9186 only stated three degrees of positive understanding, but the current analytical standard contains four. To consider this change, the comprehension category weights are modified. This revision is done to better characterize largely accurate responses that have missing key elements, as well as poorly linguistically constructed answers that partially address the visual/pictogram. In circumstances where two or three questions were asked to increase the comprehension estimation, a composite categorization (score) shows the comprehension found by the interaction of the individual answers, not just that created by the mean comprehension estimates. For each question(s) for each pictogram, the comprehension ranged between 28.8% to 96.3%, with a mean comprehensibility of 65%. The composite comprehension scores ranged between 39.8% and 85.3%, with a mean of 64.7%. The result of the study shows that only 45.8% of the individual question comprehension scores exceeded the ISO standard (67%), and only 8.3% exceeded the ANSI criteria (85%).

Similarly, the pictogram comprehension scores are below the mean 75% comprehension of the ANSI symbols tested. In total, comprehension has been much below acceptable limits. The study demonstrates the requirement for safety signs to be designed and implemented considering the novices who do not have any knowledge about the design and operation of transport aircraft, equipment for an emergency, or aircraft emergency procedures.

Caffaro and Cavallo in the year of 2015 (Caffaro & Cavallo, 2015) studied the factors influencing comprehension of signs (signs without text) associated with agricultural machinery and which sign is frequently less comprehended. A paper-pencil questionnaire study was conducted with twelve (12) safety signs and 281 users. The users/ subjects needed to choose one correct response among the four choices. The whole questionnaire took 8 minutes on average to complete. The question was closed-ended to avoid guessing the responses of the users. The correct responses have been collected from the ISO standards. The partially correct and incorrect responses were designed considering a pre-test phase where responses were collected from ten participants. The tests were corrected by two independent safety experts. Correct answers scored ‘2’, partially correct answers scored ‘1’, and incorrect answers scored ‘0’. None of the 12 signs used in the study could conform with the ANSI Z535.3-2011 comprehensibility score (85%). The highest score was (80.8%) as

achieved by one sign. It is found that the more the sign was seen previously, and the more years of experience with the machinery, the better the comprehension of the participants.

Bagagiolo and group in the year 2019 (Bagagiolo et al., 2019) studied 37 migrant workers. They conducted interviews to test the comprehension of four ISO 11684:1995-based safety signs related to agricultural machinery. The comprehension is tested using an open-ended question where users/ subjects have been asked to describe the meaning of the safety sign in their own words. Besides the authors who conducted the interviews, there was an interpreter, an expert in safety risk in the agricultural sector, who helped the migrants in the interview. The interview lasted between 20 to 40 minutes, and the responses were audio-recorded. This study implements an expert-based technique. The interviews have been transcribed, and qualitative analysis has been done. The responses are categorized into correct, which are conceptually correct with concrete terms; incorrect answers include wrong, confusing, and no answers. None of the signs secured a comprehension level of 85% suggested by ISO/ANSI standards. The study found that Romanians scored the highest comprehension performance (68.8% of correct answers), followed by Indians (35.4%), with Pakistanis being last (32.4%). The comprehension of safety signs is significantly affected by their previous experience as a farmworker. The results highlight the importance of abundant training for migrant workers on the meaning of safety signs. Besides, the study suggests redesigning signs, considering some signs' features to enhance pictorials' cross-cultural comprehension.

Arcia and group in the year 2019 (Arcia et al., 2019) studied the common methods for testing comprehension. They did two case studies. In case 1, they studied two asthma control infographics (OSH signs) with ten patients. The intention is to identify the patient's asthma control and severity level. A portable spirometer has been used to perform pulmonary function testing. Information is communicated through printed visuals for comprehension testing. After the comprehension evaluation, the patients are asked about their reactions to the information they had received and whether they would take any action based on the message presented. The questions were open-ended, and the interviewer recorded the responses textually. Probing questions have been asked to get the granular details. Two gist questions and four verbatim questions were asked. Interviews were audio-recorded and transcribed later. Two independent experts who were not present during the interview scored the responses. A third expert acts as a tiebreaker whenever there is a dispute/ or disagreement between the two experts. Four categories of responses are designed for the study as follows:

- i. Less than minimum comprehension: when one or both gist response(s) are incorrect; 0 to 4 verbatim responses are correct.
- ii. Minimum comprehension: when both gist responses are correct, but no verbatim responses are correct.
- iii. More than minimum comprehension: when both gist responses are correct, and one or more verbatim response(s) is/ are correct.
- iv. Complete comprehension: when all the responses are correct.

On average, the comprehension assessment during the interview took about 4 minutes and 23 seconds. This approach is a slight modification from the ISO 9186 method, and the final

result can be thought of as a cognitive interview. The participants reacted positively to the visuals and expressed intentions for self-management and self-care.

In case 2, Arcia groups studied comprehension of non-graph visualizations by the forty (40) heart failure patients that display longitudinal changes in “patient-reported outcomes” such as heart failure symptoms, mental health symptoms, and functional status. The comprehension assessment has been conducted under four conditions (i) text only, (ii) text plus pressure gauge icons, (iii) text plus red-yellow green gradient line, and (iv) text plus line graph. Each participant viewed every condition. The same technique/ method is used in this case as used in case 1. Interviews were conducted, probing questions were asked to get granular details for assessing their comprehension, and finally, they were transcribed; verbatims were recorded manually. The responses for their intention to do something were categorized into five based on a 5-point Likert scale with '1' as 'no intention' to '5' as 'Certainty'. Two independent respondents have been responsible for scoring the transcribed responses as per ISO 9186.

Ahmadi and his groups (2021) (Ahmadi et al., 2021) studied 351 Iranian participants to assess the relationship between the guessability of the pharmaceutical pictograms, the cognitive sign features, and prospective-user factors. Three hundred fifty-one (351) medicine consumers were considered for the study with no or little medical knowledge and were exposed to 21 pictograms (signs). Three questionnaires were used for the study. Besides the demographic questionnaire, another questionnaire based on ISO 9186 was used to assess guessability and comprehension. A third sheet for cognitive sign features included Familiarity, Concreteness, Simplicity, Meaningfulness, and Semantic Closeness. A 100-point rating scale was used to capture the participants' views about the five sign features, where '0' represents 'very unfamiliar' and '100' represents 'very familiar.' The signs were shown on the monitor, and they had to guess the answer within 10 seconds. They were provided extra time if they started guessing; otherwise, after ten seconds, the response sheet was left blank, and the following pictogram was shown. The verbal responses were audio recorded and written down by one of the research team members. Then the subjects were exposed to the second round of subjective evaluation to assess their cognitive features. A correlation study has been conducted between the sign's cognitive features. The result analysis shows that '8' out of '21' are comprehensible according to the 85% criteria recommended by ANSI, and '10' out of '21' are comprehensible according to the 67% criterion recommendations of the ISO. The rest of the four signs were not comprehensible. This study used an expert-based technique in the comprehensibility evaluation of the signs where an expert/ research team decoded the interviews. The subject-based technique has been applied in the case of the cognitive feature study, and it can be seen that their ratings were directly used for the study without the intervention of independent experts.

The other studies related to the comprehensibility evaluation of health and safety signs are tabulated (Table 1. 2).

Table 1. 2. Literature showing various studies and methods related to the evaluation of Sign

Author (Year)	No. of Subjects	Technique/ Method	Outcome
Burt et al. (1999)(Burt et al., 1999)	'21'	Initially, the appropriateness test was conducted with nine (9) symbols, filtered to four symbols of correct lifting posture for the comprehension test, open-ended questions where workers wrote the sign's meaning, and later coded using various lifting criteria.	The study helped identify correct lifting postures and showed a significant increase in the adoption of correct lifting postures in the presence of the symbol compared to a controlled condition. The study recommends using the designed lifting postures on the packaging box to communicate safety information better.
Brantley & Wogalter (1999)(Brantley & Wogalter, 1999)	'20' each for '5' conditions.	Thirty-one (31) safety symbols were shown to collect responses. Written test with probe question after all responses were provided, oral test without probe question, oral test with probe question after responses for all symbols were provided, and oral test with probe question after each partially correct or incorrect response. Two raters scored the response.	Results show that the test methods that included follow-up questioning elicited more information from participants and increased comprehension rates in both oral and written test formats. The results have implications for cost-effective symbol design and evaluation.
Monsoor and Dowse (2004) (Mansoor & Dowse, 2004)	Phase 1: '30' Phase 2: '20'	The student developers consulted graphics designers for three rounds, and initial sketches were modified. Phase 1: Target population: low literate; one researcher and one interpreter interacted with them and asked whether pictograms (10) convey the intended message. Phase 2: Six (6) pictograms, data collection follows phase 1, asked whether the revised pictogram version communicates the intended message better than the previous version of pictograms.	In phase 1, the newly designed pictogram sequence is correctly comprehended by 66.7%, which complies with ISO standards. In Phase 2, 95% of the participants correctly comprehended the new pictograms, which complied with the ANSI criterion of 85% correctness recommendations for a consultative approach to the new pictogram design.
Reis et al. (2015)(Reis et al., 2015)	'12'	A virtual environment (VE) prototype has been used to study the subject's hazard perception, simulator sickness, sense of presence, and level of engagement is assessed by observation, post hoc questionnaires, and semi-structured interviews. The safety sign questionnaire asked whether the sign influenced them and captured their attention and visibility. A 9-point scale captured their responses for sign assessment and a 7-point scale for VE experience assessment. The total experimental duration was 60 minutes.	As workers grow old, their ability to comply with such signs is adversely affected by age-related deficits. Safety signs are not always effective in communicating the intended message or necessary information to the employees.
Banares et al. (2018) (Bañares et al., 2018)	'90'	Forty (40) road signs were studied. In study 1, the responses were coded as 'correct and complete' (coded as +2), 'partially correct' (coded as +1), 'incorrect' (coded as 0), or the 'opposite of the true meaning' (coded as -2). The signs that failed to comprehend according to the ANSI (85%) standard were tested further in Study 2. In study 2, the	The comprehension evaluation reveals that more than 50% of signs fall below the ANSI standard criterion of 67%. Applying ergonomic design principles significantly improves driver comprehension levels.

same drivers rated the sign with five ergonomic principles through a scaled numerical questionnaire. In study 3, five (5) experts rated the compatibility of the three designed variants according to the correlated ergonomic principles with a significant relationship with the comprehensibility of the drivers.

Vigoroso et al. (2020)
(Vigoroso et al., 2020)

‘15’

They aimed to develop safety materials based on visual communication. They incorporated a user-centered design approach. Initially, they did need analysis using a focus group, and six (6) experts trained them using visuals. Two (2) prototypes were designed to evaluate comprehension, preferences, and suggestions for improvement. Again seven (7) different workers responded to the questionnaire study for the comprehension test. During the second focus group, three tasks were assigned to the target. In task A, four (4) alternatives for content were provided, and one has to be chosen which conveys the target message better. In Task B, they had to analyze which of the two prototypes was better comprehensible and most suitable, and in Task C, different graphical solutions were discussed. Then the sorted solution was presented to twenty-three (23) workers in a 1-hour session. The satisfaction was rated on a 6-point scale (“How satisfied are you with the quality of course materials?”: ‘0’ meaning ‘not at all satisfied’; ‘5’ meaning ‘very much satisfied’). A similar test was done for the control group, and the groups were compared.

The final version of the training material is significantly more satisfactory for trainees than traditional training material. Stereotyped designs for characters like doctors, fire safety managers, workers, etc., are found to be less effective. Regarding safety figures, employers with agricultural clothes have been given priority in the context of Italian agriculture. The participants advised accessories such as pitchforks and hats, which is as per the evidence about agricultural workers’ stereotypes.

Kaufmann & Ramirez-Andreotta (2020)
(Kaufmann & Ramirez-Andreotta, 2020)

‘11’

They used participatory design, formative evaluation, open-ended discussion, and three examiners to evaluate the factsheet.

Graphic elements outweigh text; concepts per figure should be minimal to ensure comprehension. The need for formative evaluation, information design prototyping, and a cultural model of risk communication is demonstrated.

Lin et al., (2021) (Lin et al., 2021)

‘108’

They tested Amazon crowdsourcing for participatory evaluation of the pharmaceutical pictograms. The study used fifteen (15) USP pictograms. One extra question, “What might happen if the instruction is not followed?” was added to what is usually followed in ISO 9186. Three (3) experts independently scored the responses as ‘1’ for ‘correct,’ ‘0’ for ‘incorrect,’ and ‘-1’ for ‘critical confusion.’ The final score was provided after discussion and consensus of the experts.

Seven (7) pictograms’ comprehensibility score is below 85% (ANSI standards). Again eight (8) pictograms are misinterpreted. The outcome did not conform with the ANSI (2017) standards. MTurk is a suitable, reasonable, low-priced, and time-efficient platform to evaluate pharmaceutical pictograms and find underlying design issues, even when the age group is limited to older adults. About half of the pictograms (46%) failed to satisfy the ANSI comprehension standards.

It is seen from the above studies and the existing literature that in every case of the scoring process, two or more (Lerner & Collins, 1980; Maulina et al., 2022; Ben-Bassat et al., 2021) experts are engaged in the repeated (Yeow et al., 2014) scoring process, which may induce biases due to their background, previous work experience, fatigue, etc. Can this be avoided by any means? Then the search started for other methods where the manual evaluation process for comprehensibility can be avoided. The search resulted in disappointment as no such study has been found directly; however, few computational algorithms could be found, which helped to develop the automated technique in this research work.

1.3.2. Evaluation of comprehensibility utilizing computational evaluation technique

Evaluation of comprehensibility based on computational techniques is in its infancy. A rigorous literature search revealed no studies to date where the comprehensibility evaluation of signs is automated. Although deep learning (DL) has a stellar reputation for solving construction safety problems, little literature demonstrates the connection between DL and safety management. (Pham et al., 2021). DL has demonstrated its expertise in evaluating risk and health (Ajayi et al., 2020). Deep learning based on convolution neural networks (CNN) is seen to solve safety issues when there is an interaction between autonomous vehicles and pedestrians. The approach reduces data transmission lag, improving the developed system's prediction accuracy (Zhu et al., 2022). Deep reinforcement learning coalesced with risk assessment functions has been exploited for developing a driving strategy with a DL approach to execute autonomous driving with minimum risk (Li et al., 2021). The application of DL has also been found in studying the emotions between residents of high urban areas and the built environment (Xiang et al., 2021). Transfer learning has been exploited to identify the fatigue of crane operators and construction workers.

Literature reveals numerous studies on road sign detection and recognition and their applications. Studies can be classified into color-based detection (Estevez & Kehtarnavaz, 1996; Kehtarnavaz et al., 1993; la Escalera et al., 2003; Shadeed et al., 2003); shape-based detection (Hsu & Huang, 2001; Sandoval et al., 2000; Schiekkel, 1999), color-shape based detection (Fang et al., 2003; Vitabile et al., 2002; Xu et al., 2019), and recognition and classification (de la Escalera et al., 2004; Fang, Fuh, et al., 2003; Morina & Ahma, 2022), color based extraction and texture based recognition (Balali et al., 2017), sign information extraction and identification (de la Escalera et al., 2004), sign detection and recognition using Deep Neural Network (DNN) (Avramović et al., 2020; Luo et al., 2018; Megalingam et al., 2022; Zhang et al., 2020). Character recognition to read text from an image is a buzzword of research now (Mainkar et al., 2020; Panchal et al., 2022).

The above algorithms and studies directly consider the image as the input and hence cannot be considered in this research. It is important to note that the OSH sign(s) in the form of an image cannot be used directly for evaluating comprehension of a sign from the perspective of the users. The descriptive responses/ rates/ scores the participants provided must be evaluated. The research will not replace the target audience from the study as their opinion about the sign's comprehensibility is significant here. The search is for new or existing

technologies that will cut off the repeated work of the experts engaged in the evaluation process and hence assist in the evaluation procedure. The computational method of processing text comes to mind because scores or rates again need expert intervention. An exhaustive review of the literature reveals that no published literature talks directly about the computational technique of comprehensibility evaluation of OSH messages (OSH signs) on behalf of the experts.

Further analysis of the type of participant's responses in the literature (1.3.1) for evaluating comprehensibility concluded that the participants' descriptive written responses could be considered input for the algorithms to automate the comprehensibility evaluation. The search is for technique(s) that will take text as input, score the messages, and categorize the signs as comprehensible/ incomprehensible, the same as human experts do. It is found that deep learning is a state-of-the-art machine learning technique being used for greater NLP tasks. The following discussion briefly studies the deep learning association with NLP tasks (for processing written responses).

1.3.2.1. Deep Learning and Natural Language Processing (NLP) tasks

Developing the automated system for the comprehensibility evaluation of OSH message(s) involves performing Natural Language Processing (NLP) tasks. Numerous articles demonstrate the state-of-the-art DL algorithms for NLP Tasks and their performance. There is evidence for using 'word order' for neural network-based NLP tasks. Recursive neural network (R_NN) has been successfully applied for machine translation (S. Liu et al., 2014) and paraphrase detection (Socher et al., 2011). Convolution neural networks (C_NN) have outperformed R_NNs on various NLP-related tasks (Kim, 2014). R-NET and NLNet from Microsoft Research Asia achieved human quality performance for the Stanford Question Answering Dataset evaluation task on the exact match (E.M.) and fuzzy match scores (Rajpurkar et al., 2016; Zhou et al., 2020). In the recent past, pre-trained models like bidirectional encoder representations from Transformers (BERT) (N. B. Chaudhuri et al., 2021; Sung et al., 2021) and generative pretraining of language models (Radford et al., 2018), XLnet (Yang et al., 2019) seems to exhibit good generalization results across various NLP tasks, and good proficiency in reasoning based NLP tasks (Zhou et al., 2020).

Deep neural networks (DNN) have shown their profound history in NLP tasks. According to the literature, the Encoder-Decoder (E.D.) model (Graves & Schmidhuber, 2005; Tripathy et al., 2021) has the disadvantage of losing a considerable amount of context information, which made it necessary to incorporate RNN into the current E.D. model. This additional layer modifies the previous E.D. model by taking information from the opposite direction. It allows the entire model to consider the context of the preceding and following words (Bahdanau et al., 2014). Two attention models are studied: global and local attention with different alignment techniques (Luong et al., 2015). Global attention evaluates every word in the source sequences while computing the context, whereas local attention only considers a subset of words. Global attention-like RNN-search architecture (a neural machine translation based on recurrent neural networks (RNN)) (Bahdanau et al., 2014) is computationally expensive. Transformer architecture (Vaswani et al., 2017) follows the usual prevalent

architecture of the encoder-decoder model with more mechanisms for parallelization, making it more flexible. The sentence need not be fed word by word; instead, the sentence is masked by the combination of self-attention in a way very similar to how a human reads. The positional embeddings help determine the position of each word or the distances between different words in the input. These qualities encourage the use of Transformer models as the foundation for much larger pretraining models that incorporate several layers of the Transformer along with enormous datasets (Tripathy et al., 2021).

Despite having those features, the Transformer model requires a lot of computation (compute-intensive), considering training the model in learning new concepts and using those concepts for predictions. Hence, this study has exploited the deep averaging network (DAN) based encoding model (Iyyer et al., 2015), which is computationally less intensive than transformer models. The limited dataset for training and the similar related NLP datasets present made us ideate for exploiting the pre-trained model (Cer, Yang, Kong, Hua, Limtiaco, John, et al., 2018). DAN is quite a straightforward approach that avoids needless complexities (Iyyer et al., 2015) and is a pre-trained model. The output is non-linearities based, again based on simply averaging the words. There is no requirement for deep syntax or word order analysis to get a better result, making it worthier than transformer models. The reason behind choosing the DAN model over other models for sentence recognition and comprehension is that each layer of DAN needs only a single matrix multiplication. The complexity increases only with the increased number of layers rather than the increased number of nodes. According to the literature, the DAN model performs better than neural bag-of-words (NBOW) and other complex syntactic models on sentence and document analysis and question-answering tasks with less training time. DAN has a record of training out-of-domain data in the question-answering task, which is quite difficult for other recursive neural networks. In the proposed model, DAN exploits embedding average, which works best for NLP tasks (Iyyer et al., 2015).

The three techniques (subject-based, expert-based, and computational-based) discussed are all based on subjective data that may be biased by participants' opinions, backgrounds, and feelings. Considering the biases, an objective data collection technique is considered using an eye-tracking device.

1.3.3. Evaluation of comprehensibility utilizing an eye-tracking evaluation technique

The study on using an eye-tracking device to evaluate the comprehensibility of an OSH sign has been less but found profoundly in driver safety-related pictograms, road, and traffic safety, agricultural sector. The following review mainly discusses the methodology/techniques adopted while studying sign(s) comprehensibility.

Hsu and the group in the year 2013 (Hsu et al., 2013) studied the comprehensibility status of the safety sign in an airplane (airline safety briefing cards) and its relationship with eye movement (fixation). The experiment was conducted with 51 participants using an eye-tracking device. The eye tracking device consisted of an ASL (Applied Science Laboratory)

mobile eye wearable- eye camera, scene camera, and PC with monitor. After each participant wore an eye device, they were shown the section on “Land Evacuation on the safety card”. They could look at the pictograms as long as they wanted, similar to onboard passengers. The data collected were categorized into five regions. Upon completing this phase, participants were made to evaluate their comprehension of the pictograms.

The questionnaire was open-ended, and participants were asked to elaborate on the meaning of the pictorials. They were provided unlimited time to complete the survey. The written response was often lengthy and needed manual scoring throughout. The comprehension evaluation result was scored by the technique used by Corbett, McLean, and Cosper (Corbett et al., 2008) discussed in 1.3.1. One of the signs could only meet the ISO comprehension rate of 67%, while none could meet the ANSI criteria of 85%. The study aimed to inspect the efficacy and usability of airline safety cards. It is found that safety information is not well transferred to passengers and potential passengers. When participants were exposed to pictograms relating to "how to use emergency door", eye-tracking measures such as fixation time and fixation counts were found to be the longest, while the comprehension score was greater. Other pictograms also show a positive correlation between comprehensibility, fixation time, and fixation count. Their study reveals that less interest attention (longer Time to First Fixation) is associated with lower comprehensibility. The significance of increasing the comprehensibility of passengers by spending time reading the safety card is highlighted in this study. It is also found that textual information makes safety information more meaningful.

Siswandari and Xiong in the year 2015 (Siswandari & Xiong, 2015) studied ten (10) traffic symbols with 42 participants. A practice session was designed before the original experiment to acclimatize the participants to the experimental procedure. The signs were randomized and displayed using Tobii studio software (Tobii Technology). Each sign stayed on the screen for ten seconds, followed by a question- what did the sign mean? The test was open-ended, and the participants were asked to guess the answer. The response was verbal and audio-recorded, followed by an ‘end’ word. Real-time eye movements have been recorded during the experiment, and the result reveals that comprehensibility level affects eye movement. The study reveals that as sign comprehension gets harder, the blink rate slows down, the pupil diameter increases, and the “Time to First Fixation (TTFF)” gets longer. The guessability study shows that comprehensibility scores differed significantly among different signs.

Cholewa and group in the year 2018 (Cholewa et al., 2018) developed a pictogram-based interface for medical-related communication between patients and medical practitioners or family members. The system used an eye-tracking technique to help patients scan the message in the forms of texts, pictograms, and their combinations, and they communicated with the experts through this interface. The system considers different age groups, their eye-sights, and the body and head position and adjusts accordingly. The pictograms have been selected and developed considering the intercultural nuances, and they are intuitive, which requires no extra teaching. The study reveals that images and eye-coded communication help patients overcome their communication barriers due to a lack of motor functions. The pictograms seem to improve comprehension of medical instructions and convey medical

information. People, irrespective of their educational background, can communicate with this system.

Babic and group in the year 2020 (Babic et al., 2020) studied eye movements, i.e., gaze behavior when a driver encounters a traffic sign that is unfamiliar to him. Tobii Pro 2 eye-tracking glasses have been used for the experiment. The glasses were equipped with four cameras that located the eye pupil position and an HD camera that recorded the area in front of the participant. For each participant, the following eye-tracking measures were recorded-

- i. Total gaze duration (in seconds), the span of average gaze duration (in seconds), and the number of times each participant looked at the sign (view count)
- ii. Total fixation duration (in seconds), the average fixation duration (in seconds); and the number of times a participant fixated their gaze on the area of interest (fixation count).

The study was conducted using sixty-four (64) traffic signs with sixty-two (62) participants. The participants rated their general knowledge about the sign on a 5-point Likert scale ('1' signifies 'No knowledge' and '5' signifies 'Absolute knowledge'). Each participant has been trained with 3 sample test signs different from the original ones. This is done to ensure that the participants understood the experimental procedure correctly. Eye-tracking device calibration has been done according to the manufacturer's instruction manual, and a projector is used to show the stimuli. Each slide consisted of 4 signs, and participants were made to view each slide for 10 seconds maximum. After 10 seconds, the participants can move or nod their heads and answer the question about the sign type and their meaning in a relaxed mode. The response has been categorized into three-

- i. complete knowledge (the participant knows the correct type and meaning of the sign)
- ii. partial knowledge (the participant knows either the approximate sign's context or the type of the sign correctly)
- iii. no knowledge (the participant does not know the type's name nor the traffic sign's meaning).

Five independent experts imparted a score for each sign regarding information load and simplicity (whether the sign is easily understandable/ comprehensible or hard to understand). A lower information load is assigned a value of 1.45 out of 3 points, and a fairly simple-to-understand sign is assigned 1.70 out of 3 points. An average score is calculated for each sign. Signs that consisted of more information (more content) and were unknown to the participant consistently gave the longest and average fixation and gazing durations. A greater quantity of information in visual processing produced a larger number of fixations with an overall longer duration. The participants with higher knowledge about the sign or at least know the sign type showed shorter viewing and fixation time. The participants who were unaware of the sign took more processing time.

Kumetha and group in the year 2021 (Kummetha et al., 2021) studied the driver comprehension ability of the roadway environment to keep accountability of safety during driving automation. An increase in the case of accidents with autonomous vehicles led to the study. The roadway environment included the significance of the objects, traffic signs, or hazards during driving. A simulation environment was created, and ninety (90) participants were studied. The driving study was done in a fixed base driving simulator with a dash-mounted Fovio FX3 eye-tracker. Both subjective and objective data are collected. Among a few probing questions, one is about the sign. The answer is binary: yes or no, which has been randomly placed in each task to counterbalance the answers; and to check for instances of guessing by examining gaze behavior. 'No' means that the sign means something else or no sign present. The comprehension time has been considered from the onset of the cue. A methodology for monitoring real-time driver comprehension is proposed. ANOVA revealed that with increased task difficulty, the percentage time spent gazing from the onset of the cue also increased for most cases. It is seen that as the task difficulty increases, there is a decrease in driving performance with an increasing percentage of time spent gazing to achieve comprehension.

Ding and group in the year 2022 (Ding et al., 2022) studied the effectiveness of building evacuation signs. The participants are police and student trainees of public security. All participants were blindfolded before the experiment, and their eye masks were replaced with eyeglasses during the experiment. Four groups of features linked to evacuation signs (color, position, graphics, and flashing) are considered during the studies in a building with T-junction scenarios. SMI BeGaze has been utilized to track the passengers' eye movements. The experiment started with an alarm sound, and they needed to search for evacuation signs. The experiment stopped when they found the staircase. SMI Begaze has been used for data analysis. Thirty-three (33) participants were involved in every ten experiments. The participants were not aware of the training, and each participant participated only once. The signs have been placed according to the standards. The gaze data for signs have been recorded and analyzed. There were two groups among those who detected the sign- one is those who followed the instruction/ guidance of the sign and those who did not follow. The duration of fixation on each sign, the number of participants detecting the signs during the evacuation, and following the signs while evacuating were discussed and compared for each experiment.

1.4. Summary of literature review and research gap

The literature discussed in sub-section 1.3 is kept limited for its brevity. From the above discussion, several other pieces of literature, and the recent survey, it is seen that there are studies on comprehension of safety signs and symbols and their variations/ effects due to cultural differences (Kim et al., 2012); low literate audiences (Mansoor & Dowse, 2004; Beusekom et al., 2017; Zimmerman & others, 1989); and age-groups (Malhotra et al., 2022). There are studies regarding the effects on comprehensibility due to drinking habits, frequency of attendance in onsite safety promotion activities, job positions (Tam et al., 2003); colors; design features, and driver factors (Ng & Chan, 2008); color highlights on pictograms (McDougald & Wogalter, 2014), etc.

An exhaustive survey on pictograms or visual aids (Montagne, 2013) reveals that the widely recognized and widely used safety signs and warning symbols are never developed and evaluated against any standards, nor do they meet the ANSI/ ISO standards for comprehension. More than 50% of pictograms (signs) are not correctly identified by even 30% of test participants (Bañares et al., 2018; Davies et al., 1998; Duarte & Rebelo, 2005). According to the ISO (ISO 3864-3, 2006) and ANSI (ANSI Z535.3, 2002), symbols must reach a criterion of 67% or 85% correctness, respectively, in a comprehension test to be considered “acceptable” (Chan & Chan, 2011; Ng et al., 2011). However, past experimental results indicate that there are severe issues in the study of the usability of safety signs, and it is found that the effectiveness of some safety signs is low in terms of conveying safety messages (Handcock et al., 2004; sLiu et al., 2005; Tam et al., 2003). Hence, the signs that are believed to be highly comprehensible and effective are hardly effective in real and varied contexts. The context or environment varies in terms of various industries and their level of danger involved; hence the referential metrics for one sign’s comprehensibility might vary from industry to industry. Therefore, a strict constraint on the threshold for evaluating comprehensibility is not expected. For example, the metrics for comprehensibility of a fire safety sign in highly inflammable and explosive oil and gas companies must be comparatively high compared with other food manufacturing companies. The fixing of higher metrics is due to the vulnerability of the oil and gas companies toward the fire, which must be considered. From the sign design perspective, considering such a gap is hardly seen while evaluating the signs.

According to the Caire group (Caire et al., 2013), novices outperform experts in designing symbols comprehensible by any novice group. In the present scenario, designing signs considers only experts from industries and hardly considers novices (ISO-7010, 2003). As the evaluation and design are conducted by experts, there is little scope for evaluating whether the designed signs are comprehensible by the novices before actually using the signs. This is a challenge in the evaluation strategy, which must be addressed. The signs are designed for the employees of organizations who are not designers or OSH experts; hence the design of OSH signs must consider novices for semantic transparency, a challenge that must be addressed.

There is a high recommendation to examine the people's comprehensibility regarding the meaning of the internationally developed pictograms in each country. Though ISO 9186 provides a standard for the comprehensibility evaluation of signs, there are recommendations for variations in evaluation methods of comprehensibility (Arcia et al., 2019).

The extensive literature review reveals that comprehension test captures responses that include either the subjects' written descriptions (meaning and or interpretation) of the signs or verbal interpretations of the signs, which are coded by experts or subjects are made to select the alternatives from varied choices. The subjects are asked to rate or score their opinions (ISO standard or variations) based on their understanding of the sign. The questionnaires used include various rating scales, multiple-choice questions, and forced-choice questions (a special case of MCQ- limited to binary responses: yes/ no) (Brewer et al., 2012; Liu et al., 2009; Snyder et al., 2017), short answer questions (McDowell et al., 2019; Ruiz et al., 2013;

Tao et al., 2018), forced choice and short answer paired with response time (Tao et al., 2018), task performance (Yin et al., 2017), etc. The above studies reveal that the rates or scores provided by the participants sometimes fail to give appropriate responses. This context lacks studies. This issue can be resolved by written responses from the participants, which can be scored or rated by experts. The human experts evaluate the responses manually, where there is again a chance of inconsistencies in judgment among experts due to various reasons like personal biases, ambiguity, persuasion, self-doubt, previous exposures, and unwanted errors that may creep in due to the repetitive (tedious) job of evaluation, and tiredness. There is a research gap in this context where there is an attempt to cut off or reduce the manual and time-consuming procedure of evaluating the comprehensibility of OSH signs and assist the experts in such a scenario to minimize unwanted errors, ambiguity, and biases in decision-making. The present scenario exhibits a lack of research articles focusing on the lacunae of the manual procedure (need of manpower, brain, logistics, repetition of similar work, money, and time) in assessing and evaluating OSH sign effectiveness in industries. This gap needs to be addressed.

The human experts might get fatigued by a questionnaire volume and provide ambiguous ratings or scores. Besides biases in the evaluation procedure, there can be biases in the data acquired from the participants who can be biased by their previous experiences, background, context, situation, persuasion, etc.

Evaluation of comprehension can have two dimensions (Meek & Tietz, 2022; Siswandari & Xiong, 2015). The comprehensibility analysis of subjective data based on user perception can be error-prone, which can be biased data too. The comprehensibility analysis of objective data can be a solution based on a quality measure that is unobtrusive and more reliable. There needs to be more research in analyzing the data (subjective and objective) collected for the comprehensibility study of OSH signs and various comprehensibility evaluation procedures. The evaluation strategy lacks validation which must be addressed. The overall research gap is presented in Figure 1. 6, and the research questions (Q1- Q8) are presented in section 1.5.

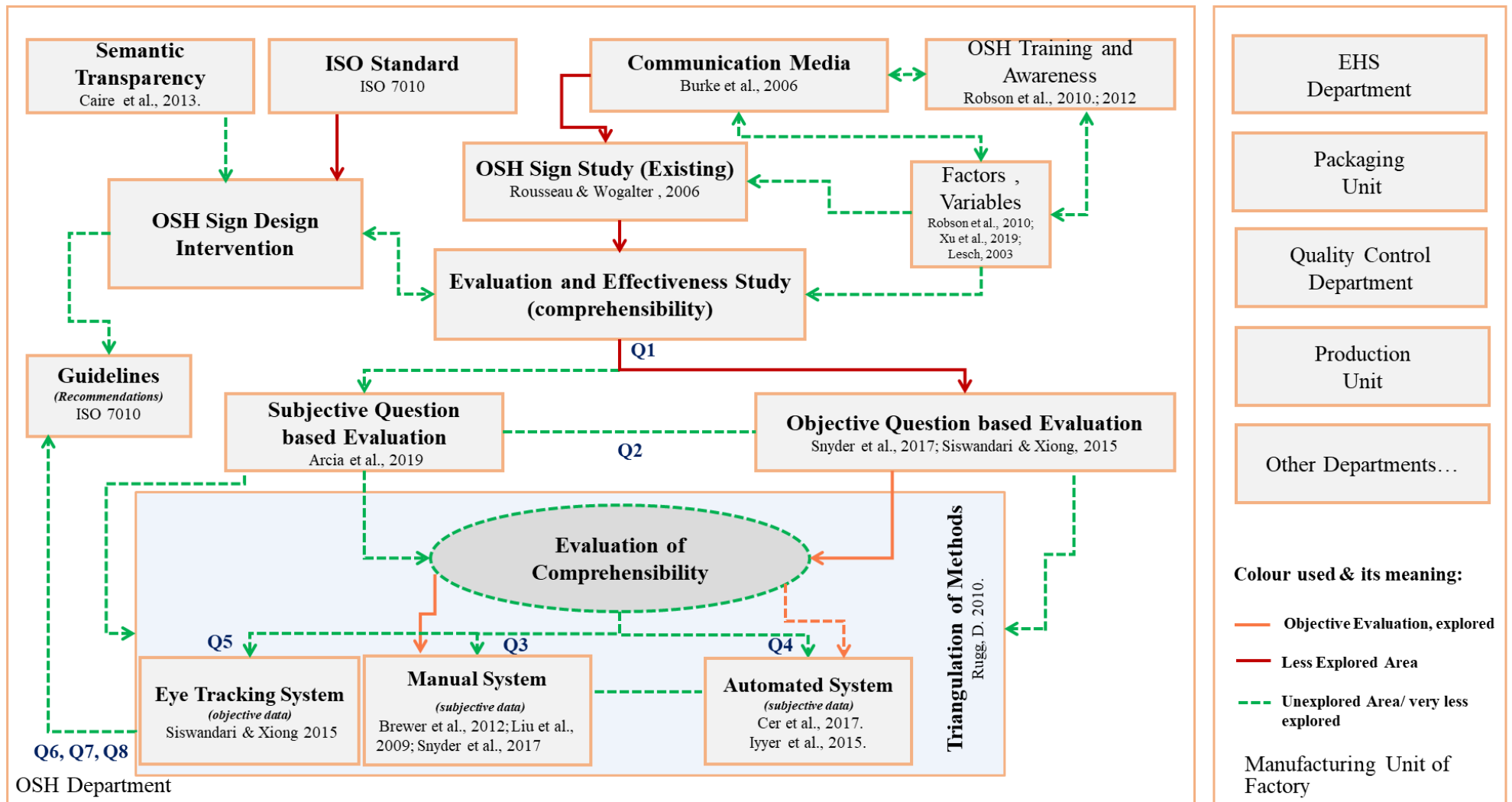


Figure 1. 6. Research Gap

1.5. Research questions (RQ)

- RQ 1. What are the various evaluation techniques available for the comprehensibility of OSH Signs?
- RQ 2. What are the various data collection techniques that can be adopted for the comprehensibility study of the OSH Sign?
- RQ 3. How to assess the level of comprehensibility of the OSH Sign by subject-based and expert-based techniques?
- RQ 4. How to assess the level of comprehensibility of the OSH Sign(s) by computational method?
- RQ 5. How to assess the level of comprehensibility of the OSH Sign by eye-tracking method?
- RQ 6. What is the difference in level of effectiveness when the computational method is compared with manual evaluation (subject-based and expert-based techniques)?
- RQ7. What is the difference in the level of effectiveness when an eye-tracking method is compared with manual evaluation (subject-based technique)?
- RQ 8. How can we develop a framework for the comprehensibility evaluation of OSH signs and testing its validity?

1.6. Problem statement

Industries use various signs and symbols to communicate OSH information to the employees of the organizations. Practical experience in Food manufacturing industries and literature study helped to get the information that the signs used by them are from standard databases (internet) or designed by the safety experts according to their needs and the context. ISO/ANSI gives a standard, but many studies reveal that most safety signs do not comply with the standards of the ISO and ANSI comprehension criteria. Hence, the signs that are believed to be highly comprehensible and effective are hardly effective in real and varied contexts. Hence it could be concluded that no methodology or guidelines till the present date are effective enough to check the effectiveness of the communication media (sign) used for generating OSH awareness. Sign comprehensibility differs depending on countries and states and many other factors.

It is seen that many times, few standard signs, as well as the designed signs, are not well understood by the employees, be it normal officials, management people, or the workers. The result can be dangerous as there are various kinds of hazards in factories that the employees must understand well for their health and safety purposes. Hence there is a need for research regarding the standardization of evaluating the comprehensibility/ understandability of signs and their effectiveness in communicating proper and intended messages. There is no proper

framework for checking whether existing, newly, or freshly designed signs are effective in terms of comprehension/ understandability. There is no framework for checking the effectiveness and simultaneous redesign of signs for the effective satisfaction of the stakeholders. This is a constraint in the sign design process. This framework is significant considering the evaluated signs, which are understandable and could easily communicate the hazard or danger and may reduce or avoid health issues/ injuries/ accidents in industries. There is a lack of studies applying formative evaluation in evaluating and designing the OSH Sign system (Kaufmann & Ramirez-Andreotta, 2020; Wogalter et al., 2002).

Again, there is a lack of literature involving novice users in the sign's design. *Semantic transparency* is the term that involves involving novice users in the evaluation and design phase so that the designed sign is understandable by ordinary people. The motive behind such an evaluation strategy is that if ordinary people understand a sign, it is implied that the employees who are trained or experts in a particular domain will understand the meaning of the sign.

There is a lack of study regarding various comprehensibility techniques simultaneously considering their advantages and disadvantages. Again, rigorous literature mining reveals that there is no mention of reducing the human experts' burden and optimizing work. Evaluation of the survey questionnaires is a repeated work done by experts for evaluating the designed sign. A question must be asked whether this work can be automated. Hardly any study could be found in this area. There are very few studies regarding the understandability/ comprehensibility of signs from the perspective of employees of organizations/ ordinary people, and there is hardly any study regarding standardization of evaluation method and framework, which must be followed for the existing signs and the redesign or fresh design of OSH Signs in workplaces.

Again the data collected from the questionnaires/ interviews are overt and obtrusive; hence subjects must be studied in such a way that there are no or minimum subjective biases (Jahedi & Méndez, 2014; Podsakoff et al., 2003; Redelmeier et al., 2003). There is a need to minimize the subjective biases affecting the result of the study.

Leaving aside the ISO standard technique for comprehensibility evaluation developed long back, there needs to be a revised proper framework or methodology for evaluating comprehension of industrial OSH Signs (existing or newly designed). A detailed study is required in this domain.

1.7. Aim

To facilitate the evaluation of Occupational Safety and Health (OSH) Signage to ensure better comprehensibility.

1.8. Objectives

- Objective 1:* To study the existing comprehensibility evaluation techniques for the OSH signs.
- Objective 2:* To collect data regarding the comprehensibility of the OSH sign through a survey and eye-tracking study.
- Objective 3:* To study the level of comprehensibility of the existing few OSH sign(s) by subject-based and expert-based techniques based on the data collected from the survey.
- Objective 4:* To study the level of comprehensibility of the existing few OSH signs (s) by a computational technique based on the data collected from surveys.
- Objective 5:* To study the level of effectiveness of comprehensibility of the existing few OSH signs (s) by eye tracking technique.
- Objective 6:* To study the level of effectiveness of the computational method in comparison to the manual evaluation (subject-based and expert-based techniques).
- Objective 7:* To study the level of effectiveness of the eye-tracking method in comparison to manual (subject-based technique) evaluation.
- Objective 8:* To develop and validate the framework for the comprehensibility evaluation of OSH signs.

1.9. Hypothesis

Based on the research questions, three main hypotheses have been framed, with one of the hypotheses having three (3) working hypotheses.

Hypothesis 1

Hypothesis 1 (H_1): Comprehensibility scores are significantly different but correlated, while evaluations are compared among the subject-based technique, expert-based technique, and computational technique.

Working Hypothesis a (H_a): Comprehensibility score is significantly different but correlated, while expert evaluation is compared with the subject-based evaluation.

Working Hypothesis b (H_b): Comprehensibility score is significantly different but correlated, while expert evaluation is compared with computational evaluation.

Working Hypothesis c (H_c): Comprehensibility score is significantly different but correlated, while subject-based evaluation is compared with computational evaluation.

Hypothesis 2

Hypothesis 2 (H₂): There is a significant correlation between the subjective response of the comprehensibility and the objective measures (eye tracking variables: average pupil diameter, visual intake count, visual intake duration, blink count, blink duration, blink frequency, saccade count, saccade amplitude, saccade latency, saccade duration, scan path length).

Hypothesis 3

Hypothesis 3 (H₃): The level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information.

1.10. Expected outcome

- i. Knowledge about various comprehensibility evaluation techniques with their pros and cons.
- ii. Introduction of a novel technique of transforming data from qualitative to quantitative binary categories (comprehensible and incomprehensible).
- iii. Assistive automated tool for the organizations' experts for large-scale and parallel comprehensibility evaluation of existing and newly designed signs.
- iv. The evaluation framework will help to design a more comprehensible sign from the users' perspective.
- v. Comprehensibility of signs can be decided based on a flexible threshold (according to the danger level/ constraint/ severity of the situation/ criteria) fixed by the organizations' experts. The threshold can be tuned, and this criterion can be used as a cutoff for determining the successful comprehensibility of a sign.
- vi. A good correlation exists between subject-based, expert-based, and computational techniques.
- vii. A correlation between eye-tracking variables and comprehension levels is elicited. Significant measures have been extracted from eye-tracking data, and their variation across comprehensibility categories (comprehensible or incomprehensible) can be seen.
- viii. Developed a new variant of a sign as a proof of concept. Developed a UCD-based sign evaluation and design/ redesign cycle.
- ix. The research contributes to the design process, design methodology, design management, and organization's EHS department by introducing an evaluation framework consisting of four different techniques which will enable methodical (all four techniques), consistent, unbiased, parallel, and fast evaluation (by computational technique) and thereby accelerate the design process. When the organization needs unbiased, unobtrusive data, the eye-tracking technique can be

used to predict the categories/ or levels of comprehensibility of signs by viewing the eye-movement data.

- x. The research assists in effective OSH sign comprehensibility evaluation, contributing to safety culture.

1.11. A Framework of the Thesis

The thesis is divided into seven chapters (Figure 1. 7) based on the flow of the work. An overview of the chapters is also provided (Figure 1. 7).

Table 1. 3 shows the research questions, the objectives, the hypothesis proposed, and based on them the articles that have been published or communicated.



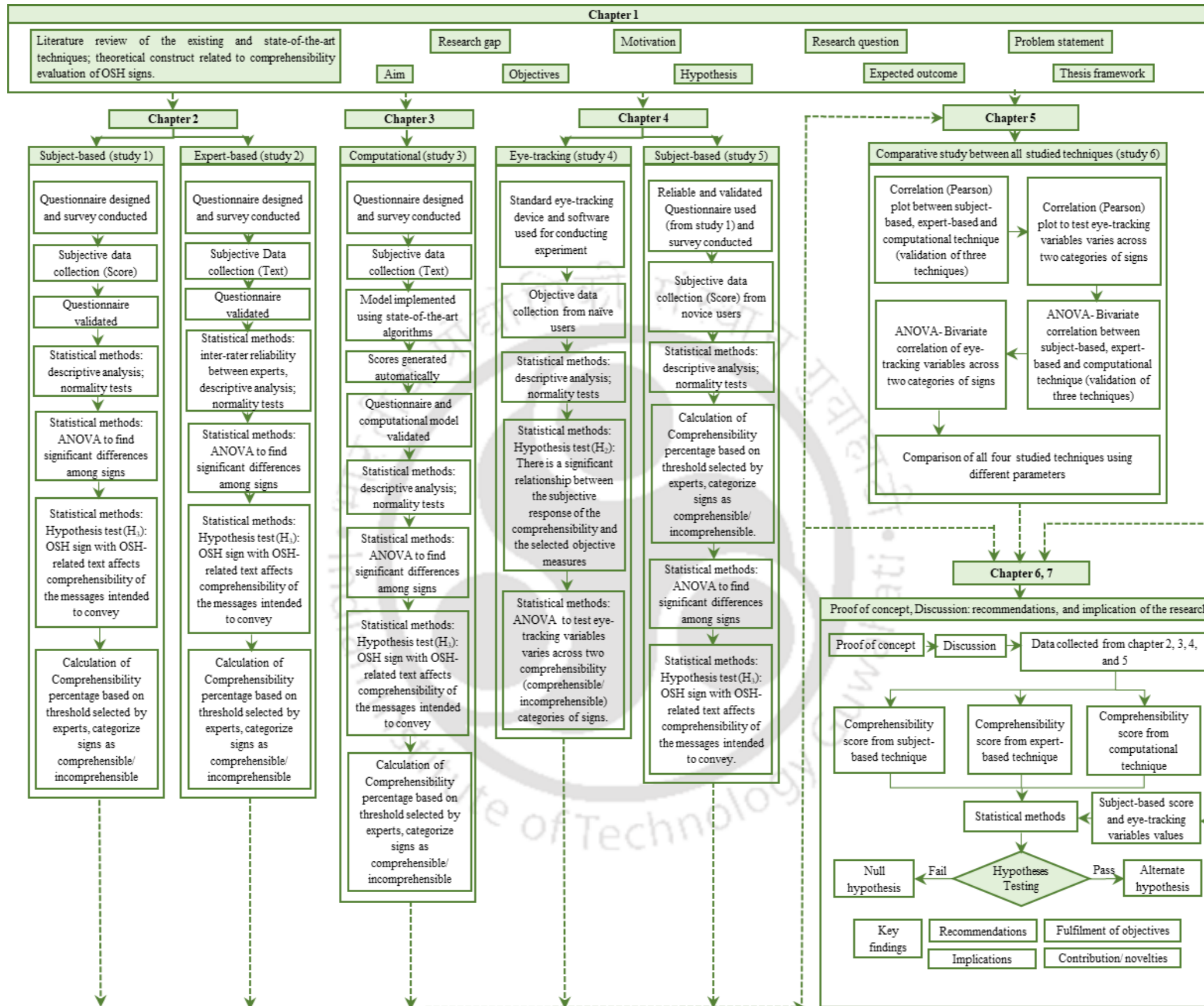


Figure 1. 7. Thesis workflow

Table 1. 3. Research questions, objectives, and hypotheses organized in various chapters and publications.

Chapter No.	Research Questions	Objectives	Hypothesis	Publications
1	RQ 1: What are the various evaluation techniques available for the comprehensibility of OSH Signs?	Obj 1. To study the existing comprehensibility evaluation techniques for the OSH signs.	Not applicable. [Note: Collected data from existing literature to study the existing evaluation techniques].	<ul style="list-style-type: none"> Chaudhuri, S.B., Majhi, M. and Karmakar, S. " A scoping review on role of communication media for effective OSH awareness and training "International Journal of Reliability and Safety (IJRS)", Vol. 15 (1-2), 1-17 (2021) https://doi.org/10.1504/IJRS.2021.119641 [Scopus Indexed, Compendex journal]
2,3,4	RQ 2: What are the various data collection techniques that can be adopted for the comprehensibility study of the OSH Sign?	Obj 2. To collect data regarding the comprehensibility of the OSH sign through a survey and eye-tracking study.	Collected data related to the comprehensibility of the OSH sign through surveys and an eye-tracking study to prove hypotheses H₁ , H₂ , and H₃ .	<ul style="list-style-type: none"> Chaudhuri, S.B., Majhi, M. and Karmakar, S. " Evaluation of Comprehensibility of a Sign by Triangulation Method", In Chakrabarti, D., Karmakar, S., Salve, U.R. (eds) Ergonomics for Design and Innovation. HWWE 2021. Lecture Notes in Networks and Systems, vol 391. Springer, Cham. https://doi.org/10.1007/978-3-030-94277-9_154 [Scopus Indexed, Compendex journal] Chaudhuri, S.B., Majhi, M. and Karmakar, S. "A Man-Machine-Hybrid Model for Comprehensibility Evaluation of Safety Messages using Assistive Deep Learning Approach." [Communicated].
2	RQ 3: How to assess the level of comprehensibility of the OSH Sign by subject-based and expert-based techniques?	Obj 3. To study the level of comprehensibility of the existing few OSH sign(s) by subject-based and expert-based techniques based on the data collected from the survey.	Collected comprehensibility data and calculated the level of comprehensibility to prove hypothesis H_a .	<ul style="list-style-type: none"> Chaudhuri, S.B., Majhi, M. and Karmakar, S. " Evaluation of Comprehensibility of a Sign by Triangulation Method", In Chakrabarti, D., Karmakar, S., Salve, U.R. (eds) Ergonomics for Design and Innovation. HWWE 2021. Lecture Notes in Networks and Systems, vol 391. Springer, Cham. https://doi.org/10.1007/978-3-030-94277-9_154 [Scopus Indexed, Compendex journal]
3	RQ 4: How to assess the level of comprehensibility of the OSH Sign(s) by computational method?	Obj 4. To study the level of comprehensibility of the existing few OSH signs (s) by a computational technique based on the data collected from surveys.	Collected comprehensibility data and computed level of comprehensibility to prove hypotheses H_b , H_c , and H₃ .	<ul style="list-style-type: none"> Chaudhuri, S.B., Majhi, M. and Karmakar, S. "An Analytical Man-Machine-Hybrid Model for Comprehensibility Evaluation of Safety Messages Using Deep Learning Approach." [Communicated]
4	RQ 5: How to assess the level of comprehensibility of the OSH Sign by eye-tracking method?	Obj 5. To study the level of effectiveness of comprehensibility of the existing few OSH signs (s) by eye tracking technique.	Collected comprehensibility data and calculated the level of comprehensibility to prove hypotheses H₂ and H₃ .	<ul style="list-style-type: none"> Chaudhuri, S.B., Majhi, M. and Karmakar, S. "Exploration and analysis of significant eye-tracking measures, subjective measures, and their inter-relationship: A study of comprehensibility of OHS signs." [Communicated]
5,6	<p>RQ 6: What is the difference in the level of effectiveness when the computational method is compared with manual evaluation (subject-based and expert-based techniques)?</p> <p>RQ 7: What is the difference in the level of effectiveness when an eye-tracking method is compared with manual evaluation (subject-based technique)?</p> <p>RQ 8: How can we develop a framework for the comprehensibility evaluation of OSH signs and testing its validity?</p>	<p>Obj 6. To study the level of effectiveness of the computational method in comparison to manual evaluation (subject-based and expert-based techniques).</p> <p>Obj 7. To study the level of effectiveness of the eye-tracking method in comparison to manual (subject-based technique) evaluation.</p> <p>Obj 8. To develop and validate the framework for the comprehensibility evaluation of OSH signs.</p>	Collected comprehensibility data from chapters 2,3 and 4 and proved hypotheses H₁ , H₂ , and H₃ .	<ul style="list-style-type: none"> Chaudhuri, S.B., Majhi, M. and Karmakar, S. " Evaluation of Comprehensibility of a Sign by Triangulation Method", [Scopus Indexed, Compendex journal] In Chakrabarti, D., Karmakar, S., Salve, U.R. (eds) Ergonomics for Design and Innovation. HWWE 2021. Lecture Notes in Networks and Systems, vol 391. Springer, Cham. https://doi.org/10.1007/978-3-030-94277-9_154 Chaudhuri, S.B., Yadav, P., Majhi, M. and Karmakar, S. "Human-Centered Sign Design/ Redesign Life Cycle for Comprehension of Safety Sign". International Joint Conference- ACED 2023, HWWE 2023 & BRICSplus HFE 2023, Mumbai, SpringerNature Chaudhuri, S.B., Majhi, M. and Karmakar, S. "An Analytical Comparative Study between Two Techniques of Data Collection for Comprehensibility Evaluation of Safety Signs" [Communicated] Chaudhuri, S.B., Majhi, M. and Karmakar, S. "An Analytical Man-Machine-Hybrid Model for Comprehensibility Evaluation of Safety Messages Using Deep Learning Approach." [Communicated] Chaudhuri, S.B., Majhi, M. and Karmakar, S. "Exploration and analysis of significant eye-tracking measures, subjective measures, and their inter-relationship: A study of comprehensibility of OHS signs." [Communicated]

1.11.1. Brief of chapters

Chapter 1: Introduction

This chapter discusses the literature on techniques or methods used to study the comprehensibility of OSH signs. There is a discussion regarding the motivation of the study, theoretical constructs, and the conceptual framework behind it. The literature assists in highlighting the research gap, followed by research questions, the problem statement, the aim of the study, and the objectives. The hypotheses have been proposed. A short highlight of the expected outcome is presented. Finally, this chapter concludes with the framework of the thesis.

Chapter 2: Evaluation of Comprehensibility of OSH Signs by Subject and Expert Techniques

This chapter discusses the two techniques, namely the subject-based technique and expert-based technique, which are both evaluated by the experts. The experts are from the OSH department of the food manufacturing company and have certification from NEBOSH IGC and PGDHSE. They are specialized in evaluating and designing OSH signs besides other OSH activities. The methodology section discusses the stimuli, the participants, the evaluation strategy, and the research design in detail. The strategy for calculating comprehensibility is discussed in detail in this chapter. The results and discussion section is divided into two depending on the techniques discussed. The result and discussion section consists of descriptive analysis and a normality test of data distribution for each technique. Then various statistical tools and techniques are used to study the inter-rater consistency (for expert-based technique), variances among the signs, and whether the textual information affects the comprehensibility of signs. The comprehensibility percentage is calculated to categorize the sign as comprehensible or incomprehensible.

Chapter 3: Evaluation of Comprehensibility of OSH Signs by Computational Technique

This chapter discusses a technique that automatically evaluates the textual responses of the participants. The methodology section discusses the stimuli, the participants, the evaluation strategy, i.e., the score generation procedure, and the research design in detail. Then there is a detailed discussion regarding the tools, techniques, and algorithms used to develop the computational model. The result and discussion section discusses the output of the score generation procedure implemented here. The discussion is followed by the descriptive analysis of the data/ score generated and the statistical tests like the normality test, the variances among sign pairs, and whether the textual information affects the comprehensibility of signs. The comprehensibility percentage is calculated to categorize the sign as comprehensible or incomprehensible. Threshold tuning is introduced, making the developed model flexible enough to be implemented for any organization with varied severity, hazards, danger, context, and like. This technique overcomes the biases of the human-based evaluation technique. It removes ambiguous decisions and unwanted errors while experts score the textual descriptive responses of the participants, or the scores directly provided by the participants or subjects during the experiment.

Chapter 4: Evaluation of Comprehensibility of OSH Signs by Eye-tracking Technique

This chapter discusses a technique capable of extracting objective data from the human eye based on an eye-tracking experiment. This study is meant to minimize the subjective biases of humans (both in terms of data and evaluation) by finding significant features that are correlated with the subjective score of the participants. Hence, subjective subject-based and eye-tracking techniques have been exploited to extract the data. The stimuli for both the techniques, participants, and research design have been discussed in this chapter. Both the data undergo descriptive analysis followed by a test for their data distribution (normality test). Comprehensibility percentage has been calculated from the subjective data, and the signs are categorized into comprehensible and incomprehensible. Various statistical tests have been performed to see the correlation between the subjective and the objective variables. Another test has been performed to check whether a sign with textual information related to OSH affects the comprehensibility of the sign under study. Finally, the signs are then statistically analyzed for their variability across categories.

Chapter 5: Comparative study between the techniques- its merits and demerits


This chapter discusses the similarities and differences between the techniques implemented and narrated in working chapters 2, 3, and 4. Statistical analyses like correlation plots (Pearson plot) and bivariate analysis have been performed between each pair of techniques to study their correlation and variability. The result and discussion section discusses the correlation between techniques, and a mini-comparative study has been displayed considering different parameters (data type, data source, instruments, experimental set-up, software/analysis tool, and ease of handling experiment). Finally, chapter 5 ends with the validation results of all the techniques discussed throughout the working chapters (Chapters 2, 3, and 4, respectively) and a conclusion derived from the comparative study.

Chapter 6: Proof of concept

The evaluation framework generated the common comprehensible and incomprehensible categories from which one of the most incomprehensible signs has been selected for redesign. A human-centered design/ redesign cycle has been proposed and implemented here. A flexible approach for improvement at the last minute before the release has been introduced to release the best possible version of the variants. The redesign phase generated four variants, out of which one has been found comprehensible and finally selected for study. Two variants for a single sign have been designed and studied for better comprehensibility.

Chapter 7: Discussion: key findings, contribution of the thesis, limitations, and future scope

The final chapter discusses the overall thesis results. There is a detailed discussion of the key findings of the present research followed by the fulfillment of the objectives. Proof of hypotheses testing and the implications of the findings have been elaborated here. There is a detailed discussion of the key contributions followed by a narration of the recommendations based on the study. Ultimately, the thesis concludes by discussing its limitations, future scope, and conclusion from the research.



2

**Evaluation of
Comprehensibility of OSH
Signs by Subject-based and
Expert-based Techniques**

Chapter 2: Evaluation of Comprehensibility of OSH Signs by Subject-based and Expert-based Techniques

Abstract

A study of the techniques of comprehensibility of a sign (message) is essential to confirm that the OSH message is communicated unambiguously to the intended users. This chapter deals with two techniques of comprehensibility evaluation of OSH sign(s)- subject-based and expert-based. In both cases, 137 participants participated in the study with no record of any eye impairments. For the subject-based technique, the subjects rated the comprehensibility level of the signs. The subject-based technique has its disadvantages, and for this, another technique- the expert-based technique has been implemented, which complements the first technique. For this technique, two experts experienced in the OSH field have been considered for the study. The experts are from the OSH department of the food manufacturing company and have certification from NEBOSH IGC and PGDHSE. They are specialized in evaluating and designing OSH signs besides other OSH activities. Their inter-rater reliability is checked for consistency of the data, which is considered for further study. Both techniques are based on subjective data, with their advantages and disadvantages discussed. For both techniques, statistical tests like descriptive statistics and normality of data have been tested for data distribution. The ANOVA test has been conducted to study the variances between the sign pairs, followed by testing whether text affects the comprehensibility of a sign. The comprehensibility percentage has been calculated for each sign. In both techniques, signs conforming to ISO standards (67%) have been considered comprehensible; otherwise, it has been categorized as incomprehensible and recommended for a redesign. The expert-based technique gave consistent results in terms of comprehensibility as compared to the subject-based technique.

2.1. Introduction

There has always been an attempt to improve the OSH environment by using safety signs as a means of communication when other methods of controlling hazard(s) fail or are infeasible (Chaudhuri et al., 2022; Jiang et al., 2022; Lehto, 1992; Wogalter, 2018b). The purpose of the sign is to communicate the desired information unambiguously, which, however, always does not succeed in doing so (Chaudhuri et al., 2021; Reis et al., 2015). The failure to communicate leads to unintended health issues or accidents that are harmful to employees in an organization(s). The study of the comprehensibility or understandability of a sign is significant considering the issues mentioned earlier, where misinterpretation of any sign/symbol may result in detrimental consequences or acute health hazards. “*Comprehensibility refers to the ability of the individual reading a label, warning, or material; safety data sheet to understand the information sufficiently to take the desired action*” (Lin et al., 2021; Sattler et al., 1997). *Comprehensibility* measures how well the receiver of the information understood it. The warning that is well comprehensible by one group of employees may be poorly comprehensible by others (Sattler et al., 1997). Comprehensibility/ understandability predominates factors like legibility, learnability, etc. (Dewar, 1988). The sign intends to

communicate the message of the hazard appropriately. Besides, signs need to be noticeable, legible, memorable, believable, and motivating to facilitate comprehension and compliance behavior (Wogalter, 2018b). Comprehension can be ensured through a standard evaluation method that confirms the comprehensibility of the existing and designed sign(s).

Evaluation of comprehensibility of OSH sign(s) is usually conducted using the ISO method of comprehension test. The signs are shown to the participants/ subjects either online using a computer (screen test page) or offline (printed test page), maintaining ISO 9186 standards. The participants chose the responses from the categories of the responses, and they were finally evaluated and analyzed (frequency of each category and percentage of the total number of responses) by the two experts. Variations of ISO standards have been used in several studies, and it is found that various items have been used with varied scales to capture the responses either verbally or in writing (Arcia et al., 2019; Berrio et al., 2022; Morina & Ahma, 2022). Even after such standardization of the evaluation procedure, studies reveal that many signs developed do not conform with the ISO/ANSI standards (Arcia et al., 2019; Montagne, 2013), and they are not comprehensible or fail to convey the intended message to the target audience (Bañares et al., 2018; Davies et al., 1998; Duarte & Rebelo, 2005; Handcock et al., 2004; L. Liu et al., 2005; Tam et al., 2003). The signs thought to be highly comprehensible are hardly effective in a real scenario. Thus, there need to be variations in the evaluation procedure of ISO standards which might help capture the subjects' minds directly as well as their thought process about the scores they might provide during evaluation. Hence, a study has been conducted that captures both the ratings and descriptive responses of the subjects.

2.2. Methodology

The detailed methodology for the two techniques is discussed in the following sub-sections.

2.2.1. Stimuli

The study's test material (stimuli) consisted of fourteen standard OSH-related signs, consisting of three prohibition sign pairs (with and without supplementary text) and four mandatory sign pairs (with and without supplementary text) (Figure 2. 1) have been used as test material (stimuli) for the experiment. Each image is colored, and the approximate size is 42 mm × 42 mm. The viewing distance is between 400 and 700 mm (Caffaro & Cavallo, 2015; Shieh & Lee, 2007). Information related to the context plays a vital role in comprehending a pictogram (Lehto, 1992), and it seems that comprehension improves when contextual information is provided (Galer, 1980; Vukelich & Whitaker, 1993). The context is described in words in oral and written format. A separate instruction page has been provided to the participants describing the purpose of the study. A demonstration of the experimental procedures has been recorded (audio-video recording for 4 mins and 58 secs) in Microsoft PowerPoint, considering the stimuli different from the proposed experiment. The recording has been given to each participant, who is instructed to view them before experimenting.

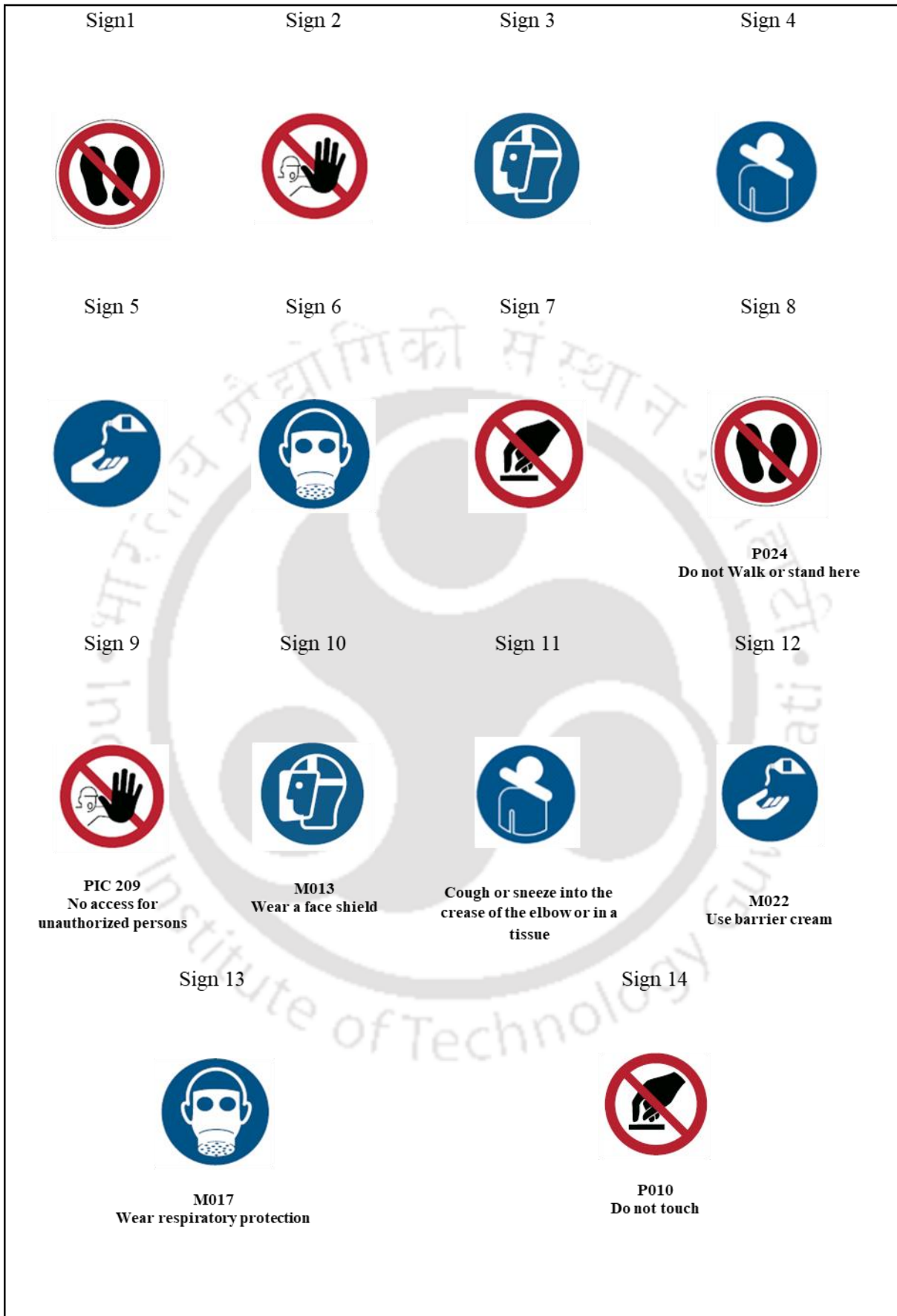


Figure 2. 1. Stimuli

2.2.2. Participants

One hundred and thirty-seven (N=137; 111 males and 26 females) employees out of four hundred (communicated through e-mails) volunteered for the study (from 15 different states of India) with no history of color deficiency and severe eye conditions and or visual impairment (self-reported). According to the International Organization for Standardization (ISO), the sample size must be 50 (ISO, 2014b, 2014a) for each symbol's variants in each country. However, in this study, Cochran's formula (Cochran, 1977; Schillaci & Schillaci, 2022) has been used to determine the sample size, which is 133 in our study.

All participants are healthy and white-collar employees of a food manufacturing company with a mean age of 33.85 and a standard deviation of 7.2. Informed consent is obtained from all the participants before the beginning of the trial. Before scribing their responses, information is provided to view each sign for a minimum duration of 160 to 200 milliseconds (Caplette et al., 2020). The experiment has been conducted following ethical guidelines and principles.

2.2.3. Research Instrument and Research Design

The designed questionnaire has been segregated into two parts, and they are online. The first part asked for the respondents' demographic data like age, place, educational background, gender, and work experience. The second part of the questionnaire helped collect data for two techniques. Each sign (stimuli) is placed on top of every page, followed by questions. Enough space has been provided for the responses.

In the *subject-based technique*, a question has been framed for rating participants' comprehension level of the sign. An eleven-point rating scale is used for collecting the response to the statement "Assign a score for your understanding of 'sign 1' that it intends to convey. Note: 0.0 is the minimum value, and 1.0 is the maximum value. (0.0 signifies no understanding, 0.1 signifies slightly better understanding than 0.0, ..., 0.5 signifies medium understanding, 1.0 signifies complete understanding.)". The rate/ score collected from the participants has been directly used for comprehensibility frequency calculation. The protocol for scoring/ rating the questionnaire follows the protocol of the expert technique as discussed below, with a slight difference in that the participants are not aware of the correct keywords or standard answers for the signs under study. The participants must guess and assign the score based on their understanding of the sign.

In the *expert-based technique*, the participants were asked to write the descriptive answer to the question, "What message does the sign convey?". This part of the question is subjective and open-ended; two experts evaluated the responses. The experts were trained and experienced in conducting behavioral experiments with human subjects and analyzing their responses (ISO, 2008). The experts are from the EHS department, Britannia Industries Ltd, a food manufacturing company in India. They are PGDHSE and have certification from NEBOSH IGC. The experts held a meeting before the experiment and decided upon a protocol (data coding/ binning threshold) while correcting descriptive answers. They are:

- i. The number of keywords written must be accounted for (leaving grammar like an/ the/ a/ into/in/at)
- ii. Let us consider an example of an OSH message “**Cough** or **sneeze** at the **crease** of the **elbow** or in a **tissue**.” There are five keywords as highlighted.
- iii. If ‘n’ is the number of keywords for the standard answer provided and the ‘n’ keywords are present in the response, then a score of ‘1.0’ can be assigned. So in this example, if the response is “Cough or sneeze at the crease of the elbow or in a tissue,” 5 out of 5 (5/5) keywords are present; hence a score of 1.0 can be assigned.
- iv. If the response contains greater than or equal to ‘n/2’ keywords when n is even or ‘(n+1)/2’ keywords when n is odd, then a score of 0.5 to 0.9 can be assigned.
- v. If the response contains less than ‘n/2’ keywords where n is even or ‘(n+1)/2’ keywords when n is odd, then a score of 0.0 to 0.4 can be assigned, and this range of values falls in the incomprehensible category, which is less than 0.5 (50%). However, 0.4 is better than 0.0, and this difference is not reflected in the final category of results in the present study. This study can be extended further with more categories besides the two categories-comprehensible/ incomprehensible.
- vi. If the keywords of the response do not match the standard response, however, if the concept almost matches, then (0.5) can be assigned, and the message (sign) can be considered comprehensible.

This whole procedure of transforming data from qualitative to quantitative binary categories (Gilula et al., 2019; Zaini et al., 2022) introduced a novel data coding technique for binning the comprehensibility scores. Finally, the mean score of the two experts was collected as the representative score (Han et al., 2011). Statistical tests have been conducted to assess interrater consistency/ reliability using Cronbach’s alpha method. The average value of the experts is being considered for the study. The bivariate correlation analysis between the data collected for the two techniques (Table 5. 2 and Table 5. 3) validates the scales (criterion-related and construct validity- correlation analysis) used for the study (Boateng et al., 2018).

Both techniques considered analyzing descriptive statistics and the normality of data (using Kolmogorov-Smirnov and Shapiro-Wilk) for its distribution before conducting any statistical tests and analysis. The internal consistency of the items, i.e., the reliability of the two items, is tested considering bivariate correlation (Pearson’s r, Kendall's tau_b, and Spearman’s rho). Bivariate correlation, an alternate form of reliability testing, is considered here as the items’ responses are of the same construct but different forms of the same test (Anderson et al., 2012). Analysis of variance (ANOVA) has been conducted for all the signs. Friedman’s ANOVA test, followed by post hoc analysis, has been done to study variances between all the signs viewed by the same group of participants. The non-parametric Friedman’s ANOVA considers the actual data, sorts the data in ranks, compares the median, and calculates the p-value. Friedman calculates overall statistical differences (p values) for all groups together and not as pairs. It will not specify which two groups are significant or not. Another test conducted following the Friedman test is the post hoc test. The post hoc test helps decide which two sign pairs are significantly different from each other.

Wilcoxon signed-rank test has been performed to check the effect of text on signs. Then, the effect size confirms the result by calculating the Z value using the following equation (2.1)

$$r = \frac{Z}{\sqrt{N}} \quad (2.1),$$

where Z is the z-score of the Wilcoxon test, and N is the size of the study (number of total observations). The effect size (Z score) is calculated for all seven (7) pairs of signs (with and without text).

‘Comprehensibility frequency’ in this study measures the total number of respondents who correctly comprehended the sign. The comprehensibility percentage (C_p) is calculated by adding the total number of participants who understood the sign (Comprehensibility frequency), dividing by the total number of participants who participated in the experiment, and multiplying by 100. The C_p equal to or exceeding 67% is considered comprehensible, and the remaining is incomprehensible. All the statistical analyses have been performed using IBM SPSS Statistics (Version 25.0). The comprehensibility evaluation model for the subject and expert-based techniques can be seen in Figure 2. 2 and Figure 2. 3, respectively.

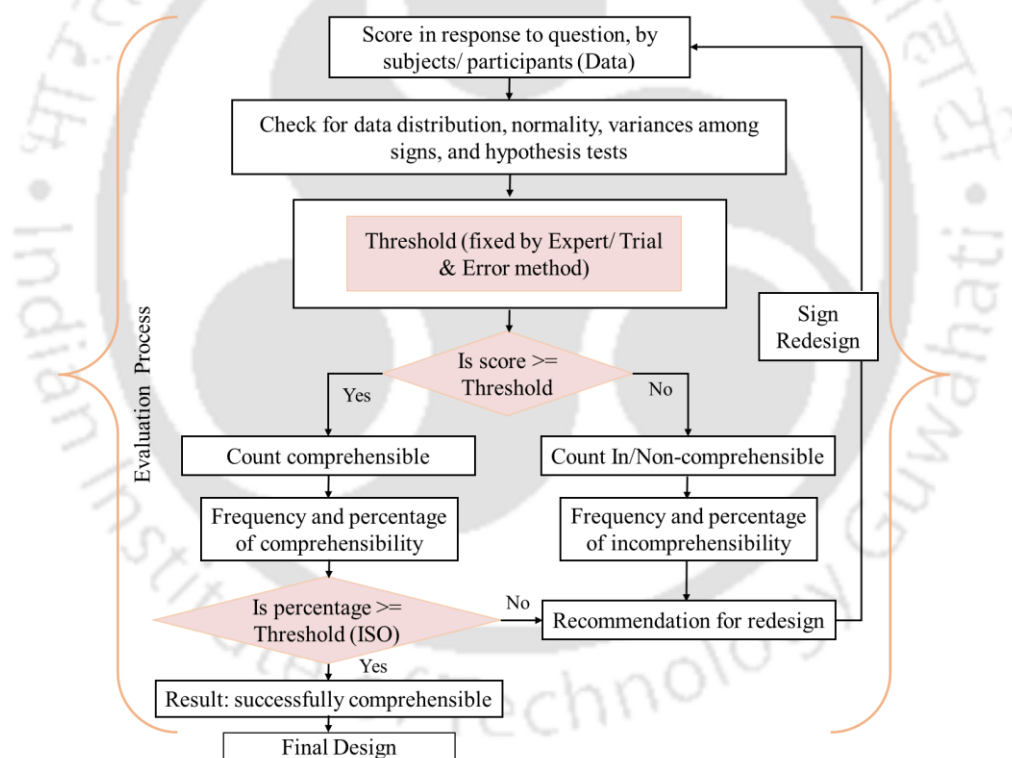


Figure 2. 2 . Comprehensibility evaluation model of OSH sign (subject-based technique)

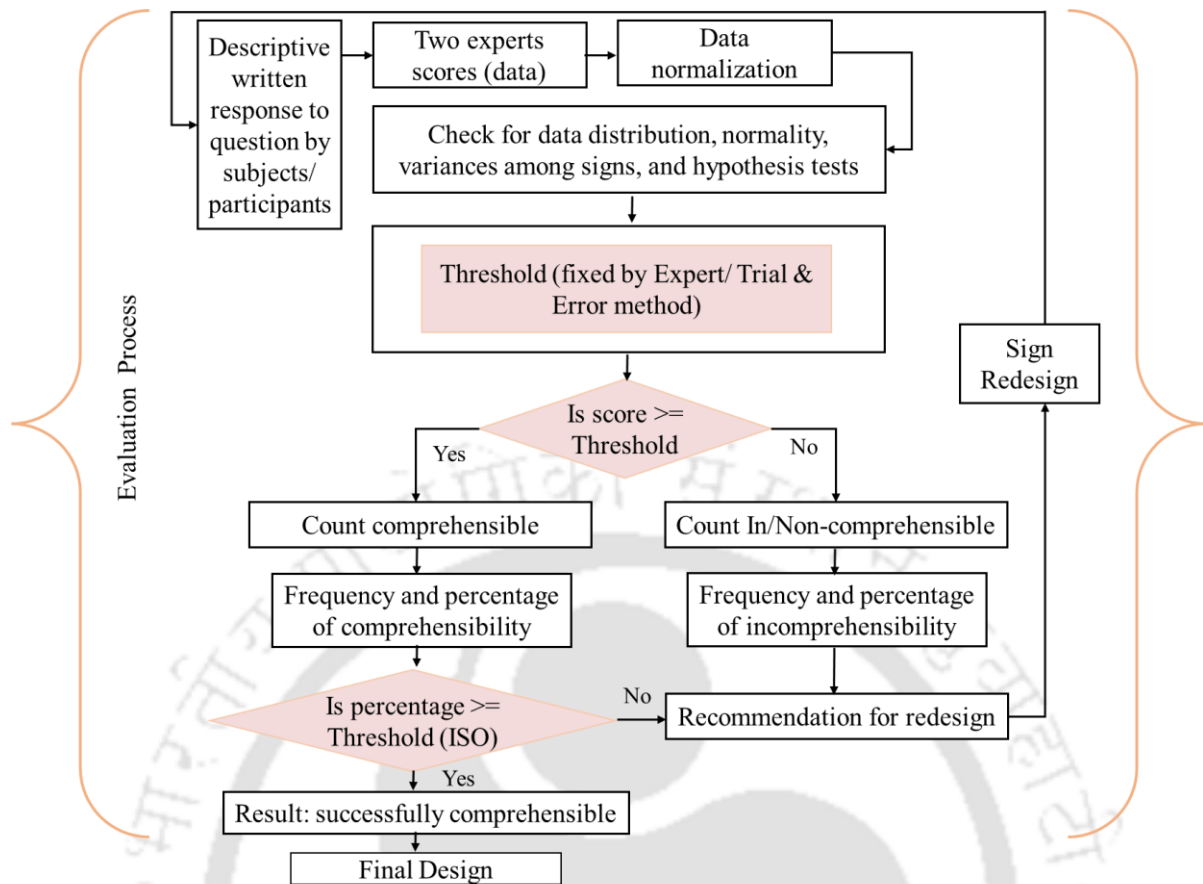


Figure 2. 3. Comprehensibility evaluation model of OSH sign (expert-based technique)

2.3. Results and Discussions

The results of the two techniques, subject-based and expert-based, and their various statistical analyses are discussed below.

2.3.1. Subject-based technique

The subject-based technique collects the comprehensibility scores directly from the subjects. The direct minds of the subjects and their comprehension level can reflect through their scores if proper training can be given for scoring procedures. Experts can use the scores directly for further studies.

2.3.1.1. Descriptive Statistics

Descriptives have been calculated using IBM SPSS 25.0 with 137 cases with no missing data (Table 2. 1, Table 2. 2).

Table 2. 1. Descriptive statistics of subject-based technique (signs without text)

		NAFUP		UBC		DNT		DNWSH		COSACOEOT		WRP		WAFS	
		Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
Mean		.6515	.02677	.5934	.02830	.5779	.02914	.6743	.02820	.2588	.02731	.4022	.03031	.7426	.02381
95% Confidence Interval for Mean	L.B.	.5985		.5374		.5203		.6185		.2048		.3423		.6955	
	U.B.	.7044		.6494		.6356		.7300		.3128		.4622		.7897	
5% Trimmed Mean		.6683		.6038		.5866		.6936		.2327		.3913		.7690	
Median		.8000		.5000		.6000		.8000		.1000		.3000		.8000	
Variance		.097		.109		.116		.108		.101		.125		.077	
Std. Deviation		.31222		.33003		.33987		.32886		.31844		.35349		.27772	
Minimum		.00		.00		.00		.00		.00		.00		.00	
Maximum		1.00		1.00		1.00		1.00		1.00		1.00		1.00	
Range		1.00		1.00		1.00		1.00		1.00		1.00		1.00	
Interquartile Range		.50		.60		.50		.60		.50		.60		.50	
Skewness		-.621	.208	-.118	.208	-.410	.208	-.777	.208	.977	.208	.493	.208	-1.049	.208
Kurtosis		-.805	.413	-1.244	.413	-1.036	.413	-.704	.413	-.335	.413	-1.088	.413	.473	.413

Here L.B.: Lower Bound; U.B.: Upper Bound

Table 2. 2. Descriptive statistics of subject-based technique (signs with text)

		DNWSH_T		NAFUP_T		WAFS_T		COSACOEOT_T		UBC_T		WRP_T		DNT_T	
		Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
Mean		.7706	.02663	.6228	.02954	.8691	.01785	.4426	.03130	.6037	.02945	.5787	.03097	.7007	.02821
95% Confidence Interval for Mean	L.B.	.7179		.5644		.8338		.3807		.5454		.5174		.6449	
	U.B.	.8233		.6812		.9044		.5046		.6619		.6399		.7565	
5% Trimmed Mean		.8007		.6364		.8977		.4363		.6152		.5874		.7230	
Median		.9000		.7000		1.0000		.4000		.6500		.6000		.8000	
Variance		.096		.119		.043		.133		.118		.130		.108	
Std. Deviation		.31058		.34447		.20817		.36507		.34349		.36115		.32897	
Minimum		.00		.00		.00		.00		.00		.00		.00	
Maximum		1.00		1.00		1.00		1.00		1.00		1.00		1.00	
Range		1.00		1.00		1.00		1.00		1.00		1.00		1.00	
Interquartile Range		.40		.60		.20		.70		.58		.68		.50	
Skewness		-1.261	.208	-.511	.208	-2.000	.208	.260	.208	-.432	.208	-.313	.208	-.880	.208
Kurtosis		.337	.413	-1.084	.413	3.953	.413	-1.404	.413	-1.053	.413	-1.335	.413	-.577	.413
Here L.B.: Lower Bound; U.B.: Upper Bound															

2.3.1.2. Test for Normality

Normality tests (Kolmogorov-Smirnov and Shapiro-Wilk in Table 2. 3) reveal that the data, i.e., comprehensibility scores of all the 14 signs, deviates from the normal distribution (p-value less than alpha (0.05)). The test fails to accept that sample data are normally distributed (Shapiro & Wilk, 1965). Getting truly, normally distributed data in the real world is hard. However, the normality of data does not cause significant issues with a sample size greater than 25 or 30 (Hogg et al., 1977).

Table 2. 3. Test for normality (subject-based technique)

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
NAFUP	.198	136	.000	.892	136	.000
UBC	.163	136	.000	.891	136	.000
DNT	.140	136	.000	.897	136	.000
DNWSH	.200	136	.000	.854	136	.000
COSACOEOT	.248	136	.000	.791	136	.000
WRP	.143	136	.000	.878	136	.000
WAFS	.177	136	.000	.838	136	.000
DNWSH_T	.248	136	.000	.749	136	.000
NAFUP_T	.152	136	.000	.877	136	.000
WAFS_T	.294	136	.000	.686	136	.000
COSACOEOT_T	.159	136	.000	.882	136	.000
UBC_T	.129	136	.000	.891	136	.000
WRP_T	.156	136	.000	.880	136	.000
DNT_T	.214	136	.000	.827	136	.000

a. Lilliefors Significance Correction

2.3.1.3. Analysis of variance (ANOVA)

Analysis of variance (ANOVA) has been conducted for all the signs. Here “sub” means subject based. The result of Friedman’s ANOVA test (Table 2. 4 and Table 2. 5), followed by post hoc analysis, shows differences in comprehensibility levels between the signs (all sign pairs) viewed by the same group of participants. Friedman’s test calculated the p-value (Table 2. 5), which is less than 0.05 and is significant here. Hence the null hypothesis, “there are no significant differences between the signs,” can be rejected. The result hypothesizes the significant differences in comprehensibility levels among all the signs considered for the study.

Table 2. 4. Rank of the signs generated by Friedman's ANOVA (subject-based technique)

Sign Name	Ranks	
	Abbreviated Sign_name (used for SPSS Analysis)	Mean Rank
No Access for Unauthorized Persons (without text)	NAFUP_sub	7.75
Use Barrier Cream (without text)	UBC_sub	7.13
Do Not Touch (without text)	DNT_sub	6.90
Do Not Walk or stand here (without text)	DNWSH_sub	8.24
Cough or Sneeze into the crease of the elbow or in a tissue (without text)	COSACOEOT_sub	3.63
Wear Respiratory Protection (without text)	WRP_sub	4.96
Wear a face shield (without text)	WAFS_sub	9.04
Do Not Walk or stand here (with text)	DNWSH_T_sub	9.66
No Access for Unauthorized Persons (with text)	NAFUP_T_sub	7.63
Wear a face shield (with text)	WAFS_T_sub	11.03
Cough or Sneeze into the crease of the elbow or in a tissue (with text)	COSACOEOT_T_s ub	5.76
Use Barrier Cream (with text)	UBC_T_sub	7.47
Wear Respiratory Protection (with text)	WRP_T_sub	7.09
Do Not Touch (with text)	DNT_T_sub	8.71

Table 2. 5. Test statistics of Friedman's ANOVA (subject-based technique)

Test Statistics ^(a)	
N	136
Chi-Square	409.895
df	13
Asymp. Sig.	.000

a. Friedman Test

Posthoc analysis (Figure 2. 4) helps to identify significant pairs. The result (Figure 2.5 – Figure 2.10) shows that forty-nine (49) pairs are significantly different, and the pairs are highlighted in yellow (p-values are the same as the Adj. Sig. value in the figure) in the figures for brevity.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distributions of NAFUP_sub, UBC_sub, DNT_sub, DNWSH_sub, COSACOEOT_sub, WRP_sub, WAFS_sub, DNWSH_T_sub, NAFUP_T_sub, WAFS_T_sub, COSACOEOT_T_sub, UBC_T_sub, WRP_T_sub and DNT_T_sub are the same.	Related-Samples Friedman's Two-Way Analysis of Variance by Ranks	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 2. 4. Post hoc analysis (subject-based technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
COSACOEOT_sub-WRP_sub	-1.338	.507	-2.638	.008	.759
COSACOEOT_sub-COSACOEOT_T_sub	-2.140	.507	-4.218	.000	.002
COSACOEOT_sub-DNT_sub	3.272	.507	6.450	.000	.000
COSACOEOT_sub-WRP_T_sub	-3.463	.507	-6.827	.000	.000
COSACOEOT_sub-UBC_sub	3.504	.507	6.907	.000	.000
COSACOEOT_sub-UBC_T_sub	-3.849	.507	-7.588	.000	.000
COSACOEOT_sub-NAFUP_T_sub	-4.000	.507	-7.885	.000	.000
COSACOEOT_sub-NAFUP_sub	4.121	.507	8.124	.000	.000
COSACOEOT_sub-DNWSH_sub	4.610	.507	9.088	.000	.000
COSACOEOT_sub-DNT_T_sub	-5.088	.507	-10.030	.000	.000
COSACOEOT_sub-WAFS_sub	-5.419	.507	-10.682	.000	.000
COSACOEOT_sub-DNWSH_T_sub	-6.037	.507	-11.900	.000	.000
COSACOEOT_sub-WAFS_T_sub	-7.408	.507	-14.603	.000	.000
WRP_sub-COSACOEOT_T_sub	-.801	.507	-1.580	.114	1.000

Figure 2. 5. Post hoc analysis result showing significantly different pairs in sequence-1 (subject-based technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
WRP_sub-DNT_sub	1.934	.507	3.812	.000	.013
WRP_sub-WRP_T_sub	-2.125	.507	-4.189	.000	.003
WRP_sub-UBC_sub	2.165	.507	4.269	.000	.002
WRP_sub-UBC_T_sub	-2.511	.507	-4.950	.000	.000
WRP_sub-NAFUP_T_sub	-2.662	.507	-5.247	.000	.000
WRP_sub-NAFUP_sub	2.783	.507	5.486	.000	.000
WRP_sub-DNWSH_sub	3.272	.507	6.450	.000	.000
WRP_sub-DNT_T_sub	-3.750	.507	-7.392	.000	.000
WRP_sub-WAFS_sub	-4.081	.507	-8.044	.000	.000
WRP_sub-DNWSH_T_sub	-4.699	.507	-9.262	.000	.000
WRP_sub-WAFS_T_sub	-6.070	.507	-11.965	.000	.000
COSACOEOT_T_sub-DNT_sub	1.132	.507	2.232	.026	1.000
COSACOEOT_T_sub-WRP_T_sub	-1.324	.507	-2.609	.009	.826
COSACOEOT_T_sub-UBC_sub	1.364	.507	2.689	.007	.653
COSACOEOT_T_sub-UBC_T_sub	-1.710	.507	-3.370	.001	.068
COSACOEOT_T_sub-NAFUP_T_sub	1.860	.507	3.667	.000	.022
COSACOEOT_T_sub-NAFUP_sub	1.982	.507	3.906	.000	.009

Figure 2. 6. Post hoc analysis result showing significantly different pairs in sequence-2 (subject-based technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
COSACOEOT_T_sub-DNWSH_sub	2.471	.507	4.870	.000	.000
COSACOEOT_T_sub-DNT_T_sub	-2.949	.507	-5.812	.000	.000
COSACOEOT_T_sub-WAFS_sub	3.279	.507	6.464	.000	.000
COSACOEOT_T_sub-DNWSH_T_sub	3.897	.507	7.682	.000	.000
COSACOEOT_T_sub-WAFS_T_sub	5.268	.507	10.385	.000	.000
DNT_sub-WRP_T_sub	-.191	.507	-.377	.706	1.000
DNT_sub-UBC_sub	.232	.507	.457	.648	1.000
DNT_sub-UBC_T_sub	-.577	.507	-1.138	.255	1.000
DNT_sub-NAFUP_T_sub	-.728	.507	-1.435	.151	1.000
DNT_sub-NAFUP_sub	.849	.507	1.674	.094	1.000
DNT_sub-DNWSH_sub	-1.338	.507	-2.638	.008	.759
DNT_sub-DNT_T_sub	-1.816	.507	-3.580	.000	.031
DNT_sub-WAFS_sub	-2.147	.507	-4.232	.000	.002
DNT_sub-DNWSH_T_sub	-2.765	.507	-5.450	.000	.000
DNT_sub-WAFS_T_sub	-4.136	.507	-8.153	.000	.000
WRP_T_sub-UBC_sub	.040	.507	.080	.936	1.000
WRP_T_sub-UBC_T_sub	.386	.507	.761	.447	1.000

Figure 2. 7. Post hoc analysis result showing significantly different pairs in sequence-3 (subject-based technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
WRP_T_sub-NAFUP_T_sub	.537	.507	1.058	.290	1.000
WRP_T_sub-NAFUP_sub	.658	.507	1.297	.195	1.000
WRP_T_sub-DNWSH_sub	1.147	.507	2.261	.024	1.000
WRP_T_sub-DNT_T_sub	-1.625	.507	-3.203	.001	.124
WRP_T_sub-WAFS_sub	1.956	.507	3.855	.000	.011
WRP_T_sub-DNWSH_T_sub	2.574	.507	5.073	.000	.000
WRP_T_sub-WAFS_T_sub	3.945	.507	7.776	.000	.000
UBC_sub-UBC_T_sub	-.346	.507	-.681	.496	1.000
UBC_sub-NAFUP_T_sub	-.496	.507	-.978	.328	1.000
UBC_sub-NAFUP_sub	.618	.507	1.218	.223	1.000
UBC_sub-DNWSH_sub	-1.107	.507	-2.181	.029	1.000
UBC_sub-DNT_T_sub	-1.585	.507	-3.124	.002	.163
UBC_sub-WAFS_sub	-1.915	.507	-3.776	.000	.015
UBC_sub-DNWSH_T_sub	-2.533	.507	-4.993	.000	.000
UBC_sub-WAFS_T_sub	-3.904	.507	-7.696	.000	.000
UBC_T_sub-NAFUP_T_sub	.151	.507	.297	.766	1.000
UBC_T_sub-NAFUP_sub	.272	.507	.536	.592	1.000

Figure 2. 8. Post hoc analysis result showing significantly different pairs in sequence-4 (subject-based technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
UBC_T_sub-DNWSH_sub	.761	.507	1.500	.134	1.000
UBC_T_sub-DNT_T_sub	-1.239	.507	-2.442	.015	1.000
UBC_T_sub-WAFS_sub	1.570	.507	3.095	.002	.179
UBC_T_sub-DNWSH_T_sub	2.188	.507	4.312	.000	.001
UBC_T_sub-WAFS_T_sub	3.559	.507	7.015	.000	.000
NAFUP_T_sub-NAFUP_sub	.121	.507	.239	.811	1.000
NAFUP_T_sub-DNWSH_sub	.610	.507	1.203	.229	1.000
NAFUP_T_sub-DNT_T_sub	-1.088	.507	-2.145	.032	1.000
NAFUP_T_sub-WAFS_sub	1.419	.507	2.797	.005	.469
NAFUP_T_sub-DNWSH_T_sub	2.037	.507	4.015	.000	.005
NAFUP_T_sub-WAFS_T_sub	-3.408	.507	-6.718	.000	.000
NAFUP_sub-DNWSH_sub	-.489	.507	-.964	.335	1.000
NAFUP_sub-DNT_T_sub	-.967	.507	-1.906	.057	1.000
NAFUP_sub-WAFS_sub	-1.298	.507	-2.558	.011	.957
NAFUP_sub-DNWSH_T_sub	-1.915	.507	-3.776	.000	.015
NAFUP_sub-WAFS_T_sub	-3.287	.507	-6.479	.000	.000
DNWSH_sub-DNT_T_sub	-.478	.507	-.942	.346	1.000

Figure 2. 9. Post hoc analysis result showing significantly different pairs in sequence-5 (subject-based technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
DNWSH_sub-WAFS_sub	-.809	.507	-1.594	.111	1.000
DNWSH_sub-DNWSH_T_sub	-1.426	.507	-2.812	.005	.448
DNWSH_sub-WAFS_T_sub	-2.798	.507	-5.515	.000	.000
DNT_T_sub-WAFS_sub	.331	.507	.652	.514	1.000
DNT_T_sub-DNWSH_T_sub	.949	.507	1.870	.062	1.000
DNT_T_sub-WAFS_T_sub	2.320	.507	4.573	.000	.000
WAFS_sub-DNWSH_T_sub	-.618	.507	-1.218	.223	1.000
WAFS_sub-WAFS_T_sub	-1.989	.507	-3.921	.000	.008
DNWSH_T_sub-WAFS_T_sub	-1.371	.507	-2.703	.007	.625

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .05. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Figure 2. 10. Post hoc analysis result showing significantly different pairs in sequence-6 (subject-based technique)

2.3.1.4. OSH sign with OSH-related textual information affects the comprehensibility of the messages intended to convey (subject-based technique)

Considering the hypothesis,

Null Hypothesis (H₀): The level of comprehensibility of the OSH sign is not significantly different when the sign is presented with or without textual information.

Alternate Hypothesis (H₃): The level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information.

To test the hypothesis, a Wilcoxon signed rank test has been conducted, and the result is shown in the following tables. Descriptive statistics (Table 2. 6) shows the mean of all the signs, indicating the significant difference between the signs with and without textual information. The mean of UBC_T is more than UBC. The case is similar for DNT_T and DNT, DNWSH_T and DNWSH, COSACOEOT_T and COSACOEOT, WRP_T and WRP, WAFS_T and WAFS. The only exception could be found in the case of NAFUP_T and

NAFUP, where the mean of the scores of NAFUP_T is slightly less than that of NAFUP; however, the difference is negligible, and further analysis reveals the actual result.

In general, it is seen that the mean value of the signs with text is more than the mean value of signs without text. This difference is significant and shows that OSH sign with OSH-related textual information affects the understanding or comprehensibility of the messages intended to be conveyed. The minimum and maximum values signify the minimum and maximum scores assigned for that sign under study.

Table 2. 6. Descriptive statistics for the subject-based comprehensibility score in the Wilcoxon Signed Rank test

Descriptive Statistics					
Sign name	N	Mean	Std. Deviation	Minimum	Maximum
NAFUP_sub	137	.6496	.31181	.00	1.00
UBC_sub	137	.5956	.32985	.00	1.00
DNT_sub	137	.5737	.34220	.00	1.00
DNWSH_sub	137	.6766	.32883	.00	1.00
COSACOEOT_sub	137	.2584	.31731	.00	1.00
WRP_sub	137	.4000	.35314	.00	1.00
WAFS_sub	137	.7445	.27757	.00	1.00
NAFUP_T_sub	137	.6255	.34471	.00	1.00
UBC_T_sub	137	.6066	.34390	.00	1.00
DNT_T_sub	136	.7007	.32897	.00	1.00
DNWSH_T_sub	137	.7723	.31006	.00	1.00
COSACOEOT_T_sub	137	.4401	.36491	.00	1.00
WRP_T_sub	137	.5818	.36162	.00	1.00
WAFS_T_sub	137	.8701	.20771	.00	1.00

Note: i. _sub denotes the subject-based technique.

ii. _T denotes textual information.

The hypothesis has been confirmed further using ranks. It is found that the mean of the positive rank and the sum of the positive ranks (Table 2. 7) are greater for the signs with textual information than the signs without textual information, except for the case of NAFUP_T and NAFUP.

Table 2. 7. Wilcoxon Signed Rank test ranks for the sign pairs: with and without text (subject-based technique)

		N	Mean Rank	Sum of Ranks
NAFUP_T_sub - NAFUP_sub	Negative Ranks	49 ^a	56.52	2769.50
	Positive Ranks	53 ^b	46.86	2483.50
	Ties	35 ^c		
	Total	137		
UBC_T_sub - UBC_sub	Negative Ranks	48 ^d	51.17	2456.00
	Positive Ranks	54 ^e	51.80	2797.00
	Ties	35 ^f		
	Total	137		
DNT_T_sub - DNT_sub	Negative Ranks	25 ^g	48.92	1223.00
	Positive Ranks	73 ^h	49.70	3628.00
	Ties	38 ⁱ		
	Total	136		
DNWSH_T_sub - DNWSH_sub	Negative Ranks	29 ^j	45.93	1332.00
	Positive Ranks	62 ^k	46.03	2854.00
	Ties	46 ^l		
	Total	137		
COSACOEOT_T_sub - COSACOEOT_sub	Negative Ranks	25 ^m	45.38	1134.50
	Positive Ranks	75 ⁿ	52.21	3915.50
	Ties	37 ^o		
	Total	137		
WRP_T_sub - WRP_sub	Negative Ranks	30 ^p	44.47	1334.00
	Positive Ranks	77 ^q	57.71	4444.00
	Ties	30 ^r		
	Total	137		
WAFS_T_sub - WAFS_sub	Negative Ranks	12 ^s	36.71	440.50
	Positive Ranks	64 ^t	38.84	2485.50
	Ties	61 ^u		
	Total	137		

- a. NAFUP_T_sub < NAFUP_sub
- b. NAFUP_T_sub > NAFUP_sub
- c. NAFUP_T_sub = NAFUP_sub
- d. UBC_T_sub < UBC_sub
- e. UBC_T_sub > UBC_sub
- f. UBC_T_sub = UBC_sub
- g. DNT_T_sub < DNT_sub
- h. DNT_T_sub > DNT_sub
- i. DNT_T_sub = DNT_sub

- j. DNWSH_T_sub < DNWSH_sub
- k. DNWSH_T_sub > DNWSH_sub
- l. DNWSH_T_sub = DNWSH_sub
- m. COSACOEOT_T_sub < COSACOEOT_sub
- n. COSACOEOT_T_sub > COSACOEOT_sub
- o. COSACOEOT_T_sub = COSACOEOT_sub

- p. WRP_T_sub < WRP_sub
- q. WRP_T_sub > WRP_sub
- r. WRP_T_sub = WRP_sub
- s. WAFS_T_sub < WAFS_sub
- t. WAFS_T_sub > WAFS_sub
- u. WAFS_T_sub = WAFS_sub

It is seen that the number of positive ranks for the NAFUP_T (NAFUP_T_sub > NAFUP_sub) is found to be 53, i.e., in 53 cases, the participants gave more scores to the NAFUP_T than the NAFUP, though the mean of the scores (descriptive statistics, Table 2. 6), mean rank and sum of the rank (Table 2. 7) is slightly more in case of NAFUP. The reason for such variation might be due to a greater number of similar kinds of comprehensibility scores in the case of NAFUP_T.

Again, the participants are novices in the scoring/ rating procedure, and it was found that they rated themselves more when they did not understand the sign. Similarly, they rated the signs less when they appropriately understood the sign. This reason sounds logical and is responsible for such variation in the case of the NAFUP and NAFUP_T pair. Table 2. 7 also shows 35 cases when the NAFUP_T and NAFUP have been ranked similarly (tie). The Z value (Table 2. 8) helped to calculate the effect size (r-value) (Table 2. 9), which assists in checking how strong is the effect of the text on the comprehensibility of the sign(s). The effect size is significant for all the pairs, which is excellent for our study.

Table 2. 8. Test statistics for Wilcoxon Signed Rank test for the sign pairs: with and without text (subject-based technique)

Test Statistics ^(a)							
	NAFUP_T_sub - NAFUP_sub	UBC_T_sub - UBC_sub	DNT_T_sub - DNT_sub	DNWSH_T_sub - DNWSH_sub	COSACOEOT_T_sub - COSACOEOT_sub	WRP_T_sub - WRP_sub	WAFS_T_sub - WAFS_sub
Z	-.479 ^b	-.571 ^c	-4.275 ^c	-3.022 ^c	-4.791 ^c	-4.850 ^c	-5.322 ^c
Asymp. Sig. (2-tailed)	.632	.568	.000	.003	.000	.000	.000

a. Wilcoxon Signed Ranks Test, b. Based on positive ranks., c. Based on negative ranks.

Table 2. 9. Effect Size for Wilcoxon Signed Rank test for the sign pairs: with and without text (subject-based technique)

	COSACOEOT_T_score - COSACOEOT_score	DNT_T_score - DNT_score	DNWSH_T_score - DNWSH_score	NAFUP_T_score - NAFUP_score	UBC_T_score - UBC_score	WAFS_T_score - WAFS_score	WRP_T_score - WRP_score
Z	-4.791	-4.275	-3.022	-0.479	-0.571	-5.322	-4.85
N	274	274	274	274	274	274	274
Root N	16.55294536	16.55294536	16.55294536	16.55294536	16.55294536	16.55294536	16.55294536
r	-0.289434895	-0.258262195	-0.182565697	-0.028937448	-0.034495372	-0.321513778	-0.292999215

Hence the interpretation of the result for all the sign pairs can be as follows:

- i. A Wilcoxon signed rank test, and the calculated effect size reveal that for all the five sign pairs, it is found that textual information positively affects the comprehensibility of message/ sign evident from their mean difference, z value, p-value, and r value (Table 2. 6 - Table 2. 9) with a minor variation for NAFUP-NAFUP_T, and UBC- UBC_T.

- ii. Hence null hypothesis can be rejected, and the alternate hypothesis has been established that the level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information.

2.3.1.5. Comprehensibility percentage based on subject-based technique

The number of comprehensible signs is added, divided by the total participants, and multiplied by 100 to get the total *Percentage of comprehensibility*. The sign whose total comprehensibility percentage is more than or equal to 67% (Berrio et al., 2022; ISO 3864-1, 2011) is considered acceptable. This means that a minimum of 67% target population found the signs comprehensible.

Besides the ANOVA and hypothesis test (sections 2.3.1.3. and 2.3.1.4.), this comprehensibility percentage reveals that signs with text are more comprehensible than signs without text. As found in the above sections (sections 2.3.1.3. and 2.3.1.4.), this calculation also found that NAFUP is more comprehensible than NAFUP_T because all the above calculations have been done based on these scores. Detailed analysis reveals that the frequency of comprehensible signs is 99 in the case of NAFUP_T and 101 in the case of NAFUP. For the present study, this difference is insignificant, considering the scoring is done by novices who are less acquainted with the scoring procedure, despite training provided to them before the original experiment. By viewing the comprehensibility percentage (Table 2. 10), it is possible to determine which sign is comprehensible and which is not, and they can be arranged according to the increasing/ decreasing order of comprehensibility. ISO 3864 standardizes and claims the signs as acceptable when 67% of the target population can understand or comprehend the sign in a comprehension test (Berrio et al., 2022; ISO 3864-1, 2011), and this has been followed in this study. The study reveals that COSACOEOT, WRP, COSACOEOT_T, and UBC are incomprehensible signs according to the ISO standard and need redesign. The other signs are comprehensible and can be considered for use. The signs in the increasing order of comprehensibility are as follows:

COSACOEOT < WRP < COSACOEOT_T < UBC < WRP_T < UBC_T < DNT < NAFUP_T < NAFUP < DNWSH < DNT_T < DNWSH_T < WAFS < WAFS_T.

This means that COSACOEOT is less comprehensible than WRP. WRP is less comprehensible than COSACOEOT_T., and so forth. Overall, the sign COSACOEOT is the least comprehensible, and WAFS_T is the most comprehensible one.

Table 2. 10. Comprehensibility percentage (subject-based technique)

Sign label	Comprehensibility	Sign label	Comprehensibility
COSACOEOT	28.46715328	COSACOEOT_T	47.44525547
DNT	71.53284672	DNT_T	76.64233577

DNWSH	74.45255474	DNWSH_T	83.94160584
NAFUP	73.72262774	NAFUP_T	72.26277372
UBC	56.20437956	UBC_T	69.34306569
WAFS	89.7810219	WAFS_T	94.89051095
WRP	39.41605839	WRP_T	65.69343066

2.3.2. Expert-based technique

The Expert-based system considers textual descriptive responses, which are evaluated and scored by two industry experts based on a pre-agreed protocol. They are later normalized (averaged) and considered for further analysis.

2.3.2.1. Interrater reliability

The interrater reliability for the two experts has been calculated for all the signs. The Cronbach's Alpha is found to be 0.872 (Table 2. 11), which is an internally consistent reliable variance, and this result is quite encouraging. The test result also demonstrated a perfect internal consistency (reliability) of the scale, the score assigned by the experts, and the internal consistency between the experts.

Table 2. 11. Reliability statistics of two experts

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.872	.877	28

The intention is to check the item statistics for the two experts for all the signs. The mean and standard deviation for each sign for the two experts are almost similar, which is the desired result (Table 2. 12). The item statistics for expert scores are studied, with their mean values. The mean value for each sign on an average reveals that expert 2 (exp2) rated more than expert 1 (exp1) or assigned the highest scores to the participants (in this study, the minimum score assigned is 0.0 (zero). The maximum score assigned is 1.0 (one) with an exception in the case of DNT. This discrepancy is evident as humans tend to get confused, biased, or tired and make such judgments because human decisions are subjective.

Table 2. 12. Descriptive statistics- Mean and Standard deviation for the scores assigned by two experts (expert-based technique)

Experts' scores item statistics			
Sign_Technique	Mean	Std. Deviation	N
NAFUP_exp1	.1562	.18225	137
NAFUP_exp2	.1723	.20355	137
UBC_exp1	.2307	.15556	137
UBC_exp2	.3117	.16586	137
DNT_exp1	.6153	.42511	137
DNT_exp2	.6139	.42516	137
DNWSH_exp1	.4350	.26996	137
DNWSH_exp2	.5015	.27891	137
COSACOEOT_exp1	.0788	.21366	137
COSACOEOT_exp2	.0737	.20083	137
WRP_exp1	.2263	.23865	137
WRP_exp2	.2847	.27783	137
WAFS_exp1	.6606	.29415	137
WAFS_exp2	.7219	.28690	137
DNWSH_T_exp1	.6613	.32067	137
DNWSH_T_exp2	.7000	.31600	137
NAFUP_T_exp1	.4613	.36607	137
NAFUP_T_exp2	.4993	.37661	137
WAFS_T_exp1	.8672	.22397	137
WAFS_T_exp2	.8978	.20017	137
COSACOEOT_T_exp1	.4416	.37625	137
COSACOEOT_T_exp2	.4591	.43480	137
UBC_T_exp1	.5380	.36058	137
UBC_T_exp2	.5942	.34508	137
WRP_T_exp1	.5080	.37711	137
WRP_T_exp2	.5591	.37974	137
DNT_T_exp1	.8314	.34101	137
DNT_T_exp2	.8460	.32855	137

The inter-expert score correlation result for the two experts is tabulated diagonally in (Table 2. 13) which shows a significant correlation between the experts' assigned scores. This correlation strength has been possible as the experts agreed upon a protocol for the scoring procedure. The other portions of the matrix are removed for brevity.

Table 2. 13. Inter-Expert Correlation Matrix

Inter-Expert Correlation matrix

	NAFUP_ exp2	UBC_ exp2	DNT_ exp2	DNWSH_ exp2	COSACOEOT_ exp2	WRP_ exp2	WAFS_ _exp2	DNWSH_ T_exp2	NAFUP_T_ _exp2	WAFS_ T_exp2	COSACOEOT_ T_exp2	UBC_T_ exp2	WRP_T_ _exp2	DNT_T_ exp2	
NAFUP_exp1	0.970														
UBC_exp1		0.969													
DNT_exp1			0.993												
DNWSH_exp1				0.948											
COSACOEOT_exp1					0.979										
WRP_exp1						0.992									
WAFS_exp1							0.985								
DNWSH_T_exp1								0.978							
NAFUP_T_exp1									0.991						
WAFS_T_exp1										0.982					
COSACOEOT_T_exp1											0.822				
UBC_T_exp1												0.990			
WRP_T_exp1													0.991		
DNT_T_exp1															0.963

Here exp1 is expert 1, and exp2 is expert 2.

The intraclass correlation (ICC) is a reliability measure that assesses either the degree of consistency or absolute agreement (Kim, 2013). The model for the ICC (Table 2. 14) chosen is two-way mixed, as the raters or experts are fixed here and not random. Absolute agreement is chosen as the method evaluates how closely the experts scored each response. The ICC, and the coefficient for the lower and upper bound at a 95% confidence interval for the average measure, are above .7 (>.7), which is acceptable. Both the lower and upper bound at a 95% confidence interval are found to be greater than .7, which is acceptable. This value confirms the high interrater reliability between the two raters or experts.

Table 2. 14. Intraclass correlation coefficient (ICC) (expert-based technique)

Intraclass Correlation Coefficient							
	Intraclass Correlation (b)	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.123 ^a	.089	.167	7.799	136	3672	.000
Average Measures	.798 ^c	.733	.849	7.799	136	3672	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

- a. The estimator is the same, whether the interaction effect is present or not.
- b. Type A intraclass correlation coefficients using an absolute agreement definition.
- c. This estimate is computed assuming the interaction effect is absent because it is not estimable otherwise.

2.3.2.2. Descriptive Statistics for the final score

The final score coded by the experts is used for calculating the descriptives using IBM SPSS 25.0 with 137 cases with no missing data (Table 2. 15 and Table 2. 16). Descriptive statistics characterize data distribution.

Table 2. 15. Descriptive statistics of expert-based technique (signs without text)

		NAFUP		UBC		DNT		DNWSH		COSACOEOT		WRP		WAFS	
		Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
Mean		0.1730	0.01707	0.3124	0.01418	0.6263	0.03582	0.5007	0.02350	0.0803	0.01827	0.2839	0.02372	0.7212	0.02448
95% Confidence Interval for Mean	L.B.	0.1392		0.2844		0.5554		0.4543		0.0442		0.2370		0.6728	
	U.B.	0.2068		0.3405		0.6971		0.5472		0.1164		0.3309		0.7696	
5% Trimmed Mean		0.1443		0.3041		0.6403		0.5008		0.0455		0.2669		0.7457	
Median		0.1000		0.3000		0.8000		0.5000		0.0000		0.3000		0.8000	
Variance		0.040		0.028		0.176		0.076		0.046		0.077		0.082	
Std. Deviation		0.19982		0.16603		0.41926		0.27507		0.21380		0.27766		0.28657	
Minimum		0.00		0.00		0.00		0.00		0.00		0.00		0.00	
Maximum		1.00		1.00		1.00		1.00		1.00		1.00		1.00	
Range		1.00		1.00		1.00		1.00		1.00		1.00		1.00	
Interquartile Range		0.20		0.00		0.80		0.40		0.00		0.55		0.40	
Skewness		2.439	0.207	1.199	0.207	-0.488	0.207	0.063	0.207	2.684	0.207	0.445	0.207	-1.327	0.207
Kurtosis		7.160	0.411	4.543	0.411	-1.476	0.411	-0.683	0.411	6.360	0.411				

Here L.B.: Lower Bound; U.B.: Upper Bound

Table 2. 16. Descriptive statistics of expert-based technique (signs with text)

		DNWSH_T		NAFUP_T		WAFS_T		COSACOEOT_T		UBC_T		WRP_T		DNT_T	
		Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
Mean		0.7015	0.02686	0.5000	0.03215	0.8978	0.01710	0.4693	0.03388	0.5949	0.02949	0.5591	0.03244	0.8438	0.02803
95% Confidence Interval for Mean	L.B.	0.6483		0.4364		0.8640		0.4023		0.5366		0.4950		0.7884	
	U.B.	0.7546		0.5636		0.9316		0.5363		0.6532		0.6233		0.8992	
5% Trimmed Mean		0.7238		0.5000		0.9321		0.4611		0.6054		0.5657		0.8820	
Median		0.8000		0.6000		1.0000		0.6000		0.6000		0.6000		1.0000	
Variance		0.099		0.142		0.040		0.157		0.119		0.144		0.108	
Std. Deviation		0.31436		0.37632		0.20017		0.39660		0.34520		0.37974		0.32807	
Minimum		0.00		0.00		0.00		0.00		0.00		0.00		0.00	
Maximum		1.00		1.00		1.00		1.60		1.00		1.00		1.00	
Range		1.00		1.00		1.00		1.60		1.00		1.00		1.00	
Interquartile Range		0.50		0.80		0.20		0.80		0.70		0.65		0.00	
Skewness		-1.010	0.207	0.023	0.207	-3.017	0.207	-0.013	0.207	-0.227	0.207	-0.307	0.207	-1.943	0.207
Kurtosis		0.005	0.411	-1.537	0.411	10.382	0.411	-1.282	0.411	-1.216	0.411	-1.375	0.411	2.196	0.411
Here L.B.: Lower Bound; U.B.: Upper Bound															

2.3.2.3. Test for Normality

Normality tests (Kolmogorov-Smirnov and Shapiro-Wilk in Table 2. 17) reveal that the data, i.e., comprehensibility scores of all the 14 signs, deviates from the normal distribution (p-value less than alpha (0.05)). The test fails to accept that sample data are normally distributed (Shapiro & Wilk, 1965). Getting truly, normally distributed data in the real world is hard. However, the normality of data does not cause significant issues with a sample size greater than 25 or 30 (Hogg et al., 1977).

Table 2. 17. Test for normality (expert-based technique)

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
NAFUP	0.293	137	0.000	0.710	137	0.000
UBC	0.369	137	0.000	0.719	137	0.000
DNT	0.310	137	0.000	0.759	137	0.000
DNWSH	0.100	137	0.002	0.955	137	0.000
COSACOEOT	0.500	137	0.000	0.429	137	0.000
WRP	0.256	137	0.000	0.850	137	0.000
WAFS	0.207	137	0.000	0.810	137	0.000
DNWSH_T	0.171	137	0.000	0.839	137	0.000
NAFUP_T	0.206	137	0.000	0.865	137	0.000
WAFS_T	0.359	137	0.000	0.547	137	0.000
COSACOEOT_T	0.247	137	0.000	0.834	137	0.000
UBC_T	0.177	137	0.000	0.879	137	0.000
WRP_T	0.149	137	0.000	0.857	137	0.000
DNT_T	0.449	137	0.000	0.518	137	0.000

a. Lilliefors Significance Correction

2.3.2.4. Analysis of variance (ANOVA)

Analysis of variance (ANOVA) has been conducted for all the signs. The result of Friedman's ANOVA test (Table 2. 18 and Table 2. 19), followed by post hoc analysis, shows differences in comprehensibility levels between the signs (all sign pairs) viewed by the same group of participants. Friedman's test calculated the p-value (Table 2. 19), which is less than 0.05 and is significant here. Hence the null hypothesis, "there are no significant differences between the signs," can be rejected. The result hypothesizes the significant differences in comprehensibility levels among all the signs considered for the study.

Table 2. 18. Rank of the signs generated by Friedman's ANOVA (expert-based technique)

Ranks	
Sign_Technique	Mean Rank
NAFUP_exp	4.14
UBC_exp	5.44
DNT_exp	8.85
DNWSH_exp	7.35
COSACOEOT_exp	2.81
WRP_exp	4.82
WAFS_exp	9.47
DNWSH_T_exp	9.24
NAFUP_T_exp	7.31
WAFS_T_exp	11.64
COSACOEOT_T_exp	6.57
UBC_T_exp	8.47
WRP_T_exp	7.69
DNT_T_exp	11.19

Table 2. 19. Test statistics of Friedman's ANOVA (expert-based technique)

Test Statistics ^(a)	
N	137
Chi-Square	731.768
df	13
Asymp. Sig.	.000

a. Friedman Test

Posthoc analysis (Figure 2. 11) helps to identify significant pairs. The result (Figure 2. 12 - Figure 2. 17) shows that sixty-five (65) pairs are significantly different, and the pairs are highlighted in yellow (p-values are the same as the Adj. Sig. value in the figure) in the figures for brevity.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distributions of NAFUP_exp, UBC_exp, DNT_exp, DNWSH_exp, COSACOEOT_exp, WRP_exp, WAFS_exp, DNWSH_T_exp, NAFUP_T_exp, WAFS_T_exp, COSACOEOT_T_exp, UBC_T_exp, WRP_T_exp and DNT_T_exp are the same.	Related-Samples Friedman's Two-Way Analysis of Variance by Ranks	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 2. 11. Post hoc analysis (expert-based technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
COSACOEOT_exp-NAFUP_exp	1.328	.505	2.628	.009	.781
COSACOEOT_exp-WRP_exp	-2.004	.505	-3.964	.000	.007
COSACOEOT_exp-UBC_exp	2.628	.505	5.199	.000	.000
COSACOEOT_exp-COSACOEOT_T_exp	-3.755	.505	-7.430	.000	.000
COSACOEOT_exp-NAFUP_T_exp	-4.493	.505	-8.889	.000	.000
COSACOEOT_exp-DNWSH_exp	4.536	.505	8.975	.000	.000
COSACOEOT_exp-WRP_T_exp	-4.872	.505	-9.640	.000	.000
COSACOEOT_exp-UBC_T_exp	-5.661	.505	-11.199	.000	.000
COSACOEOT_exp-DNT_exp	6.036	.505	11.943	.000	.000
COSACOEOT_exp-DNWSH_T_exp	-6.427	.505	-12.716	.000	.000
COSACOEOT_exp-WAFS_exp	-6.661	.505	-13.178	.000	.000
COSACOEOT_exp-DNT_T_exp	-8.380	.505	-16.579	.000	.000
COSACOEOT_exp-WAFS_T_exp	-8.825	.505	-17.460	.000	.000

Figure 2. 12. Post hoc analysis result showing significantly different pairs in sequence-1 (expert-based technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
NAFUP_exp-WRP_exp	-.675	.505	-1.336	.182	1.000
NAFUP_exp-UBC_exp	-1.299	.505	-2.571	.010	.924
NAFUP_exp-COSACOEOT_T_exp	-2.427	.505	-4.802	.000	.000
NAFUP_exp-NAFUP_T_exp	-3.164	.505	-6.260	.000	.000
NAFUP_exp-DNWSH_exp	-3.208	.505	-6.347	.000	.000
NAFUP_exp-WRP_T_exp	-3.544	.505	-7.011	.000	.000
NAFUP_exp-UBC_T_exp	-4.332	.505	-8.571	.000	.000
NAFUP_exp-DNT_exp	-4.708	.505	-9.315	.000	.000
NAFUP_exp-DNWSH_T_exp	-5.099	.505	-10.087	.000	.000
NAFUP_exp-WAFS_exp	-5.332	.505	-10.549	.000	.000
NAFUP_exp-DNT_T_exp	-7.051	.505	-13.950	.000	.000
NAFUP_exp-WAFS_T_exp	-7.496	.505	-14.831	.000	.000
WRP_exp-UBC_exp	.624	.505	1.235	.217	1.000

Figure 2. 13. Post hoc analysis result showing significantly different pairs in sequence-2 (expert-based technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
WRP_exp-COSACOEOT_T_exp	-1.752	.505	-3.466	.001	.048
WRP_exp-NAFUP_T_exp	-2.489	.505	-4.924	.000	.000
WRP_exp-DNWSH_exp	2.533	.505	5.011	.000	.000
WRP_exp-WRP_T_exp	-2.869	.505	-5.675	.000	.000
WRP_exp-UBC_T_exp	-3.657	.505	-7.235	.000	.000
WRP_exp-DNT_exp	4.033	.505	7.979	.000	.000
WRP_exp-DNWSH_T_exp	-4.423	.505	-8.751	.000	.000
WRP_exp-WAFS_exp	-4.657	.505	-9.214	.000	.000
WRP_exp-DNT_T_exp	-6.376	.505	-12.614	.000	.000
WRP_exp-WAFS_T_exp	-6.821	.505	-13.495	.000	.000
UBC_exp-COSACOEOT_T_exp	-1.128	.505	-2.231	.026	1.000
UBC_exp-NAFUP_T_exp	-1.865	.505	-3.690	.000	.020
UBC_exp-DNWSH_exp	-1.909	.505	-3.776	.000	.014
UBC_exp-WRP_T_exp	-2.245	.505	-4.441	.000	.001
UBC_exp-UBC_T_exp	-3.033	.505	-6.000	.000	.000
UBC_exp-DNT_exp	-3.409	.505	-6.744	.000	.000

Figure 2. 14. Post hoc analysis result showing significantly different pairs in sequence-3 (expert-based technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
UBC_exp-DNWSH_T_exp	-3.799	.505	-7.517	.000	.000
UBC_exp-WAFS_exp	-4.033	.505	-7.979	.000	.000
UBC_exp-DNT_T_exp	-5.752	.505	-11.380	.000	.000
UBC_exp-WAFS_T_exp	-6.197	.505	-12.261	.000	.000
COSACOEOT_T_exp-NAFUP_T_exp	.737	.505	1.459	.145	1.000
COSACOEOT_T_exp-DNWSH_exp	.781	.505	1.545	.122	1.000
COSACOEOT_T_exp-WRP_T_exp	-1.117	.505	-2.210	.027	1.000
COSACOEOT_T_exp-UBC_T_exp	-1.905	.505	-3.769	.000	.015
COSACOEOT_T_exp-DNT_exp	2.281	.505	4.513	.000	.001
COSACOEOT_T_exp-DNWSH_T_exp	2.672	.505	5.286	.000	.000
COSACOEOT_T_exp-WAFS_exp	2.905	.505	5.748	.000	.000
COSACOEOT_T_exp-DNT_T_exp	-4.624	.505	-9.149	.000	.000
COSACOEOT_T_exp-WAFS_T_exp	5.069	.505	10.029	.000	.000
NAFUP_T_exp-DNWSH_exp	.044	.505	.087	.931	1.000
NAFUP_T_exp-WRP_T_exp	-.380	.505	-.751	.453	1.000
NAFUP_T_exp-UBC_T_exp	-1.168	.505	-2.311	.021	1.000

Figure 2. 15. Post hoc analysis result showing significantly different pairs in sequence-4 (expert-based technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
NAFUP_T_exp-DNT_exp	1.544	.505	3.054	.002	.205
NAFUP_T_exp-DNWSH_T_exp	1.934	.505	3.827	.000	.012
NAFUP_T_exp-WAFS_exp	2.168	.505	4.289	.000	.002
NAFUP_T_exp-DNT_T_exp	-3.887	.505	-7.690	.000	.000
NAFUP_T_exp-WAFS_T_exp	-4.332	.505	-8.571	.000	.000
DNWSH_exp-WRP_T_exp	-.336	.505	-.664	.506	1.000
DNWSH_exp-UBC_T_exp	-1.124	.505	-2.224	.026	1.000
DNWSH_exp-DNT_exp	1.500	.505	2.968	.003	.273
DNWSH_exp-DNWSH_T_exp	-1.891	.505	-3.740	.000	.017
DNWSH_exp-WAFS_exp	-2.124	.505	-4.202	.000	.002
DNWSH_exp-DNT_T_exp	-3.843	.505	-7.603	.000	.000
DNWSH_exp-WAFS_T_exp	-4.288	.505	-8.484	.000	.000
WRP_T_exp-UBC_T_exp	.788	.505	1.560	.119	1.000
WRP_T_exp-DNT_exp	1.164	.505	2.303	.021	1.000
WRP_T_exp-DNWSH_T_exp	1.555	.505	3.076	.002	.191
WRP_T_exp-WAFS_exp	1.788	.505	3.538	.000	.037

Figure 2. 16. Post hoc analysis result showing significantly different pairs in sequence-5 (expert-based technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
WRP_T_exp-DNT_T_exp	-3.507	.505	-6.939	.000	.000
WRP_T_exp-WAFS_T_exp	3.953	.505	7.820	.000	.000
UBC_T_exp-DNT_exp	.376	.505	.744	.457	1.000
UBC_T_exp-DNWSH_T_exp	.766	.505	1.516	.129	1.000
UBC_T_exp-WAFS_exp	1.000	.505	1.978	.048	1.000
UBC_T_exp-DNT_T_exp	-2.719	.505	-5.379	.000	.000
UBC_T_exp-WAFS_T_exp	3.164	.505	6.260	.000	.000
DNT_exp-DNWSH_T_exp	-.391	.505	-.773	.440	1.000
DNT_exp-WAFS_exp	-.624	.505	-1.235	.217	1.000
DNT_exp-DNT_T_exp	-2.343	.505	-4.636	.000	.000
DNT_exp-WAFS_T_exp	-2.788	.505	-5.517	.000	.000
DNWSH_T_exp-WAFS_exp	.234	.505	.462	.644	1.000
DNWSH_T_exp-DNT_T_exp	-1.953	.505	-3.863	.000	.010
DNWSH_T_exp-WAFS_T_exp	-2.398	.505	-4.744	.000	.000
WAFS_exp-DNT_T_exp	-1.719	.505	-3.401	.001	.061
WAFS_exp-WAFS_T_exp	-2.164	.505	-4.282	.000	.002
DNT_T_exp-WAFS_T_exp	.445	.505	.881	.378	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
Significance values have been adjusted by the Bonferroni correction for multiple tests.

Figure 2. 17. Post hoc analysis result showing significantly different pairs in sequence-6 (expert-based technique)

2.3.2.5. OSH sign with OSH-related textual information affects the comprehensibility of the messages intended to convey (expert-based)

Considering the hypothesis,

Null Hypothesis (H₀): The level of comprehensibility of the OSH sign is not significantly different when the sign is presented with or without textual information.

Alternate Hypothesis (H₃): The level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information.

To test the hypothesis, a Wilcoxon signed rank test has been conducted, and the result is shown in the following tables. Descriptive statistics (Table 2. 20) shows the mean of all the signs, indicating the significant difference between the signs with and without textual information. The mean of the signs with text is more than the mean of the signs without text. This difference is significant, and the result is the same as expected from this technique. The minimum and maximum values signify the minimum and maximum scores assigned for that sign under study.

Table 2. 20. Descriptive statistics for the expert-based comprehensibility score in the Wilcoxon Signed Rank test

Descriptive Statistics					
Sign_Technique	N	Mean	Std. Deviation	Minimum	Maximum
NAFUP_exp	137	.1730	.19982	.00	1.00
UBC_exp	137	.3124	.16603	.00	1.00
DNT_exp	137	.6263	.41926	.00	1.00
DNWSH_exp	137	.5007	.27507	.00	1.00
COSACOEOT_exp	137	.0803	.21380	.00	1.00
WRP_exp	137	.2839	.27766	.00	1.00
WAFS_exp	137	.7212	.28657	.00	1.00
NAFUP_T_exp	137	.5000	.37632	.00	1.00
UBC_T_exp	137	.5949	.34520	.00	1.00
DNT_T_exp	137	.8438	.32807	.00	1.00
DNWSH_T_exp	137	.7015	.31436	.00	1.00
COSACOEOT_T_exp	137	.4693	.39660	.00	1.60
WRP_T_exp	137	.5591	.37974	.00	1.00
WAFS_T_exp	137	.8978	.20017	.00	1.00

Note: i. _exp: denotes the expert-based technique.

ii. _T: denotes textual information.

The hypothesis is confirmed further by considering ranks. It is found that the mean of the positive rank and the sum of the positive ranks (Table 2. 21) are more for the signs with textual information than the signs without textual information.

Table 2. 21. Wilcoxon Signed Rank test ranks for the sign pairs: with and without text (expert-based technique)

		N	Mean Rank	Sum of Ranks
NAFUP_T_exp - NAFUP_exp	Negative Ranks	30 ^a	19.70	591.00
	Positive Ranks	81 ^b	69.44	5625.00
	Ties	26 ^c		
	Total	137		
UBC_T_exp - UBC_exp	Negative Ranks	15 ^d	31.23	468.50
	Positive Ranks	92 ^e	57.71	5309.50
	Ties	30 ^f		
	Total	137		
DNT_T_exp - DNT_exp	Negative Ranks	8 ^g	15.13	121.00
	Positive Ranks	51 ^h	32.33	1649.00
	Ties	78 ⁱ		
	Total	137		
DNWSH_T_exp - DNWSH_exp	Negative Ranks	32 ^j	39.28	1257.00
	Positive Ranks	85 ^k	66.42	5646.00
	Ties	20 ^l		
	Total	137		
COSACOEOT_T_exp - COSACOEOT_exp	Negative Ranks	3 ^m	30.83	92.50
	Positive Ranks	81 ⁿ	42.93	3477.50
	Ties	53 ^o		
	Total	137		
WRP_T_exp - WRP_exp	Negative Ranks	15 ^p	16.90	253.50
	Positive Ranks	69 ^q	48.07	3316.50
	Ties	53 ^r		
	Total	137		
WAFS_T_exp - WAFS_exp	Negative Ranks	8 ^s	23.81	190.50
	Positive Ranks	77 ^t	44.99	3464.50
	Ties	52 ^u		
	Total	137		

a. NAFUP_T_exp < NAFUP_exp

h. DNT_T_exp > DNT_exp

o. COSACOEOT_T_exp = COSACOEOT_exp

b. NAFUP_T_exp > NAFUP_exp

i. DNT_T_exp = DNT_exp

p. WRP_T_exp < WRP_exp

c. NAFUP_T_exp = NAFUP_exp

j. DNWSH_T_exp < DNWSH_exp

q. WRP_T_exp > WRP_exp

d. UBC_T_exp < UBC_exp

k. DNWSH_T_exp > DNWSH_exp

r. WRP_T_exp = WRP_exp

e. UBC_T_exp > UBC_exp

l. DNWSH_T_exp = DNWSH_exp

s. WAFS_T_exp < WAFS_exp

The number of positive ranks for all the signs with text is more than for signs without text. For example, the number of positive ranks for the NAFUP_T (NAFUP_T_sub > NAFUP_sub) is found to be 81 (Table 2. 21), i.e., 81 cases the participants gave more scores

to the NAFUP_T than the NAFUP. The table also shows 26 cases when the NAFUP_T and NAFUP have been ranked similarly (tie).

The experts in this technique are experienced in scoring the written descriptive responses of the safety messages, and they scored according to the pre-determined protocol. Hence the scores are bound to be less ambiguous than the scores assigned by the participants in the subject-based technique.

The Z value (Table 2. 22) helped to calculate the effect size (r-value) (Table 2. 23), which assists in checking how strong is the effect of the text on the comprehensibility of the sign(s). The effect size is found to be significant for all the pairs, which is excellent for our study.

Table 2. 22. Test statistics for Wilcoxon Signed Rank test for the sign pairs: with and without text (expert-based technique)

Test Statistics ^(a)							
	NAFUP_T_sub - NAFUP_sub	UBC_T_sub - UBC_sub	DNT_T_sub - DNT_sub	DNWSH_T_sub - DNWSH_sub	COSACOEOT_T_sub - COSACOEOT_sub	WRP_T_sub - WRP_sub	WAFS_T_sub - WAFS_sub
Z	-7.421 ^b	-7.554 ^b	-5.784 ^b	-5.990 ^b	-7.579 ^b	-6.845 ^b	-7.235 ^b
Asymp. Sig.	0.000	0.000	0.000	0.000	0.000	0.000	0.000

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

c. Based on negative ranks.

Table 2. 23. Effect Size for Wilcoxon Signed Rank test for the sign pairs: with and without text (expert-based technique)

	COSACOEOT_T_score - COSACOEOT_score	DNT_T_score - DNT_score	DNWSH_T_score - DNWSH_score	NAFUP_T_score - NAFUP_score	UBC_T_score - UBC_score	WAFS_T_score - WAFS_score	WRP_T_score - WRP_score
Z	-7.579	-5.784	-5.99	-7.421	-7.554	-7.235	-6.845
N	274	274	274	274	274	274	274
Root N	16.55294536	16.55294536	16.55294536	16.55294536	16.55294536	16.55294536	16.55294536
r	-0.457864135	-0.349424219	-0.361869134	-0.448319005	-0.456353829	-0.437082335	-0.413521573

Hence the interpretation of the result for all the sign pairs can be as follows:

- i. A Wilcoxon signed rank test, and the calculated effect size reveals that for all the seven sign pairs, textual information positively affects the comprehensibility of message/ sign evident from their mean difference, z value, p-value, and r value (Table 2. 20 – Table 2. 23).
- ii. Hence null hypothesis can be rejected, and the alternate hypothesis has been established that the level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information.

2.3.2.6. Comprehensibility percentage based on expert-based technique

The comprehensibility percentage for each sign is calculated based on the rating/ score of the experts (Table 2. 24). The sign whose total comprehensibility percentage is more than or equal to 67% (Berrio et al., 2022; ISO 3864-1, 2011) is considered acceptable. This means that a minimum of 67% target population found the signs comprehensible.

Besides the ANOVA and hypothesis test (sections 2.3.2.4. and 2.3.2.5.), this comprehensibility percentage reveals that signs with text are more comprehensible than signs without text.

By viewing the Comprehensibility Percentage (Table 2. 24), it is possible to determine which sign is comprehensible and which is not, and they can be arranged according to the increasing/ decreasing order of comprehensibility. ISO 3864 standardizes and claims the signs as acceptable when 67% of the target population can understand or comprehend the sign in a comprehension test (Berrio et al., 2022; ISO 3864-1, 2011), and this has been followed in this study. The study reveals that COSACOEOT, COSACOEOT_T, DNT, DNWSH, NAFUP, NAFUP_T, UBC, UBC_T, WRP, and WRP_T are incomprehensible signs according to the ISO standard, and they need redesign. The other signs, DNT_T, DNWSH_T, WAFS, and WAFS_T, are comprehensible and can be considered for use. It is seen that the sign NAFUP is less comprehensible than COSACOEOT which in turn is less comprehensible than UBC, with WAFS_T as the most comprehensible sign. It is to be noted that the sign DNT and UBC_T have the same comprehensibility percentage of 63.50364964 (Table 2.24). Similarly, the comprehensibility percentage for DNWSH and NAFUP_T is similar i.e., 54.74452555 (Table 2.24).

Table 2. 24. Comprehensibility percentage (expert-based technique)

Sign label	Comprehensibility Percentage [%]	Sign label	Comprehensibility Percentage [%]
COSACOEOT	9.489051095	COSACOEOT_T	56.93430657
DNT	63.50364964	DNT_T	86.86131387
DNWSH	54.74452555	DNWSH_T	82.48175182
NAFUP	6.569343066	NAFUP_T	54.74452555
UBC	11.67883212	UBC_T	63.50364964
WAFS	89.7810219	WAFS_T	96.35036496
WRP	32.84671533	WRP_T	64.23357664

2.4. Conclusion

The two evaluation techniques have been implemented and applied based on subjective data. These subject-based and expert-based techniques are both supervised by experts. The subject-based technique suffers from a few drawbacks. Among them, one is the biased score. Despite training the subjects formally, they are found to provide ambiguous scores. As all the participants rate or score the questionnaire based on their understanding, their scores are biased based on their previous experiences, education, work environment, previous exposure to that sign or related sign, the self-judgment of their comprehension level, and their actual comprehension level at that moment, etc. The participants were pre-trained on the scoring procedure before the actual experiment; however, the result reveals that there are times when participants rated themselves more when they comprehended the sign less and rated less when they comprehended the sign successfully.

Another technique, the expert-based technique, is implemented to overcome the drawback of the subject-based technique. The experts scoring the responses are experienced and qualified for the process. They are less biased and rated the response of the subjects based on pre-agreed terms.

Data are checked for descriptives and normality for both techniques before statistical analyses. The study found significant differences in comprehensibility levels among the signs under study. The study also found that the level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information.

A complement to the existing ISO method has been introduced, a variation of the ISO 3186 standard that tries to capture the participants' minds more by considering an 11-point rating scale. The two new categories generated from an 11-point scale introduced a novel technique of coding data into two categories- comprehensible and incomprehensible based on a binning threshold. The comprehensible category ranges from 0.5 to 1.0. The incomprehensible category contains a range of values that ranges from 0.0 to 0.4. A score of 0.4 is better than 0.0, and this difference is not reflected in the final category of results in the present study, which is out of the scope of the thesis. This study can be extended further by adding more categories besides the two categories-comprehensible/ incomprehensible. Future studies can be conducted by adding more categories that will help to capture more minds and ideas of the end users involved in the study, and it might assist the designers in further improvement in the sign evaluation process.

The subject-based evaluation found UBC_T, DNT, NAFUP_T, NAFUP, DNWSH, DNT_T, DNWSH_T, WAFS, and WAFS_T as the comprehensible signs, with WAFS_T as the most comprehensible one. The expert-based evaluation found DNWSH_T, DNT_T, WAFS, and WAFS_T comprehensible, with WAFS_T as the most comprehensible similar to the subject-based technique.



3

Evaluation of Comprehensibility of OSH Signs by Computational Technique

Chapter 3: Evaluation of Comprehensibility of OSH Signs by Computational Technique

Abstract

This chapter overcomes the drawbacks of the expert-based technique. A novel computational model is developed capable of automatically evaluating the comprehensibility of an OSH sign without much expert intervention. The automated model evaluates the questionnaire responses in the form of written and descriptive text; and categorizes them as comprehensible or non-comprehensible. If executed simultaneously for innumerable signs, the model can evaluate them in a fraction of a second, thus eliminating the expert-based technique scoring and the burden of repetitive scoring or rating of enormous questionnaires/surveys. This technique uses a deep neural network (DNN) and deep averaging network (DAN) to interpret participants' responses. The coded responses, i.e., data, are tested for their distribution and normality. Two hypothesis tests have been conducted considering the effect of text on the comprehensibility of a sign and variances among all the signs.

This technique uses the same binning threshold for converting the 11-point scale into binary categories as has been followed in the subject and expert-based techniques, evaluates the sign according to the ISO standard, and categorizes them as comprehensible/incomprehensible based on the ISO standard (67%). This computational model generates systematic, consistent, and unbiased comprehensibility scores. The automated model's generated score is validated by comparing it with the English language model (existing STS benchmark) and the separate expert-based technique (score). The Pearson correlation r for both is found to be more than 0.80 and the p -value 0.0, which is significant for the present study.

3.1. Introduction

There are a variety of methods or techniques applied for the evaluation of the comprehensibility of a sign like multiple choice questions, short questions, ISO methods, matching items, true/ false, etc., written mode, verbal mode, or their combination(s) (Arcia et al., 2019; Chaudhuri et al., 2022). All the techniques to date focus on and discuss the result and interpretation based on the above-discussed techniques; however, there is hardly discussion on administering challenges of the methods. This study addresses one of the existing issues/challenges from the existing techniques and those studied in Chapter 2. Though the expert-based technique is superior to the subject-based technique, it suffers a significant drawback. The written documents are evaluated by human experts who score and decide whether participants/ subjects can comprehend the sign. This manual procedure is time-consuming and monotonous. It is not considered that using this method can give rise to unwanted errors when the evaluators are human beings and can get tired due to prolonged work. The evaluators can also commit unwanted errors due to personal biases, self-doubt, previous experiences, age, context, or persuasion. This research gap is considered in this study, and an attempt has been made to eradicate these human errors and speed up the evaluation procedure using a novel computational technique.

Dowse, in one of his studies, mentioned: *“The success of pictograms in communicating a message demands a rigorous, multistage, consultative pictogram design-and-test process which, although time-consuming, is the only way to generate quality new pictograms or to modify existing ones”* (Dowse, 2021). Among all the aspects of successful message communication discussed by Dowse, a complement to the test procedure has been considered and designed by keeping all the other aspects similar. The attempt is to minimize the evaluation time to fractions of seconds by automating the evaluation and accelerating the process using the proposed novel technique. This automated model is named a “Deep Averaging Network based Man-Machine-Hybrid (DAN-MMH)” model based on its algorithm DAN used for text classification and, finally, mapping semantic textual similarity (STS) between the texts. STS measures the equivalence of meaning between two textual segments – the subjects' interpretation of the OSH sign in the form of text and the standard interpretation provided by the OSH experts/ standards. Besides, STS has been successfully applied in Machine Translation, where the semantics information of the translated output has been compared with the input text used for translation (Comelles & Atserias, 2019). STS finds its extensive application in the field of information extraction. It has been widely used in discovering all possible relations from the text (Shinyama & Sekine, 2006). STS has its wide application in text summarization or sentence ranking (Alami et al., 2019), question answering (Soares & Parreiras, 2020), textual entailment, which determines if one text is implied from another text (Bentivogli et al., 2016), and paraphrase detection that identifies gradation of meaning overlap (D. Chen & Dolan, 2011; W. Xu et al., 2015). The concept of STS is detailed more in the methodology section.

The reason behind selecting DAN is that it has an established high baseline performance record for text classification jobs (Cer, Yang, Kong, Hua, Limtiaco, St. John, et al., 2018). Another apparent reason proved from the previous literature is that DAN is a simple model without any irrelevant complexity (Iyyer et al., 2015). DAN benefits itself with the performances or accuracies of the syntactic composition by exploiting the unordered composition function “Neural bag of Words” (NBOW) model (Iyyer et al., 2015). The output of DAN is non-linear, which is based on simple averaging of the words. There is no necessity to deeply analyze the syntax or the word order to obtain a better result. The complexity is enhanced only when there is an increment in the number of layers rather than an increment in the number of nodes. Each layer of DAN goes through a single matrix multiplication, and this is the cause why DAN has been chosen here for sentence recognition and comprehension over other models (Zhou et al., 2020). This model transforms sentences into embedding vectors that further use ‘transfer learning’ for NLP tasks. The deep learning method seems to outperform other statistical learning methods, and hence the same has been exploited here in this study (Cer, Yang, Kong, Hua, Limtiaco, John, et al., 2018; Iyyer et al., 2015; Zhou et al., 2020).

The DAN-MMH model developed can work cooperatively with organizations' experts to evaluate the comprehensibility of the messages. The model will assist the experts in decision making whether a designed sign is comprehensible to the participants for whom the study has been conducted. The developed model can act as an assistive tool in delivering solutions to the OSH experts regarding the comprehensibility of a sign from the users' perspective and the

decision-making process from the experts' perspective on whether to accept the designed sign or consider it for a redesign. The model cuts off the repeating unintelligent work that is not worthwhile for such human experts. Such expert brains can be channelized for better sustainable design and other intellectual safety-related work. The greater perspective can be the technology intervention in preventing OSH hazards in the workplace.

3.2. Methodology

The diagram for the DAN-MMH model (Figure 3. 1) depicts the overall flow of the technique. The methodology has been described in detail in the following sub-sections.

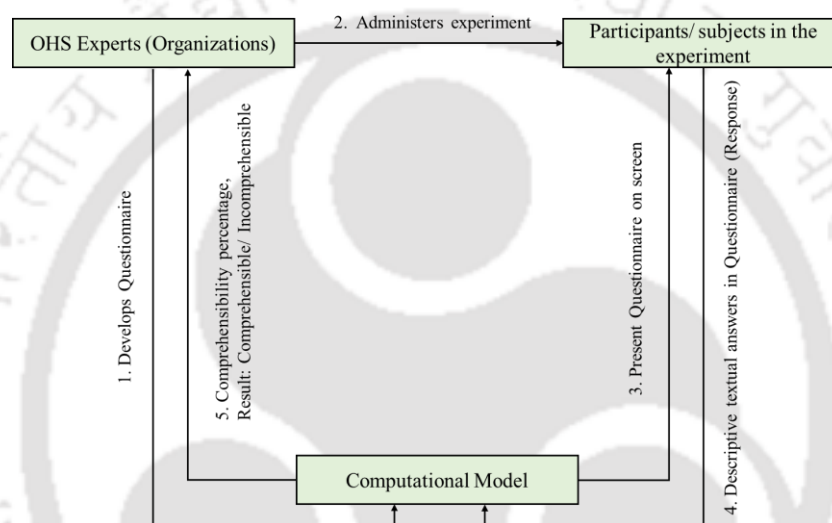


Figure 3. 1. Overview of DAN-MMH model

3.2.1. Stimuli

Fourteen (14) signs, seven with written text and seven same pairs without any written text have been considered as stimuli for the study. They are standard OSH-related signs (three pairs of prohibition signs and four pairs of mandatory signs) as used in the previous studies (Figure 2.1). They are used as test material for the experiment. Each image size is approximately 42 mm × 42 mm in line with the ISO standards and is colored. Participants are kept at a viewing distance between 400 mm and 700 mm (Caffaro & Cavallo, 2015; Shieh & Lee, 2007). Comprehension of a pictogram is found to improve when information regarding context is provided (Galer, 1980; Lehto, 1992; Vukelich & Whitaker, 1993). The context is elaborated in the form of words in written textual format. A separate instruction page has been given to the participants describing the intention of the experiment. The detailed procedure for the trial has been recorded (4 mins and 58 secs audio-video recording) in Microsoft PowerPoint, keeping the image different from the experimental image. The participants have been provided with the recording and instructed to watch it before the experiment.

3.2.2. Participants

One hundred and thirty-seven (N=137; 111 males and 26 females) employees out of four hundred (communicated through e-mails) volunteered for the study (from 15 different states of India) with no history of color deficiency and severe eye conditions and or visual impairment (self-reported). According to the International Organization for Standardization (ISO), the sample size must be 50 (ISO, 2014b, 2014a) for each symbol's variants in each country. However, in this study, Cochran's formula (Cochran, 1977; Schillaci & Schillaci, 2022) has been used to determine the sample size, which is 133.

All participants selected are healthy and white-collar employees of a food manufacturing company with a mean age of 33.85 and a standard deviation of 7.2. Informed consent was obtained from all the participants before the trial. Before scribing their responses, information was provided to view each sign for a minimum duration of 160 to 200 milliseconds (Caplette et al., 2020). The experiment has been conducted following ethical guidelines and principles.

3.2.3. Research Instrument and Research Design

The designed questionnaire has been segregated into two parts, and they are online. The first part asked for the respondents' demographic data like age, place, educational background, gender, and work experience. The second part of the questionnaire helped collect sign comprehensibility data for this technique. Each sign (stimuli) is placed on top of every page, followed by questions. Enough space has been provided for writing the responses.

In *computational technique*, the participants were asked to write the descriptive textual answer to the question, "*What message does the sign convey?*". This part of the question is subjective and open-ended, and the computer (machine) evaluates the responses, generates the scores, categorizes them into comprehensible or non-comprehensible, and returns the results to the OSH experts. This technique acts as an assistive tool for the experts who, based on the results, decide whether to accept the sign or recommend a redesign and repeat the evaluation process.

The computational technique proposed here for evaluating the comprehensibility of the OSH sign is novel. The DAN-MMH model proposed and developed is a new concept. No such model has been found so far in the literature on OSH, which can evaluate the OSH message automatically, within a few seconds, and unambiguously with an accuracy of more than 80%, in line with human experts. This capability of evaluating the comprehensibility of innumerable signs in a fraction of time without human intervention or committing a mistake is a novel idea.

Automated Evaluation Strategy:

The model evaluates the comprehensibility of OSH messages using a deep neural network (DNN). The overall diagram for the comprehensibility evaluation model for the computational technique is shown in Figure 3. 2. The proposed model takes descriptive

textual responses (OSH messages) as the raw input, which is digitized and further pre-processed according to the system requirement. Pre-processing includes grammatical syntax checking and spelling corrections.

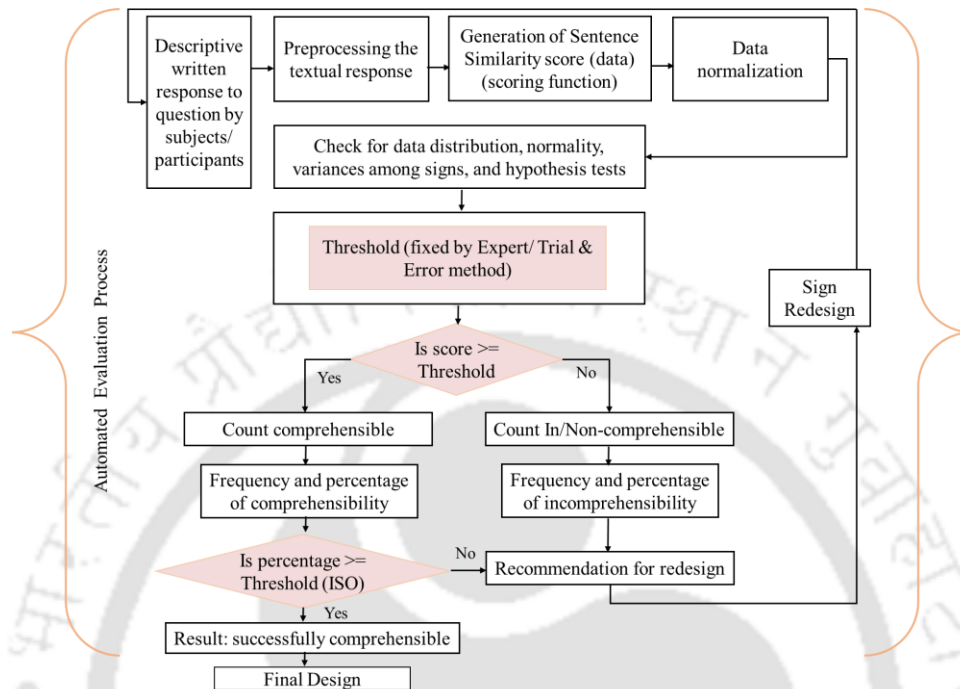


Figure 3. 2. Comprehensibility evaluation model of OSH sign (computational technique)

The scoring function calculates the sentence similarity between the standard interpretation (S.I.) of the sign and the textual response provided by the experiment participants. S.I. for each sign is collected by consulting the OSH experts or standard safety database (ISO/ANSI).

An algorithm has been proposed to compute comprehensibility, named COMPUTE COMPREHENSIBILITY, as provided below.

Algorithm: COMPUTE COMPREHENSIBILITY

Input: $M.S.$ score for each participant, Threshold Value ($T.H.$), $Count_c := 0$, $Count_{nc} := 0$, where $M.S.$ score is the semantic similarity score generated from equation (3.3); Threshold Value ($T.H.$) is set by the expert; $Count_c$ is the variable for counting comprehensibility in terms of how many participants found it comprehensible; $Count_{nc}$ is the variable for counting incomprehensibility in terms of how many participants found it incomprehensible.

- i. $S.V. := M.S.$; /* Scoring function: Model generated similarity score (Sim_{score}), refer to equation (3.3) */
- ii. IF $SV \geq TH$ /*Compare Function*/
- iii. THEN,

- iv. $C_y = 1$; /* C_y signifies comprehensibility, 1 signifies comprehensible*/
- v. $Count_c = Count_c + 1$; /* Counts the total number of ($C_y = 1$) */
- vi. ELSE
- vii. $C_y = 0$; /*0 signifies incomprehensible*/
- viii. $Count_{nc} = Count_{nc} + 1$; /* Counts the total number of ($C_y = 0$) */

In the scoring function, the scoring variable (S.V.) illustrates the normalized score or model-generated score (M.S.) (3.2) computed by the proposed model. M.S. score signifies the semantic similarity between two sentences. The OSH expert sets the standard for converting the 11-point scale to a binary scale assuming the severity of the danger is less. The algorithm for producing the value of comprehensibility (0 or 1) for an OSH message has been represented here as COMPUTE COMPREHENSIBILITY. The *M.S.* scores have been normalized up to one digit. Depending upon the value of *M.S.*, the 'compare function' in COMPUTE COMPREHENSIBILITY determines the value of C_y to be 0 or 1. A new term is coined for comprehension analysis, i.e., comprehensibility decision making (*CDM*) in terms of percentage, when $C_y = 0$. The details about *CDM* are discussed in phase 4. This *CDM* value decides whether a sign needs to be considered for a redesign and reevaluation or whether it is comprehensible and suitable for use in organizations. Finally, a decision is made where the proposed model decides whether the sign is selected or rejected.

3.2.3.1. Phase-wise description of the Model

The comprehensibility evaluation model considers the “OSH message” as the system input. The “OSH message” is the descriptive textual response of the participants in sentence form, which is collected from the experiment. The OSH messages are recorded, digitized, and fed into the proposed DAN- MMH model. The model evaluates the messages, checks semantic similarity between the standard S.I. and the OSH message, generates the scores, and categorizes the output as “Comprehensible” or “incomprehensible.”

Semantic similarity checks the similarity of the message a participant seems to convey compared to the standard interpretation (S.I.). S.I. is considered the standard message/ sentence provided by OSH experts. Those two sentences/ paragraphs are compared so that their similarity or matching scores can be computed. The higher the matching score, the more comprehensible the OSH message is. The computational details can be found in the subsequent phases.

To be more precise, sentences are semantically similar if they have responses of similar distribution. Considering an example like “How old is your father?” and “What is your father’s age?” these are both questions about a father’s age that can be responded to by a similar response like “My father is 75 years old”.

In contrast, “How is your father?” and “How old is your father?” have different meanings, resulting in different responses regardless of containing many identical words.

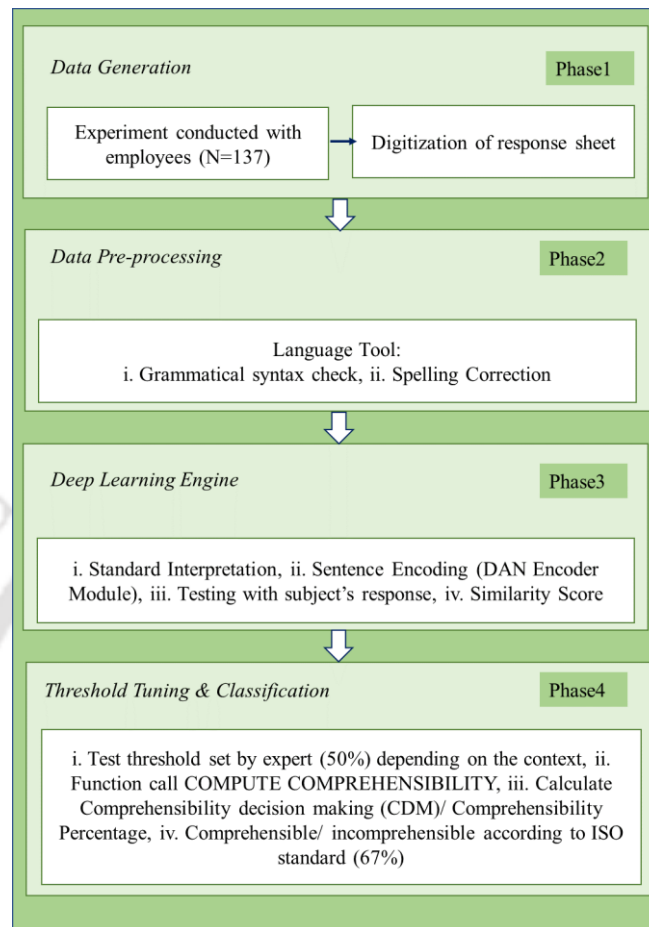


Figure 3. 3. Phase-wise diagram of the computational model

The DAN-based encoder model encodes a sentence into fixed-length 512-dimensional embeddings (Cer, Yang, Kong, Hua, Limtiaco, St. John, et al., 2018). The work of the encoder is to transform the categorical values of the relevant features (texts and paragraphs) into numerical values understood by machine learning models. The overall phases of the proposed model are presented in Figure 3. 3. The phases are detailed below:

Phase 1

Standard interpretations (S.I.) are collected for each sign (7 pairs) and initialized to value 1. The participants attended an online experiment and wrote answers on the computer. They spent a maximum of twenty minutes completing the experiment. The responses have been automatically recorded in an Excel sheet. S.I. and the written responses are used for further processing.

Phase 2

The responses are assessed for syntax and semantics (spelling and grammar) using an online language tool (<https://languagetool.org/>). The errors are then corrected before feeding the inputs to the proposed DNN-based MMH model. Initially, the inputs (S.I. and responses) are tokenized (Cer, Yang, Kong, Hua, Limtiaco, St. John, et al., 2018). Tokenization splits sentences into smaller components like words, punctuation marks, numbers, and other discrete items (Kibble, 2013). This model uses Penn Treebank (PTB) for the tokenization method.

Phase 3

The DAN encoding model is then fed with these PTB-tokenized inputs to encode them into embedded vectors (Iyyer et al., 2015). Tensorflow has been exploited to implement the model (Abadi et al., 2016). TensorFlow is an open-source platform for machine learning applications. Each input text or message produces an embedding sentence output vector of dimension 512. The encoder has been trained using unsupervised training data from various web sources, like web question-answer pages, web news, discussion forums, and Wikipedia. Wide and deep learning generated distributed representations from categorical features (Abadi, Barham, et al., 2016). In TensorFlow, distributed representation embeds the training tensors as a pattern of activity across many neurons, which is learned by backpropagation (Hinton, 1986). The encoder is further augmented with supervised training data from the Standard Natural Language Inference (SNLI) corpus (Bowman et al., 2015). These pre-trained models (Conneau et al., 2017) using sentence-level embeddings have been used here for the natural language processing (NLP) tasks. The model utilizes the deep averaging network (DAN) technique, where the word input embeddings and bigrams are first averaged together. As the vector average (V.A.) passes through the multiple hidden layers, the tiny but meaningful differences in the V.A. get magnified (Iyyer et al., 2015). The V.A.s are fed into a feedforward deep neural network (DNN) to produce sentence embeddings (Cer, Yang, Kong, Hua, Limtiaco, St. John, et al., 2018; Skansi, 2018). The logic behind using feedforward DNN is that each layer learns to represent the input more abstractly than the previous layer (Bengio et al., 2013). The proposed model has been implemented in TensorFlow using Python as a programming language, and the execution of the code is GPU-based.

The model trained with SNLI helped to improve transfer performance (Conneau et al., 2017). The use of pre-trained models helped to avoid expensive training. Word pairs from Psychology literature are also considered in the model to characterize human-like bias (Caliskan et al., 2017). The encoder is adaptable enough to handle sentences, phrases, and paragraphs as it has been optimized for greater-than-word length text. Two encoders have been used to compute the pairwise semantic similarity between the sentences. The sentence embeddings ‘ u ’ and ‘ v ’ have been normalized, and the ‘dot product’ calculates cosine similarity (3.2). Cosine similarity averages the word input embeddings and bigrams together through multiple layers so that the ordering of the words remains intact.

$$\cos_{sim}(u,v) = \arccos\left(\frac{uv}{\|u\|\|v\|}\right) \quad (3.2)$$

Finally, the similarity score is calculated using equation (3.3)

$$Sim_{score} = \left(\frac{1 - \cos_{sim}(u,v)}{\pi} \right) \quad (3.3)$$

Normalization has been applied to the Sim_score values in the range of 0 to 1. Negative correlations have been normalized to 0, thus reorienting the negative correlation to ‘no’ or absence of correlations between the two texts. This model employs a semantic textual similarity (STS) benchmark. The real-valued semantic similarity score has been found to exhibit fantastic performance on the STS Benchmark (Cer et al., 2017; Cer, Yang, Kong, Hua, Limtiaco, St. John, et al., 2018). STS measures the degree of semantic equivalence (equivalence of meaning representation) between two sentences, similar to the human OSH experts trying to find the similarity between the SI and the participants' interpretation. STS in the English language is a well-studied problem. It has accomplished a cutting-edge system, possessing pragmatic and world knowledge, scaling heights for more than 70% correlation with human expert judgment (Cer et al., 2017). The present study reached an even better correlation (more than 80%) with the current test dataset.

Deep Averaging Network (DAN) based Framework

DAN utilizes the unordered composition function “Neural bag of Words” (NBOW) model to its advantage with the accuracies of the syntactic composition, circumventing the complexities of the syntactic composition (Iyyer et al., 2015). Simply put, a composition function is a mathematical procedure combining and accommodating multiple words in a single vector. To explain NBOW, let us consider an example of text classification: mapping a sequence of input tokens X to one of the k labels. A composition function h is applied to the sequence of the word embeddings u_w for $w \in X$. The output of the composition function is vector o , and this vector seriatim is used as an input to a logistic regression function. Here h average word embeddings (3.4),

$$o = h(w \in X) = \frac{1}{|X|} \sum_{w \in X} u_w \quad (3.4)$$

o is fed into a softmax layer which induces estimated probabilities \hat{y} for each output label (3.5)

$$\hat{y} = \mathit{softmax}(W_s \cdot o + b) \quad (3.5),$$

where the SoftMax function is (equation (3.6))

$$\mathit{softmax}(q) = \frac{\exp q}{\sum_{j=1}^k \exp q_j} \quad (3.6)$$

W_s is described as $k \times d$ matrix for a dataset with k output labels, and b as a bias. NBOW model is trained for minimizing cross-entropy error $E(\hat{y})$. This minimized error makes the

model a better version of itself. A single training instance with ground-truth label y is expressed in equation (3.7)

$$E(\hat{y}) = \sum_{p=1}^k y_p \log(\hat{y}_p) \quad (3.7)$$

DAN model (Figure 3. 4) learns layer by layer, and each layer acquires a more abstract representation of the input than its previous layer. This concept has been borrowed from the NBOV model. The concept is that each layer increasingly magnifies small but meaningful differences in the word embedding average. To give an illustration, consider a statement s_1 to be "I love cake" and create s_2 and s_3 by substituting 'love' with 'like' and then with 'hate.' The vector averages of the statements are roughly identical; however, the synonymous statements s_1 and s_2 tend to be more identical to each other than they are to the average of s_3 . It is clear from equation (3.4) that considering the input text X ; the vector representation o can be calculated by averaging the word vectors $u_{w \in X}$. This representation is not directly passed to an output layer; instead, o is transformed by adding more layers before applying softmax. Considering n layers, say $o_{1...n}$, each layer is calculated by equation (3.8).

$$o_i = h(o_{i-1}) = f(W_i \cdot o_{i-1} + b_i) \quad (3.8)$$

o_n is the final layer representation, and this is fed into the SoftMax layer for prediction (Figure 3. 4).

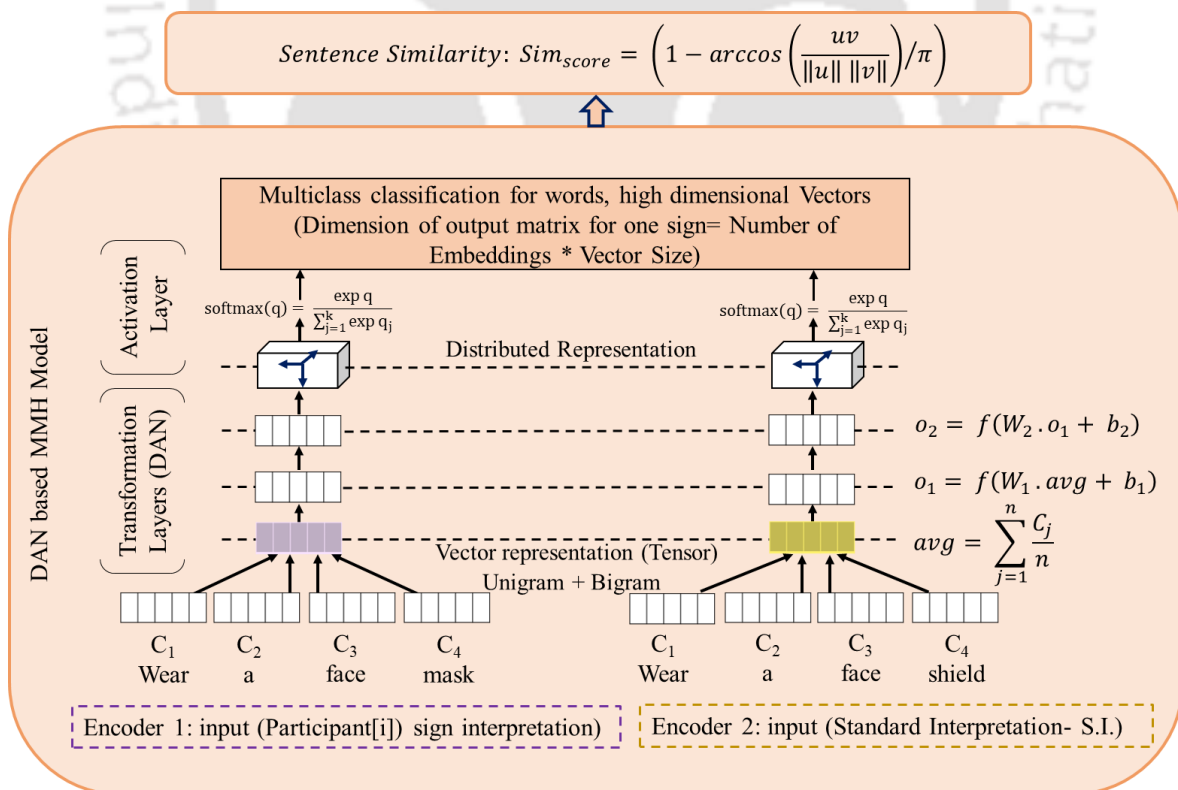


Figure 3. 4. DAN-based DNN

Phase 4

The OSH experts have chosen a value of 0.0 to 1.0, with an interval of 0.1 (11-point scale), for marking the answers (Darbyshire & McDonald, 2004; Kwong & Pathak, 2007; Lazaridou et al., 2018). 0.0 indicates that the sign is not at all comprehensible or understandable, and 1.0 indicates that the sign is fully comprehensible or understandable. The threshold for classification is based on the severity of the context, 50% fixed by the experienced experts. The binning threshold value is decided by experts based on the severity of the context or the level of danger or threats that can affect the OSH of the organization if the message is not understood correctly. The comprehensibility threshold for sign selection is considered based on the ISO standard (67%).

A new and novel approach based on the incomprehensibility (non-comprehensibility) of the OSH message (s) has been attempted here. The intention is to compute the Comprehensibility decision-making factor (CDM or NC_p , a term newly coined in this study, in terms of percentage as shown in equation (3.9)), where 'n' represents the total number of records/ participants, 'i' is the index of the record/ participant, 'SI.' is the standard interpretation whose value is always 1. The value of $Count_nc$ (which counts the total number of ($C_y = 0$); the variable for counting incomprehensibility in terms of how many participants found it incomprehensible) is derived from the algorithm COMPUTE COMPREHENSIBILITY. It is to be noted that the terms CDM and NC_p have been used synonymously in this study, and both terms signify the same.

$$CDM = \left(\frac{1}{n} \sum_{i=1}^{Count_nc} (SI - MS)^2 \right) \times 100 \quad (3.9)$$

Equation (3.9) can be rewritten as

$$CDM = \left(\frac{1}{n} \sum_{i=1}^{Count_nc} (SI - C_y)^2 \right) \times 100 \quad (3.10)$$

Considering $C_y = 0$, equation (3.10) can be written as

$$CDM = \left(\frac{1}{n} \sum_{i=1}^{Count_nc} (SI)^2 \right) \times 100 = \left(\frac{1}{n} \sum_{i=1}^{Count_nc} SI \right) \times 100. \quad (3.11)$$

Here $(SI)^2$ is equal to $SI \because$ the value of $SI = 1$.

The percentage of participants who successfully comprehended the message, C_p , is calculated by the equation (3.12)

$$C_p = 100 - CDM \quad (3.12)$$

The accuracy of the system or comprehensibility of OSH message(s) decreases as the CDM/ NC_p increases. The value must be less for equation 3.11 to achieve good comprehensibility. This equation decides whether the OSH sign's message is comprehensible and can go for the consequent post-design phase. C_p in equation (3.12)(3.12) is the reverse of equation (3.11). C_p directly measures the percentage of participants who successfully comprehended the message(s), and NC_p/CDM calculates the percentage of participants who failed to comprehend the message(s). The proposed technique implementing the existing DAN algorithm to develop such an assisting tool for automated evaluation is a new and novel contribution to the domain of OSH.

Once the similarity scores are generated, data is normalized as discussed above, and this whole procedure of transforming data from qualitative to quantitative binary categories (Zaini et al., 2022) introduced a novel data coding technique in this study.

Once data is generated, descriptive statistics are considered, and the normality of data (using Kolmogorov-Smirnov and Shapiro-Wilk) is checked for distribution before statistical tests and analysis. Analysis of variance (ANOVA) has been conducted for all the signs. Friedman's ANOVA test, followed by post hoc analysis, has been done to study variances between all the sign pairs viewed by the same group of participants. The post hoc test helps decide which two sign pairs are significantly different from each other.

Wilcoxon signed-rank test has been performed to check the effect of text on signs. Then, the effect size using the Z value has been calculated to confirm the result. The effect size is calculated using the following equation (3.13)

$$r = \frac{Z}{\sqrt{N}} \quad (3.13),$$

where Z is the z-score of the Wilcoxon test, and N is the size of the study (number of total observations). The effect size (Z score) is calculated for all seven (7) pairs of signs- with and without text. All the statistical analyses have been performed using IBM SPSS Statistics (Version 25.0).

Two thresholds have been used for evaluation. The 'Binning threshold' has been used to convert the 11-point rating scale to binary categories. 'Binning threshold' is introduced to tune the threshold at levels according to the context of the sign's use. Here 50% binning threshold has been considered by the experts in food processing industries where the severity is less compared to industries with more hazards. This is important to highlight that this Binning threshold is used for binning the 11-point rating scale into two categories. This 'Binning threshold' is different and has no relation with the ISO standard threshold (67%). 'Comprehensibility frequency' in this study measures the total number of respondents who correctly interpreted the sign. The comprehensibility percentage (C_p) is calculated by adding the total number of participants who understood the sign (Comprehensibility frequency), dividing by the total number of participants who participated in the experiment, and

multiplying by 100. The (C_p) equal to or exceeding 67% is considered comprehensible, and the remaining is incomprehensible.

3.3. Results and Discussions

The results of the computational technique and their various statistical analyses are discussed below.

3.3.1. Results for the computational model (Automated score generation)

The computational model considers the written response (OSH message) as input and generates the score (Figure 3.5), simulating human expert scoring in the comprehensibility evaluation process. The model uses a cloud-based environment known as ‘Colaboratory’ to generate the score. The DAN-based model combines the accuracy of syntactic functions with the speed of unordered functions (Iyyer et al., 2015) with linear time complexity $O(n)$ and constant space complexity (Cer, Yang, Kong, Hua, Limtiaco, John, et al., 2018). However, the DAN model requires more space than other models as it stores unigram and bigram embeddings. The model evaluates the sentence-level semantic similarity (similarity between the two sentences- standard interpretation (*SI*) and the participant’s interpretation (response) to understand the comprehensibility of the message in comparison to the standard interpretation) and generates the comprehensibility scores.

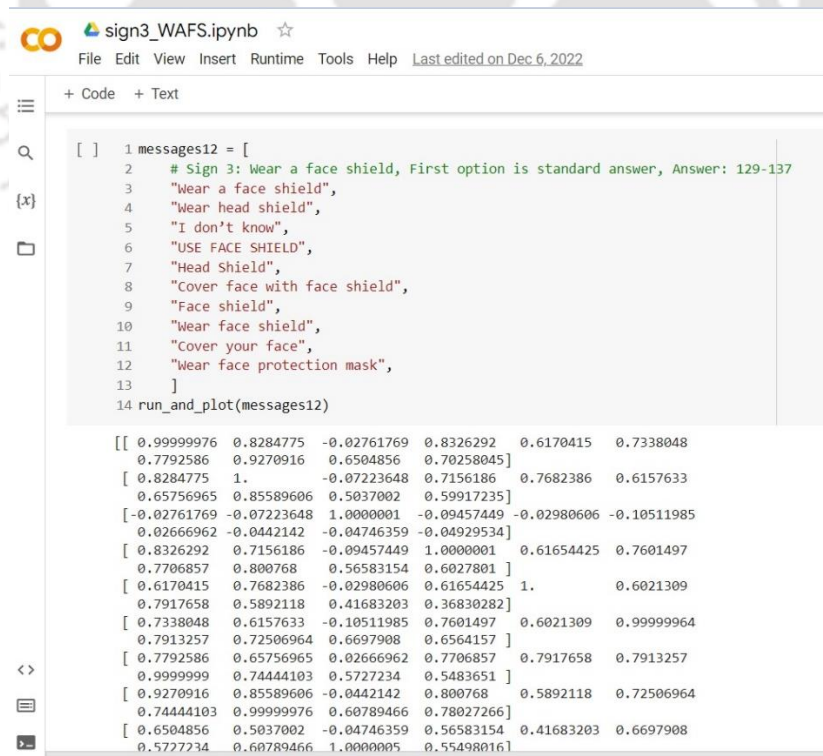


Figure 3. 5. Score generation by the computational DAN-based MMH model

One thousand nine hundred eighteen (1918) such scores have been generated for the study (137 responses for 14 signs). An instance of the scores being generated can be seen in Figure 3. 5. A heat map is generated where '1' (intense dark shade) represents a complete match between the standard interpretation (SI) of the OSH sign and the interpretation of the participants. A '0' (lightest shade) represents a mismatch, where the participants' responses did not match the SI. The shades get lighter as the mismatch between the sentences increases. An instance of the output of the twelve participants for sign 3 (WAFS) showing semantic textual similarity between the SI and their responses is shown in Figure 3. 6.

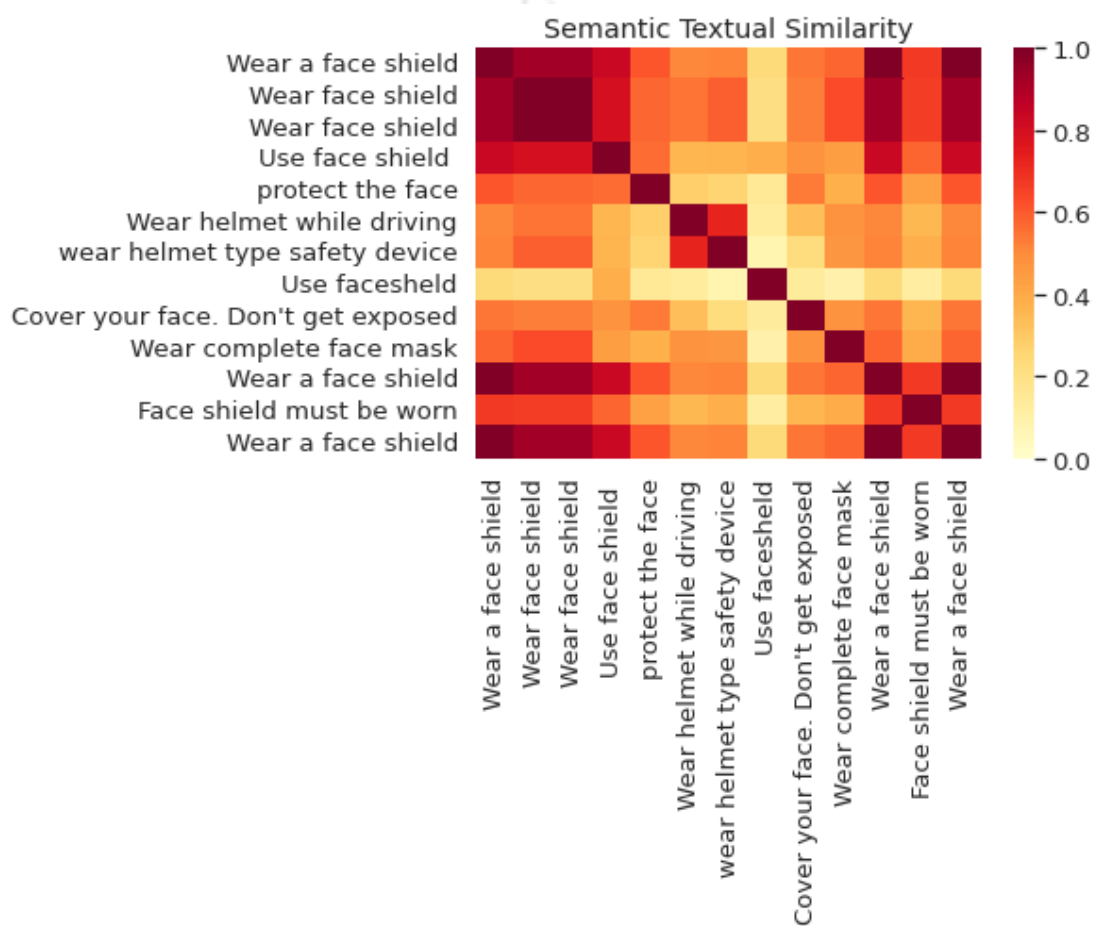


Figure 3. 6. Heat map showing semantic textual similarity generated by the computational technique

Fourteen (14) signs (7 pairs with and without text) recommended by the experts are considered for investigation, and each sign is associated with one hundred and thirty-seven (137) results divided into twelve (12) figures. Hence a total of one hundred and sixty-eight (168) figures/ heat maps have been generated for the results of fourteen (14) signs. A single sign (sign 1-DNWSH) is shown for brevity, combining all the heatmaps for all 137 participants (Figure 3. 7 and Figure 3. 8).

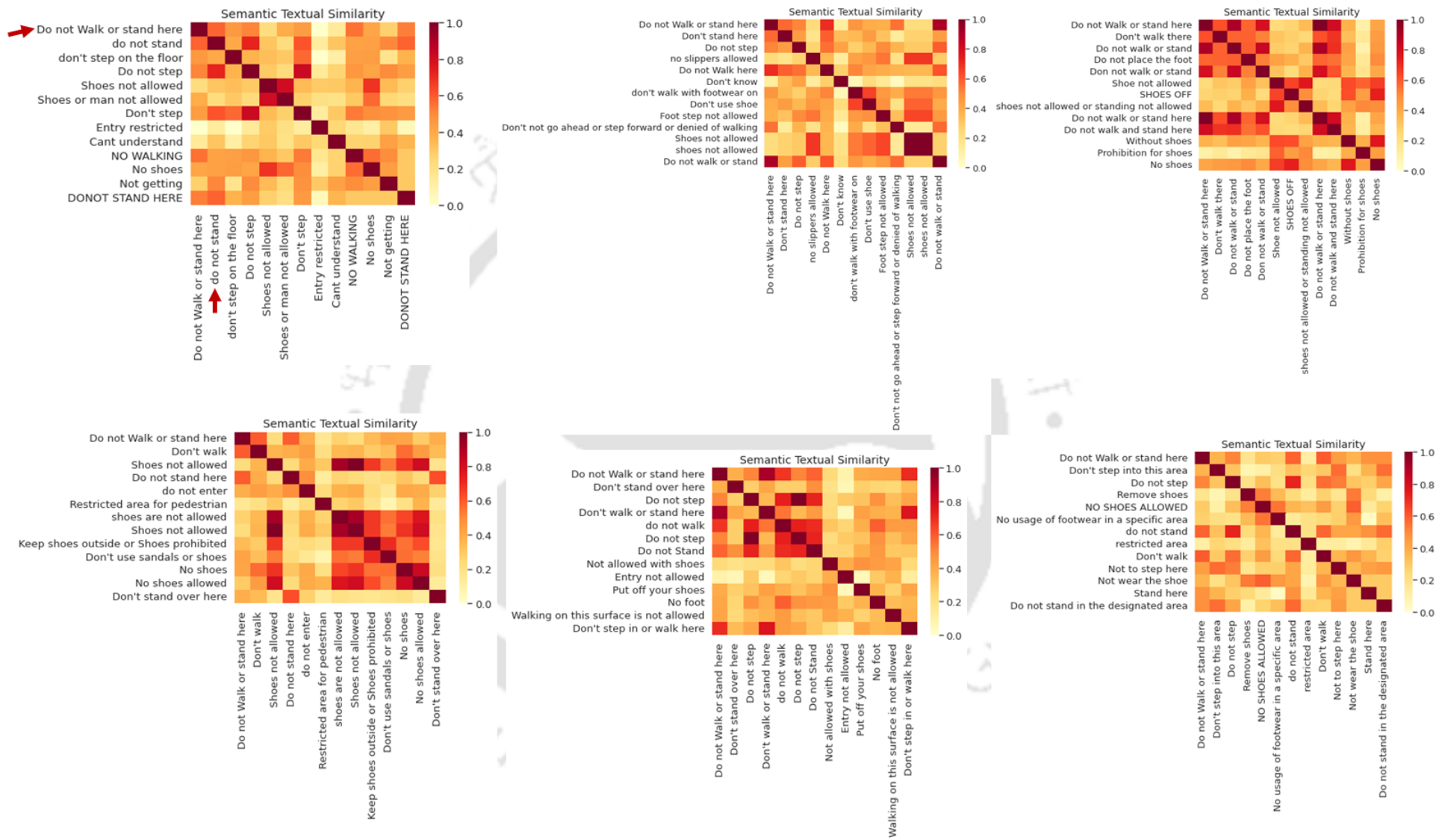


Figure 3. 7. Heat map generated for the first 72 participants (72/137) for Sign 1 (DNWSH)

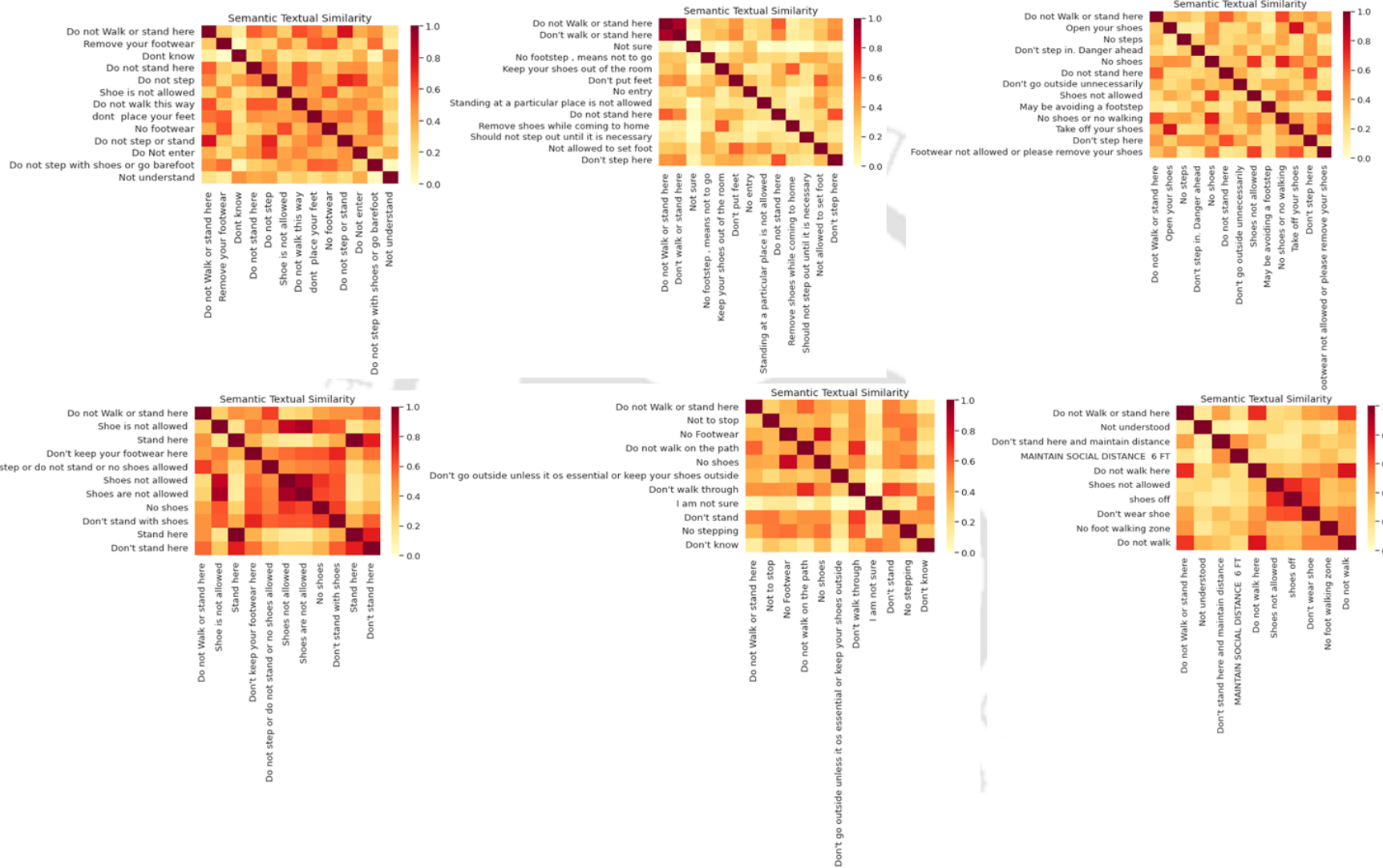


Figure 3. 8. Heat map generated for the remaining 65 participants (65/137) for Sign 1 (DNWSH)

Considering the standard interpretation “*Do not Walk or stand here*” and the subject’s written response “*do not stand*” (marked with red arrows in Figure 3. 7), both have a fair match in the embeddings and a cosine similarity of 0.5827668, which is rounded to approximately (0.6). This scoring method is similar to the way the experts assign a score, and upon validating the same with the expert score, it is found that the average score assigned by the experts is 0.7 which is in line with the model score (machine-computed). The Pearson correlation for the model computed score and the expert score for the whole set of 137 samples are found to be more than 0.8 for all 14 signs with a p-value less than 0.05. The result is significant, and this validates the model performance. Training with the STS benchmark helped in achieving this high semantic similarity. The automated system calculates the score based on the cosine similarity and hence is not error-prone like human beings who might score the responses based on heuristics. The data generated has been checked for its descriptives and distribution before conducting other analyses (statistical and computational).

3.3.2. Descriptive Statistics

Descriptives have been calculated using IBM SPSS 25.0 with 137 cases with no missing data (Table 3. 1 and Table 3. 2).

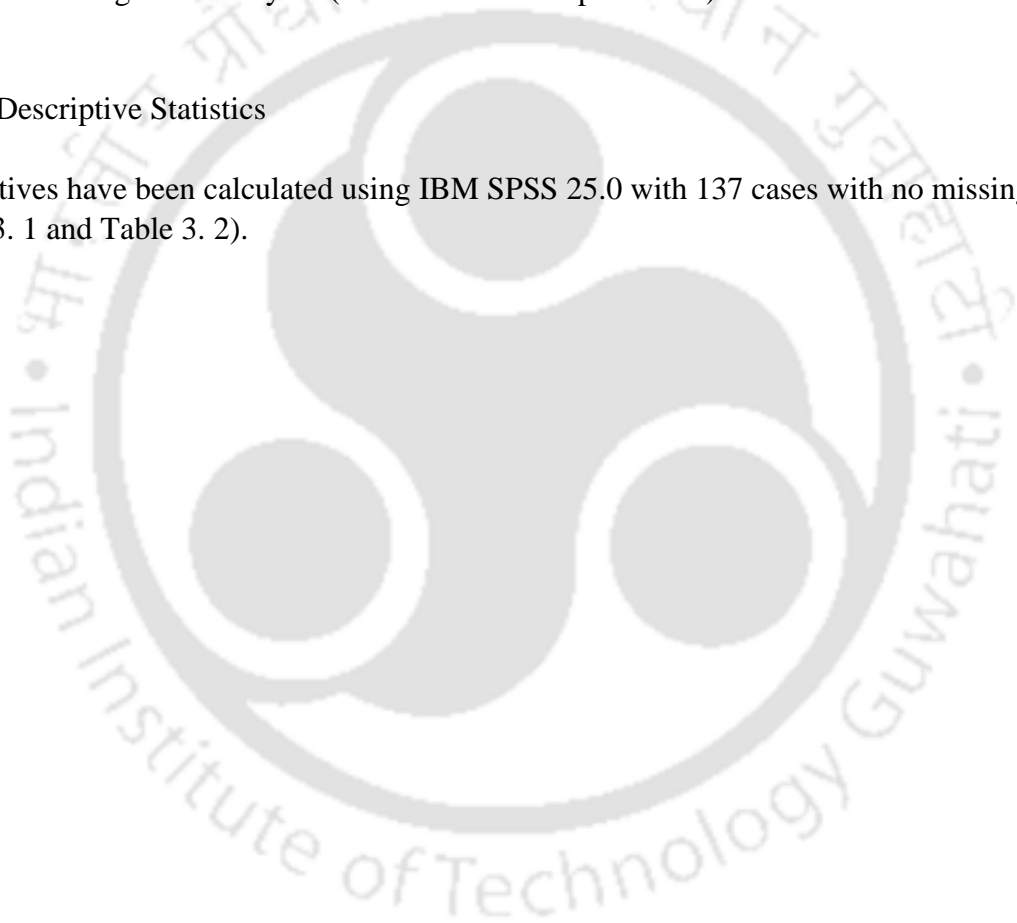


Table 3. 1. Descriptive statistics of computational technique (signs without text)

		NAFUP		UBC		DNT		DNWSH		COSACOEOT		WRP		WAFS	
		Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
Mean		0.1708	0.01359	0.2759	0.01321	0.5942	0.02600	0.4146	0.01701	0.1000	0.01827	0.3080	0.02020	0.6774	0.02085
95% Confidence Interval for Mean	L.B.	0.1439		0.2498		0.5427		0.3810		0.0639		0.2681		0.6361	
	U.B.	0.1977		0.3020		0.6456		0.4482		0.1361		0.3480		0.7186	
5% Trimmed Mean		0.1515		0.2676		0.5991		0.4066		0.0672		0.3008		0.6938	
Median		0.1000		0.3000		0.6000		0.4000		0.0000		0.4000		0.8000	
Variance		0.025		0.024		0.093		0.040		0.046		0.056		0.060	
Std. Deviation		0.15912		0.15460		0.30432		0.19909		0.21386		0.23641		0.24404	
Minimum		0.00		0.00		0.10		0.00		0.00		0.00		0.00	
Maximum		0.90		1.00		1.00		1.00		1.00		1.00		1.00	
Range		0.90		1.00		0.90		1.00		1.00		1.00		1.00	
Interquartile Range		0.10		0.10		0.60		0.20		0.10		0.40		0.20	
Skewness		2.402	0.207	1.635	0.207	-0.045	0.207	0.465	0.207	2.426	0.207	0.198	0.207	-1.227	0.207
Kurtosis		7.221	0.411	6.010	0.411	-1.436	0.411	0.076	0.411	5.441	0.411	-1.011	0.411	0.869	0.411

Here L.B.: Lower Bound; U.B.: Upper Bound

Table 3. 2. Descriptive statistics of computational technique (signs with text)

		DNWSH_T		NAFUP_T		WAFS_T		COSACOEOT_T		UBC_T		WRP_T		DNT_T	
		Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
Mean		0.6234	0.02306	0.4591	0.02806	0.8387	0.01649	0.4409	0.03132	0.5387	0.02881	0.5679	0.02895	0.7876	0.02404
95% Confidence Interval for Mean	L.B.	0.5778		0.4036		0.8061		0.3789		0.4817		0.5106		0.7400	
	U.B.	0.6690		0.5146		0.8713		0.5028		0.5957		0.6251		0.8351	
5% Trimmed Mean		0.6331		0.4546		0.8655		0.4343		0.5430		0.5754		0.8141	
Median		0.6000		0.4000		0.9000		0.5000		0.5000		0.6000		0.9000	
Variance		0.073		0.108		0.037		0.134		0.114		0.115		0.079	
Std. Deviation		0.26988		0.32844		0.19298		0.36653		0.33721		0.33889		0.28139	
Minimum		0.00		0.00		0.00		0.00		0.00		0.00		0.00	
Maximum		1.00		1.00		1.00		1.00		1.00		1.00		1.00	
Range		1.00		1.00		1.00		1.00		1.00		1.00		1.00	
Interquartile Range		0.50		0.65		0.20		0.70		0.65		0.50		0.40	
Skewness		-0.389	0.207	0.169	0.207	-2.277	0.207	-0.038	0.207	0.048	0.207	-0.280	0.207	-1.240	0.207
Kurtosis		-0.613	0.411	-1.307	0.411	6.091	0.411	-1.563	0.411	-1.229	0.411	-1.140	0.411	0.288	0.411

Here L.B.: Lower Bound; U.B.: Upper Bound

3.3.3. Test for Normality

Normality tests (Kolmogorov-Smirnov and Shapiro-Wilk in Table 3. 3) reveal that the data, i.e., comprehensibility scores of all the 14 signs, deviates from the normal distribution (p-value less than alpha (0.05)). The test fails to accept that sample data are normally distributed (Shapiro & Wilk, 1965). In the real world, it is acceptable to get non-normally distributed data.

Table 3. 3. Test for normality (computational technique)

Tests of Normality						
	Kolmogorov-Smirnov (a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
NAFUP	.314	137	.000	.705	137	.000
UBC	.241	137	.000	.846	137	.000
DNT	.171	137	.000	.901	137	.000
DNWSH	.135	137	.000	.960	137	.000
COSACOEOT	.410	137	.000	.543	137	.000
WRP	.176	137	.000	.910	137	.000
WAFS	.211	137	.000	.853	137	.000
DNWSH_T	.144	137	.000	.940	137	.000
NAFUP_T	.155	137	.000	.915	137	.000
WAFS_T	.253	137	.000	.730	137	.000
COSACOEOT_T	.199	137	.000	.853	137	.000
UBC_T	.155	137	.000	.907	137	.000
WRP_T	.136	137	.000	.903	137	.000
DNT_T	.298	137	.000	.760	137	.000

a. Lilliefors Significance Correction

3.3.4. Analysis of variance (ANOVA)

Analysis of variance (ANOVA) for the computational technique has been conducted for all the signs. The result of Friedman's ANOVA test (Table 3.4 and 3.5) followed by post hoc analysis has been done to study variances between all the sign pairs viewed by the same group of participants. Friedman's test calculated the p-value (Table 3.5), which is less than 0.05 and is significant here. The result shows significant differences in comprehensibility levels among all the signs considered for the study.

Table 3. 4. Rank of the signs generated by Friedman's ANOVA (computational technique)

Ranks	
Sign_Technique	Mean Rank
NAFUP_mac	3.66
UBC_mac	5.07
DNT_mac	9.12
DNWSH_mac	6.90
COSACOEOT_mac	2.65
WRP_mac	5.19
WAFS_mac	9.82
DNWSH_T_mac	9.08
NAFUP_T_mac	7.05
WAFS_T_mac	11.89
COSACOEOT_T_mac	6.63
UBC_T_mac	8.23
WRP_T_mac	8.31
DNT_T_mac	11.41

Table 3. 5. Test statistics of Friedman's ANOVA (computational technique)

Test Statistics(a)	
N	137
Chi-Square	795.234
df	13
Asymp. Sig.	.000

a. Friedman Test

A post hoc analysis (Figure 3. 9) following the Friedman test helps to identify the significant and non-significant pairs. The result (Figure 3. 10 - Figure 3. 15) shows that sixty-three (63) pairs are significantly different, and the pairs are highlighted in yellow (p-values are similar to the Adj. Sig.) in the figures.

Hypothesis Test Summary

Null Hypothesis	Test	Sig.	Decision
The distributions of NAFUP_mac, UBC_mac, DNT_mac, DNWSH_mac, COSACOEOT_mac, WRP_mac, WAFS_mac, DNWSH_T_mac, NAFUP_T_mac, WAFS_T_mac, COSACOEOT_T_mac, UBC_T_mac, WRP_T_mac and DNT_T_mac are the same.	Related-Samples Friedman's Two-Way Analysis of Variance by Ranks	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 3. 9. Post hoc analysis (computational technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
COSACOEOT_mac-NAFUP_mac	1.011	.505	2.000	.045	1.000
COSACOEOT_mac-UBC_mac	2.420	.505	4.787	.000	.000
COSACOEOT_mac-WRP_mac	-2.547	.505	-5.040	.000	.000
COSACOEOT_mac-COSACOEOT_T_mac	-3.982	.505	-7.878	.000	.000
COSACOEOT_mac-DNWSH_mac	4.252	.505	8.412	.000	.000
COSACOEOT_mac-NAFUP_T_mac	-4.401	.505	-8.708	.000	.000
COSACOEOT_mac-UBC_T_mac	-5.588	.505	-11.055	.000	.000
COSACOEOT_mac-WRP_T_mac	-5.661	.505	-11.199	.000	.000
COSACOEOT_mac-DNWSH_T_mac	-6.434	.505	-12.730	.000	.000
COSACOEOT_mac-DNT_mac	6.478	.505	12.817	.000	.000
COSACOEOT_mac-WAFS_mac	-7.175	.505	-14.196	.000	.000
COSACOEOT_mac-DNT_T_mac	-8.766	.505	-17.344	.000	.000
COSACOEOT_mac-WAFS_T_mac	-9.241	.505	-18.283	.000	.000
NAFUP_mac-UBC_mac	-1.409	.505	-2.787	.005	.484
NAFUP_mac-WRP_mac	-1.536	.505	-3.040	.002	.215

Figure 3. 10. Post hoc analysis result showing significantly different pairs in sequence-1 (computational technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
NAFUP_mac-COSACOEOT_T_mac	-2.971	.505	-5.878	.000	.000
NAFUP_mac-DNWSH_mac	-3.241	.505	-6.412	.000	.000
NAFUP_mac-NAFUP_T_mac	-3.391	.505	-6.708	.000	.000
NAFUP_mac-UBC_T_mac	-4.577	.505	-9.055	.000	.000
NAFUP_mac-WRP_T_mac	-4.650	.505	-9.199	.000	.000
NAFUP_mac-DNWSH_T_mac	-5.423	.505	-10.730	.000	.000
NAFUP_mac-DNT_mac	-5.467	.505	-10.817	.000	.000
NAFUP_mac-WAFS_mac	-6.164	.505	-12.196	.000	.000
NAFUP_mac-DNT_T_mac	-7.755	.505	-15.344	.000	.000
NAFUP_mac-WAFS_T_mac	-8.230	.505	-16.283	.000	.000
UBC_mac-WRP_mac	-.128	.505	-.253	.800	1.000
UBC_mac-COSACOEOT_T_mac	-1.562	.505	-3.090	.002	.182
UBC_mac-DNWSH_mac	-1.832	.505	-3.625	.000	.026
UBC_mac-NAFUP_T_mac	-1.982	.505	-3.921	.000	.008
UBC_mac-UBC_T_mac	-3.168	.505	-6.268	.000	.000
UBC_mac-WRP_T_mac	-3.241	.505	-6.412	.000	.000

Figure 3. 11. Post hoc analysis result showing significantly different pairs in sequence-2 (computational technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
UBC_mac-DNWSH_T_mac	-4.015	.505	-7.943	.000	.000
UBC_mac-DNT_mac	-4.058	.505	-8.029	.000	.000
UBC_mac-WAFS_mac	-4.755	.505	-9.408	.000	.000
UBC_mac-DNT_T_mac	-6.347	.505	-12.557	.000	.000
UBC_mac-WAFS_T_mac	-6.821	.505	-13.495	.000	.000
WRP_mac-COSACOEOT_T_mac	-1.434	.505	-2.838	.005	.413
WRP_mac-DNWSH_mac	1.704	.505	3.372	.001	.068
WRP_mac-NAFUP_T_mac	-1.854	.505	-3.668	.000	.022
WRP_mac-UBC_T_mac	-3.040	.505	-6.015	.000	.000
WRP_mac-WRP_T_mac	-3.113	.505	-6.159	.000	.000
WRP_mac-DNWSH_T_mac	-3.887	.505	-7.690	.000	.000
WRP_mac-DNT_mac	3.931	.505	7.777	.000	.000
WRP_mac-WAFS_mac	-4.628	.505	-9.156	.000	.000
WRP_mac-DNT_T_mac	-6.219	.505	-12.304	.000	.000
WRP_mac-WAFS_T_mac	-6.693	.505	-13.243	.000	.000
COSACOEOT_T_mac-DNWSH_mac	.270	.505	.534	.593	1.000

Figure 3.12. Post hoc analysis result showing significantly different pairs in sequence-3 (computational technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
COSACOEOT_T_mac-NAFUP_T_mac	.420	.505	.830	.406	1.000
COSACOEOT_T_mac-UBC_T_mac	-1.606	.505	-3.177	.001	.135
COSACOEOT_T_mac-WRP_T_mac	-1.679	.505	-3.321	.001	.081
COSACOEOT_T_mac-DNWSH_T_mac	2.453	.505	4.852	.000	.000
COSACOEOT_T_mac-DNT_mac	2.496	.505	4.939	.000	.000
COSACOEOT_T_mac-WAFS_mac	3.193	.505	6.318	.000	.000
COSACOEOT_T_mac-DNT_T_mac	-4.785	.505	-9.466	.000	.000
COSACOEOT_T_mac-WAFS_T_mac	5.259	.505	10.405	.000	.000
DNWSH_mac-NAFUP_T_mac	-.150	.505	-.296	.767	1.000
DNWSH_mac-UBC_T_mac	-1.336	.505	-2.643	.008	.748
DNWSH_mac-WRP_T_mac	-1.409	.505	-2.787	.005	.484
DNWSH_mac-DNWSH_T_mac	-2.182	.505	-4.318	.000	.001
DNWSH_mac-DNT_mac	2.226	.505	4.405	.000	.001
DNWSH_mac-WAFS_mac	-2.923	.505	-5.784	.000	.000
DNWSH_mac-DNT_T_mac	-4.515	.505	-8.932	.000	.000
DNWSH_mac-WAFS_T_mac	-4.989	.505	-9.871	.000	.000

Figure 3. 13. Post hoc analysis result showing significantly different pairs in sequence-4 (computational technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
NAFUP_T_mac-UBC_T_mac	-1.186	.505	-2.347	.019	1.000
NAFUP_T_mac-WRP_T_mac	-1.259	.505	-2.491	.013	1.000
NAFUP_T_mac-DNWSH_T_mac	2.033	.505	4.022	.000	.005
NAFUP_T_mac-DNT_mac	2.077	.505	4.109	.000	.004
NAFUP_T_mac-WAFS_mac	2.774	.505	5.488	.000	.000
NAFUP_T_mac-DNT_T_mac	-4.365	.505	-8.636	.000	.000
NAFUP_T_mac-WAFS_T_mac	-4.839	.505	-9.575	.000	.000
UBC_T_mac-WRP_T_mac	-.073	.505	-.144	.885	1.000
UBC_T_mac-DNWSH_T_mac	.847	.505	1.675	.094	1.000
UBC_T_mac-DNT_mac	.891	.505	1.762	.078	1.000
UBC_T_mac-WAFS_mac	1.588	.505	3.141	.002	.153
UBC_T_mac-DNT_T_mac	-3.179	.505	-6.289	.000	.000
UBC_T_mac-WAFS_T_mac	3.653	.505	7.228	.000	.000
WRP_T_mac-DNWSH_T_mac	.774	.505	1.531	.126	1.000
WRP_T_mac-DNT_mac	.818	.505	1.617	.106	1.000
WRP_T_mac-WAFS_mac	1.515	.505	2.997	.003	.248

Figure 3. 14. Post hoc analysis result showing significantly different pairs in sequence-5 (computational technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
WRP_T_mac-DNT_T_mac	-3.106	.505	-6.145	.000	.000
WRP_T_mac-WAFS_T_mac	3.580	.505	7.083	.000	.000
DNWSH_T_mac-DNT_mac	.044	.505	.087	.931	1.000
DNWSH_T_mac-WAFS_mac	.741	.505	1.466	.143	1.000
DNWSH_T_mac-DNT_T_mac	-2.332	.505	-4.614	.000	.000
DNWSH_T_mac-WAFS_T_mac	-2.807	.505	-5.553	.000	.000
DNT_mac-WAFS_mac	-.697	.505	-1.379	.168	1.000
DNT_mac-DNT_T_mac	-2.288	.505	-4.527	.000	.001
DNT_mac-WAFS_T_mac	-2.763	.505	-5.466	.000	.000
WAFS_mac-DNT_T_mac	-1.591	.505	-3.148	.002	.149
WAFS_mac-WAFS_T_mac	-2.066	.505	-4.087	.000	.004
DNT_T_mac-WAFS_T_mac	.474	.505	.939	.348	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.

Asymptotic significances (2-sided tests) are displayed. The significance level is .05. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Figure 3. 15. Post hoc analysis result showing significantly different pairs in sequence-6 (computational technique)

3.3.5. OSH sign with OSH-related textual information affects the understanding or comprehensibility of the messages intended to convey (computationally computed data)

Considering the hypothesis,

Null Hypothesis (H₀): The level of comprehensibility of the OSH sign is not significantly different when the sign is presented with or without textual information.

Alternate Hypothesis (H₃): The level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information.

To test the hypothesis, a Wilcoxon signed rank test has been conducted, and the result is shown in the following tables. Descriptive statistics (Table 3. 6) shows the mean of all the signs, indicating the significant difference between the signs with and without textual information. The mean of the signs with text is more than the mean of the signs without text. This difference is significant, and the result is the same as expected from this technique. The minimum and maximum values signify the minimum and maximum scores assigned for that sign under study.

Table 3. 6. Descriptive statistics for the computational-based comprehensibility score in the Wilcoxon Signed Rank test

Descriptive Statistics					
	N	Mean	Std. Deviation	Minimum	Maximum
NAFUP_mac	137	.1708	.15912	.00	.90
UBC_mac	137	.2759	.15460	.00	1.00
DNT_mac	137	.5942	.30432	.10	1.00
DNWSH_mac	137	.4146	.19909	.00	1.00
COSACOEOT_mac	137	.1000	.21386	.00	1.00
WRP_mac	137	.3080	.23641	.00	1.00
WAFS_mac	137	.6774	.24404	.00	1.00
NAFUP_T_mac	137	.4591	.32844	.00	1.00
UBC_T_mac	137	.5387	.33721	.00	1.00
DNT_T_mac	137	.7876	.28139	.00	1.00
DNWSH_T_mac	137	.6234	.26988	.00	1.00
COSACOEOT_T_mac	137	.4409	.36653	.00	1.00
WRP_T_mac	137	.5679	.33889	.00	1.00
WAFS_T_mac	137	.8387	.19298	.00	1.00

Note: i. _mac: denotes the computational technique. ii. _T: denotes textual information.

The hypothesis is confirmed further by considering ranks. It is found that the mean of the positive rank and the sum of the positive ranks (Table 3. 7) are more for the signs with textual information than the signs without textual information.

Table 3. 7. Wilcoxon Signed Rank test ranks for the sign pairs: with and without text (computational technique)

Ranks		N	Mean Rank	Sum of Ranks
NAFUP_T_mac - NAFUP_mac	Negative Ranks	15 ^a	18.20	273.00
	Positive Ranks	83 ^b	55.16	4578.00
	Ties	39 ^c		
	Total	137		
UBC_T_mac - UBC_mac	Negative Ranks	18 ^d	26.42	475.50
	Positive Ranks	90 ^e	60.12	5410.50
	Ties	29 ^f		
	Total	137		
DNT_T_mac - DNT_mac	Negative Ranks	13 ^g	23.50	305.50
	Positive Ranks	69 ^h	44.89	3097.50
	Ties	55 ⁱ		
	Total	137		
DNWSH_T_mac - DNWSH_mac	Negative Ranks	17 ^j	25.24	429.00
	Positive Ranks	81 ^k	54.59	4422.00
	Ties	39 ^l		
	Total	137		
COSACOEOT_T_mac - COSACOEOT_mac	Negative Ranks	7 ^m	16.36	114.50
	Positive Ranks	75 ⁿ	43.85	3288.50
	Ties	55 ^o		
	Total	137		
WRP_T_mac - WRP_mac	Negative Ranks	14 ^p	20.43	286.00
	Positive Ranks	74 ^q	49.05	3630.00
	Ties	49 ^r		
	Total	137		
WAFS_T_mac - WAFS_mac	Negative Ranks	9 ^s	21.50	193.50
	Positive Ranks	75 ^t	45.02	3376.50
	Ties	53 ^u		
	Total	137		

a. NAFUP_T_mac < NAFUP_mac

b. NAFUP_T_mac > NAFUP_mac

c. NAFUP_T_mac = NAFUP_mac

d. UBC_T_mac < UBC_mac

e. UBC_T_mac > UBC_mac

f. UBC_T_mac = UBC_mac

g. DNT_T_mac < DNT_mac

h. DNT_T_mac > DNT_mac

i. DNT_T_mac = DNT_mac

j. DNWSH_T_mac < DNWSH_mac

k. DNWSH_T_mac > DNWSH_mac

l. DNWSH_T_mac = DNWSH_mac

m. COSACOEOT_T_mac <

COSACOEOT_mac

n. COSACOEOT_T_mac >

COSACOEOT_mac

o. COSACOEOT_T_mac = COSACOEOT_mac

p. WRP_T_mac < WRP_mac

q. WRP_T_mac > WRP_mac

r. WRP_T_mac = WRP_mac

s. WAFS_T_mac < WAFS_mac

t. WAFS_T_mac > WAFS_mac

u. WAFS_T_mac = WAFS_mac

It is seen that the number of positive ranks for all the signs with text is found to be much more than for signs without text. For example, the number of positive ranks for the NAFUP_T (NAFUP_T_sub > NAFUP_sub) is found to be 83 (Table 3. 7), i.e., 83 cases the participants gave more scores to the NAFUP_T than the NAFUP. The table also shows 39 cases when the NAFUP_T and NAFUP have been ranked similarly (tie).

The machine or the computational technique scored unambiguously in this technique. Hence the scores are bias-free, unlike the previous techniques, and they are consistent. The Z value (Table 3. 8) helped to calculate the effect size (r-value) in Table 3.9, which assists in checking how strong the effect of the text is on the comprehensibility of the sign(s). The calculated effect size is significant for all the pairs, which is excellent for our study.

Table 3. 8. Test statistics for Wilcoxon Signed Rank test for the sign pairs: with and without text (computational technique)

Test Statistics ^(a)							
	NAFUP_T_sub - NAFUP_sub	UBC_T_sub - UBC_sub	DNT_T_sub - DNT_sub	DNWSH_T_sub - DNWSH_sub	COSACOEOT_T_sub - COSACOEOT_sub	WRP_T_sub - WRP_sub	WAFS_T_sub - WAFS_sub
Z	-7.643b	-7.584b	-6.472b	-7.096b	-7.356b	-6.973b	-7.141b
Asymp. Sig.	0.000	0.000	0.000	0.000	0.000	0.000	0.000

a. Wilcoxon Signed Ranks Test, b. Based on positive ranks, c. Based on negative ranks.

Table 3. 9. Effect Size for Wilcoxon Signed Rank test for the sign pairs: with and without text (computational technique)

	COSACOEOT_T_score - COSACOEOT_score	DNT_T_score - DNT_score	DNWSH_T_score - DNWSH_score	NAFUP_T_score - NAFUP_score	UBC_T_score - UBC_score	WAFS_T_score - WAFS_score	WRP_T_score - WRP_score
Z	-7.356	-6.472	-7.096	-7.643	-7.584	-7.141	-6.973
N	274	274	274	274	274	274	274
Root N	16.55294536	16.55294536	16.55294536	16.55294536	16.55294536	16.55294536	16.55294536
r	-0.444392212	-0.390987819	-0.428685037	-0.461730516	-0.458166196	-0.431403587	-0.421254336

Hence the interpretation of the result for all the sign pairs can be as follows:

- ii. A Wilcoxon signed rank test and the calculated effect size reveal that for all the seven sign pairs, textual information positively affects the comprehensibility of message/ sign evident from their mean difference, z value, p-value, and r value (Table 3. 6 - Table 3. 9).
- iii. Hence null hypothesis can be rejected, and the alternate hypothesis has been established that the level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information.

3.3.6. Comprehensibility percentage based on computationally generated data

The comprehensibility percentage for each sign is calculated based on the score assigned by the automated computational model (Table 3.10). The 11-point scale has been converted into two categories consulting the experts. The calculation is provided in section 2.2.3. The number of comprehensible signs is added, divided by the total participants, and multiplied by 100 to get the total Percentage of comprehensibility. The sign whose total comprehensibility percentage is more than or equal to 67% (Berrio et al., 2022; ISO 3864-1, 2011) is considered acceptable. This means that a minimum of 67% target population found the signs comprehensible. Besides the ANOVA and hypothesis test (sections 3.3.4. and 3.3.5.), this comprehensibility percentage reveals that signs with text are more comprehensible than signs without text.

By viewing the Comprehensibility Percentage (Table 3. 10), it is possible to determine which sign is comprehensible and which is not, and they can be arranged according to the increasing/ decreasing order of comprehensibility. ISO 3864 standardizes and claims the signs as acceptable when 67% of the target population can understand or comprehend the sign in a comprehension test (Berrio et al., 2022; ISO 3864-1, 2011), and this has been followed in this study. The study reveals that COSACOEOT, COSACOEOT_T, DNT, DNWSH, NAFUP, NAFUP_T, UBC, UBC_T, WRP, and WRP_T are incomprehensible signs according to the ISO standard, and they need redesign. These are the same signs also found to be incomprehensible in the expert-based technique. The other signs, DNT_T, DNWSH_T, WAFS, and WAFS_T, are comprehensible same as in the expert-based technique, and can be considered for use. The signs in the increasing order of comprehensibility are as follows:

NAFUP < COSACOEOT < UBC < WRP < DNWSH < NAFUP_T < COSACOEOT_T < UBC_T < DNT < WRP_T < DNWSH_T < DNT_T < WAFS < WAFS_T.

The above sequence denotes that the sign NAFUP is less comprehensible than COSACOEOT which in turn is less comprehensible than UBC and the comparison goes on. WAFS_T is the most comprehensible sign found in this study (Table 3.10).

Table 3. 10. Comprehensibility percentage (computational technique)

Sign label	Comprehensibility Percentage [%]	Sign label	Comprehensibility Percentage [%]
COSACOEOT	9.489051095	COSACOEOT_T	56.20437956
DNT	62.04379562	DNT_T	83.21167883
DNWSH	39.41605839	DNWSH_T	74.45255474
NAFUP	5.839416058	NAFUP_T	48.90510949
UBC	11.67883212	UBC_T	56.20437956

WAFS	86.13138686	WAFS_T	94.16058394
WRP	32.84671533	WRP_T	63.50364964

3.3.7. Tuning of Threshold and CDM (NC_p)

Experts have fixed the 'Binning Threshold'. 'Binning threshold' is introduced to tune the threshold at levels according to the context of the sign's use. The result has been displayed in Table 3.10. The binning threshold can be tuned or adjusted to any other values (1%-100%) depending upon the context of uses, the severity of the danger/ warning, or the importance of the message to be conveyed. The result of all such tuning is out of the scope.

Similarly, the equation (3.11) proposed here assists in calculating the comprehensibility/ incomprehensibility automatically by using the CDM/ NC_p and the algorithm COMPUTE COMPREHENSIBILITY. The result of the comprehensibility decision using ' NC_p ' is shown in Table 3.11. It is observed that the lesser the value of CDM, the better the sign's comprehensibility. WAFS_T is the most comprehensible sign with CDM/ ' NC_p ' value 5.839416058. The least comprehensible sign NAFUP has the maximum ' NC_p ' or CDM value (94.16058394) according to the ISO standard (67% cutoff).

Table 3. 11. Comprehensibility decision (CDM) based on NC_p

Sign label	Incomprehensibility Percentage [%] based on CDM / NC_p	Sign label	Incomprehensibility Percentage [%] based on CDM/ NC_p
COSACOEOT	90.51094891	COSACOEOT_T	43.79562044
DNT	37.95620438	DNT_T	16.78832117
DNWSH	60.58394161	DNWSH_T	25.547445
NAFUP	94.16058394	NAFUP_T	51.09489051
UBC	88.32116788	UBC_T	43.79562044
WAFS	13.86861314	WAFS_T	5.839416058
WRP	67.15328467	WRP_T	36.49635036

3.3.8. Model Flexibility

The model has been tested for its flexibility in scoring when the inputs (sentences) are provided in uppercase letters, lowercase, and their combination. This has been possible as the model has been trained with vast and varied datasets besides the STS dataset. The automated DAN-based MMH model seems to be flexible enough to assign scores consistently irrespective of uppercase or lowercase letters or their combination, as shown in Figure 3. 16. It can be seen in the figure that the subjects' responses "DO NOT TOUCH", and "Do NoT TouCh", generates scores 0.99999994 and 1 which are almost similar. They have the same-colored box in the heatmap, thus showing the model's flexibility. Hence the respondents are free to write in any case (uppercase or lowercase texts) they want, and the data needs no further conversion (upper to lower or vice-versa) before feeding it to the automated model to get the score and, finally, the category of comprehensibility.

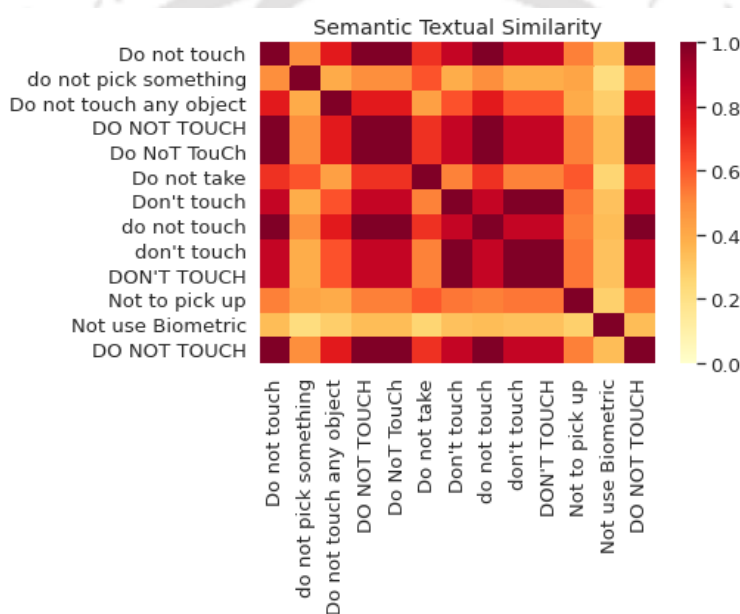


Figure 3. 16. Result reflecting upper- and lower-case flexibility of the model

3.3.9. Robustness of the Model

The model is robust enough to handle test data, different from the training data. The training data is from various sources of the web, like web question-answer pages, web news, discussion forums, Wikipedia (unsupervised data), and Standard Natural Language Inference (SNLI) corpus (supervised data). Data augmentation helped to avoid any overfitting during model training. The test data is the OSH message. The result is acceptable, considering the quality of the automated model's similarity scores against human judgments, which is encouraging (Pearson's score is more than 0.80 for all the signs with a p-value less than 0.05, which shows the result's significance, Figure 3.17). This fact validates the performance, technique, and robustness of the model.

```

File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
exec(code_obj, self.user_global_ns, self.user_ns)
Evaluate Sentence Embeddings
1 sts_data = sts_dev #@param ["sts_dev", "sts_test"] {type:"raw"}
2
3 def run_sts_benchmark(batch):
4     sts_encode1 = tf.nn.l2_normalize(embed(tf.constant(batch['sent_1'].tolist())), axis
5     sts_encode2 = tf.nn.l2_normalize(embed(tf.constant(batch['sent_2'].tolist())), axis
6     cosine_similarities = tf.reduce_sum(tf.multiply(sts_encode1, sts_encode2), axis=1)
7     clip_cosine_similarities = tf.clip_by_value(cosine_similarities, -1.0, 1.0)
8     scores = 1.0 - tf.acos(clip_cosine_similarities) / math.pi
9     """Returns the similarity scores for the human expert and the computational techni
10    return scores
11
12 dev_scores = sts_data['sim'].tolist()
13 scores = []
14 for batch in np.array_split(sts_data, 10):
15     scores.extend(run_sts_benchmark(batch))
16
17 pearson_correlation = scipy.stats.pearsonr(scores, dev_scores)
18 print('Pearson correlation coefficient = {0}\np-value = {1}'.format(
19     pearson_correlation[0], pearson_correlation[1]))
sts_data: sts_dev
Pearson correlation coefficient = 0.8036389324177564
p-value = 0.0

```

Figure 3. 17. Pearson correlation coefficient for the automated model (computational technique)

3.4. Conclusion

A novel technique of automated evaluation of the comprehensibility level of OSH signs has been proposed and developed here. This model addresses the issue concerning the repeated task of the OSH experts/ sign designers in manually evaluating the questionnaires responded to by the participants for their comprehensibility of the OSH sign. The existing manual technique for evaluating surveys/ questionnaires demands the brains of OSH designers/ experts for a long, which may lead to errors in comprehensibility evaluation and scoring and biased decision-making. A wrong decision may result in using imperfect signs for organizations whose ambiguous interpretation might become a catastrophe for the employees resulting in wrong or inappropriate behavior. Hence, an automated model (DAN-based MMH) has been proposed here to eliminate such inconvenience. This model can assist sign designers/ experts in evaluating the comprehensibility of OSH signs (which are newly designed or existing or in their early developmental phase) by assessing the subjective and descriptive responses of the participants with the application of deep learning techniques and DNN. Evaluation of comprehensibility of the OSH sign is still a manual procedure where the proposed system recommends maneuvering to the automated technique. Tuning of threshold (a thresholding function) has been proposed, which makes the system flexible. The proposed system can be tuned and deployed in any organization with diverse contexts of severity, and there lies the novelty of creating such a system. However, the present study considers only a only a 50% threshold depending upon the context of the study. A scoring function is developed, which scores the textual responses consistently. This standardizes the

comprehensibility evaluation, thus eliminating the errors, ambiguity, and biases in expert-based techniques.

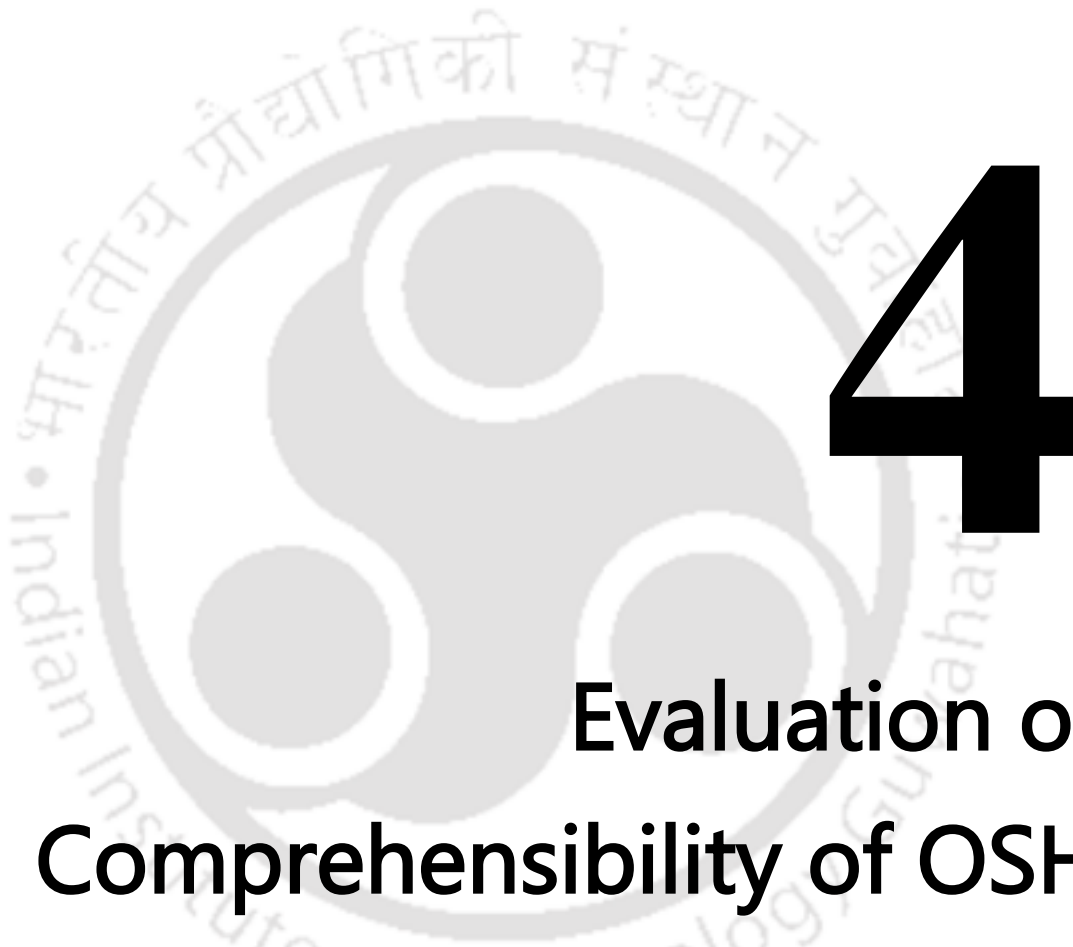
A 50% Binning Threshold has been introduced depending on the context of the study. A scoring function is developed, which scores the textual responses consistently. This standardizes the comprehensibility evaluation, thus eliminating the errors, ambiguity, and biases in expert-based techniques.

This model, created with the collaboration of human experts, makes it flexible to adapt to work in various contexts and environments. The model is trained with the standard databases. The present study portrays a new strategy for evaluating the comprehensibility of a sign's message and validates the existing pre-trained model with this new OSH dataset. For testing the model, representative data are collected from different states of India. The data generated are pre-processed and checked for reliability. Finally, the generated scores are checked for descriptives and normality before statistical analyses. The study found significant differences in comprehensibility levels among the signs under study. The study also found that OSH sign with OSH-related textual information affects the understanding or comprehensibility of the messages intended to be conveyed.

The contribution of this work lies in the computational technique, which has been designed and developed; that supports the evaluation of the comprehensibility of OSH sign(s) automatically and assists in the decision-making process. The experts can use the scores assigned by the model to check the subjects' (participants') views about the signs and work on them. This way, this technique can reduce the amount of work and time the experts spend on menial tasks and thus help firms leverage extra human resources on other unavoidable duties where real intelligent expertise is needed. A new term for comprehension analysis has been coined, such as Comprehensibility decision making (*CDM*), which acts as the decision-making factor for good design (in terms of comprehensibility). This DAN-MMH model trained on the STS benchmark is developed based on expert knowledge. It is capable of generating consistent and unbiased results. This developed model requires minimum human intervention for their decision-making regarding the comprehensibility of a sign. This model hence can act as an assistive tool for the human experts designing the OSH signs and symbols. The system's efficiency is compared with the experts' judgment (Pearson correlation r), which is greater than 80%, with a p-value less than 0.05. This result is significant to ensure the model's performance when it simulates human expert judgment.

The model is tuned for a 50% binning threshold for the present context. This computational technique-based evaluation found DNWSH_T, DNT_T, WAFS, and WAFS_T as the comprehensible signs, with WAFS_T as the most comprehensible, a similar result obtained in the previous two techniques. The rest of the signs may be considered for a redesign process.

The future work lies in fine-tuning the model, which has been out of the scope at this stage. The study is limited to an evaluation technique based on comprehensibility and does not consider design elements of the OSH sign under study. The model can be extended to assess the comprehensibility of other OSH-related communication media, such as posters and banners.



4

Evaluation of Comprehensibility of OSH Signs by Eye-tracking Technique

Chapter 4: Evaluation of Comprehensibility of OSH Signs by Eye-tracking Technique

Abstract

This study addresses the subjective biases of participants (target users), which may be present in subject-based, expert-based, and computational techniques. The intention is to develop an objective measurement paradigm that assesses the relationship between subjective measures (comprehensibility score) and the objective measures gained from the eye-tracking experiment related to OSH signs and symbols. An experiment has been conducted to find the effect of comprehensibility of OSH-related signs on eye movements. Two experimental setups have been designed. One is for the eye-tracking experiment, and the other is for the questionnaire-based study. The study is mixed-mode research. Descriptive statistics and the normality of data have been checked for both methods. For subjective data, statistical tests have been done to study significant differences in comprehensibility levels among signs and whether OSH signs with OSH-related textual information affect the understanding or comprehensibility of the messages. The correlation between the subjective comprehensibility score and the significant objective measures is studied. Then the variability of signs across categories of data (comprehensible and incomprehensible) is also studied. All the statistical analyses have been conducted using IBM SPSS 25.0. The result from the correlation study encourages including eye-tracking technology as one of the evaluation strategies for the comprehensibility of OSH sign(s).

4.1. Introduction

“*Comprehension is an ability to understand the meaning of a concept or an action based on the intelligent power of abstract thought and reasoning*” (Wang & Gafurov, 2003). Studying signs and symbols related to occupational safety and health(OSH) is an attempt to improve the health and safety of the workplace in an environment (Ahmed et al., 2023) where other methods of controlling hazard(s) are infeasible (Lehto, 1992). The sign is designed to convey the desired message unambiguously, which does not always happen in reality (Reis et al., 2015). This miscommunication brings unwanted accidents or health issues that become detrimental to employees of an organization(s). The study of comprehensibility or understandability of a sign is crucial considering the issues mentioned above where misinterpretation of any sign/ symbol may result in disaster or acute health hazard. The sign intends to communicate the message of the hazard appropriately; this could be ensured through a standard evaluation method that confirms the comprehensibility of the existing and designed sign(s).

The assessment where comprehensibility scores are generated manually or automatically from participants' written guesses based on subjective data can be biased due to many factors like their mental health, stress, previous exposures to such conditions and the signs, and many other unknown factors. So there emanates the idea of objective data collection through unobtrusive research where the subjective biases of humans can be avoided. For this, there must be some relation between the subjective and objective measures. *Objective measures* are measurements made with diagnostic instruments, including eye tracking devices, electroencephalograms (EEG), etc. (Cholewa et al., 2018; Siswandari et al., 2014). The

existence of the eye-mind relationship made it possible to use eye movement measurements to analyze something about human behavior, in this case, the comprehensibility of an OSH sign. The present research uses an eye-tracking device to capture the objective data from the participants. There are numerous eye-tracking experiments conducted and can be found in the literature for studying cognitive activities during learning, the process of learning, the cognitive process during reading, scene perception, music reading, visual perception, reading comprehension, human-computer interaction, media communication, strategy for problem-solving, multimedia learning, etc. (Lai et al., 2013). A recent eye-tracking survey reporting eye-tracking study in various aspects such as instrument, environment, participant, methodology, etc., and how they affect the quality of eye-tracking data has been found (Holmqvist et al., 2023). However, there is a dearth of literature studying the evaluation of the comprehensibility of OSH signs using eye-tracking techniques to find significant eye-tracking measures of comprehensibility. There is no report of how this technique is more advantageous or disadvantageous than any other in studying comprehension evaluation of OSH signs. Hence a study has been designed where subjective and objective data have been collected to find a relation between them.

Literature reveals that pupil dilation is a reliable and valid psychophysiological marker of attentional allocation (Adachi et al., 2014; Kahneman, 1973; Scharinger et al., 2020). Similarly, blink rate, time to first fixation (TTFF) (Siswandari & Xiong, 2015), and saccade (Krieger et al., 2016) are significant factors. Hence, these measures have been studied here to assess a sign's comprehensibility from the perspective of eye-tracking measures.

A hypothesis has been proposed as follows:

H₂: There is a significant relationship between the subjective response of the comprehensibility and the objective measures (eye tracking variables: average pupil diameter, visual intake count, visual intake duration, blink count, blink duration, blink frequency, saccade count, saccade amplitude total, saccade amplitude average, saccade duration total, scan path length).

4.2. Methodology and Experimental Setup

This part discusses the whole methodology and the experimental setup used for the eye-tracking technique of comprehensibility evaluation. The methodology consists of two separate experiments discussed in the following methods.

4.2.1. Stimuli

For both subjective and eye-tracking experiments, signs embedded with and without written text (Peckham, 2017) have been considered for the evaluation. Standard fourteen OSH-related signs (three prohibition sign pairs- with and without supplementary text and four mandatory sign pairs- with and without supplementary text, Figure 2.1) have been used as test material/ stimuli for the experiment. Each image is colored, and the approximate size is 80 mm × 80 mm. The size of the text in the sign is kept at twenty-two points (22 pt), and the

font style is Calibri (Body). The viewing distance is between 400 mm and 700 mm (Caffaro & Cavallo, 2015; Shieh & Lee, 2007) for collecting subjective data and 960 cm for objective data. Information related to the context plays a vital role in comprehending a pictogram (Lehto, 1992), and it seems that comprehension improves when contextual information is provided (Galer, 1980; Vukelich & Whitaker, 1993). The context for the study is described in words in written format and provided earlier before the experiment. A separate instruction page has been provided describing the purpose of the study. The experimental procedures with examples different from the proposed research have been recorded (audio-video recording for 4 mins and 58 secs) in a Microsoft PowerPoint. They were provided to the participants and were instructed to view them before experimenting.

4.2.2. Participants

The novices who are not from pure industry backgrounds were considered for this study. The motive behind such sample selection is that if the novice users can understand and comprehend any sign unknown or less known to them (semantic transparency), then the people acquainted with such, or related signs can easily comprehend the message conveyed by the sign. *Semantic transparency* is a term that involves novice users in the evaluation and design phase so that the designed sign is understandable or comprehensible by ordinary people (Caire et al., 2013; Polančič et al., 2020).



Figure 4. 1. Experimental setup: Test for eye defect using Snell's chart

Forty-four (N=44) participants from the Indian Institute of Technology Guwahati, India, participated in the experiment, out of which forty-one (N=41) participants' data was taken. Three data were corrupted and not considered for the study. The participants volunteered for

the study with no history of color deficiency, severe eye conditions, or visual impairment (self-reported). They are selected based on purposive sampling with a mean age of 29.23 and a standard deviation of 5.62. They do not have any eye defect (tested by Snell's chart) (Figure 4. 1) or color blindness (checked using Ishihara Test for Colour Blindness) (Figure 4. 2). The participants were asked to view a five-minute video showing the experimental protocol with an example of a trial with signs of different contexts, which in no way matched with the original sign. Informed consent was obtained from all the participants before the beginning of the trial. The experiment has been conducted following ethical guidelines and principles.



Figure 4. 2. Ishihara Test for color blindness

4.2.3. Research Instrument and Research Design

The objective of the study is to evaluate the comprehensibility of OSH signs using the eye tracking device so that the comprehensibility results truly reflect the opinion of the target audience. The present research is a within-subjects study. A six-month interval has been considered to minimize the carryover effect between the two experiments. Besides, the stimuli presentation is randomized to minimize the carryover effect. Two methods have been used for data collection.

4.2.3.1. Method1 (subject-based study)

In Method 1, the subjects were asked to answer demographic questions with the main study-related questions. The subjects were asked to assign a score/ rate about their comprehension of the sign. The score ranged between 0.0 and 1.0, with an interval of 0.1. 0.0 signifies “not comprehensible,” 0.5 is “partially comprehensible,” and 1.0 is “fully comprehensible.” The expert considers a 0.5 scale for converting the 11-point rating scale to the binary category, and finally, the comprehensibility percentage is calculated based on the ISO standard (67%). This data is subjective and is based on human perception. All the participants gave consent before facing the experiment. Before scribing their responses, they are informed to look at each sign for a minimum duration of 160 to 200 milliseconds (Caplette et al., 2020). The data collected from this study is further analyzed. Friedman's ANOVA test, followed by post hoc analysis, has been done to study variances between all the signs viewed by the same group of

participants. The non-parametric Friedman's ANOVA considers the actual data, sorts the data in ranks, compares the median, and calculates the p-value. Friedman calculates overall statistical differences (p values) for all groups together and not as pairs. Another test conducted following the Friedman test is the post hoc test. The post hoc test helps decide which two sign pairs are significantly different from each other. Further analysis on subject-based data is done using the Wilcoxon signed-rank test followed by post hoc analysis to check whether signs with text affect the participants' comprehensibility for the sign under study.

4.2.3.2. Method 2 (eye-tracking study)

The second experiment using an eye-tracking device was designed after six months. The eye-tracking experiment uses a device based on sensor-based technology, and the data collected are objective, involuntary, and natural. Each time the experiment is conducted, the stimuli have been shuffled using VBA (Visual Basic for Applications) programming. During the experiment, the participant wears the eye tracking glass and sits in front of the big monitor where the stimuli are displayed (Figure 4. 3). The chin rests at about 110 cm, approximately from the ground level. The hand rest table height is approximately 79.5 cm (Figure 4. 4). The distance between the big screen and the eye is approximately 96 cm (viewing distance of the stimuli) (Figure 4. 5). One point calibration has been done each time before every experiment with each participant so that their accurate eye movements and gaze points can be captured. Figure 4. 6. shows the sequence of the calibration stages. Each sign appears and lasts 10 seconds (Babić et al., 2020; Siswandari & Xiong, 2015). A dot sign has been kept (for 3 seconds) after and before each sign to minimize the previous effect. During the 10 seconds of sign comprehension, the physiological data of eye movements have been captured with an SMI eye tracking device. The details about the eye-tracking device are provided in Appendix A.



Figure 4. 3. Participant in front of a big screen wearing an eye-tracking glass

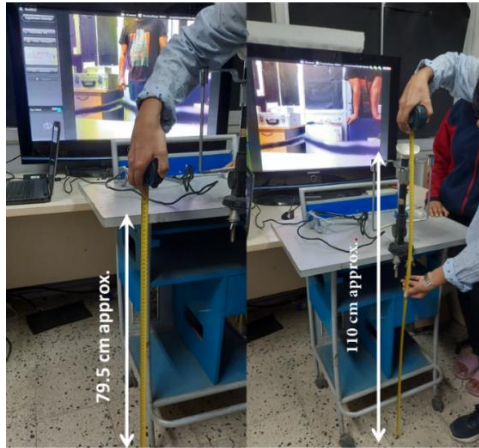


Figure 4. 4. Chin and hand rest in an eye-tracking set-up



Figure 4. 5. Viewing distance of the stimuli (eye-tracking setup)

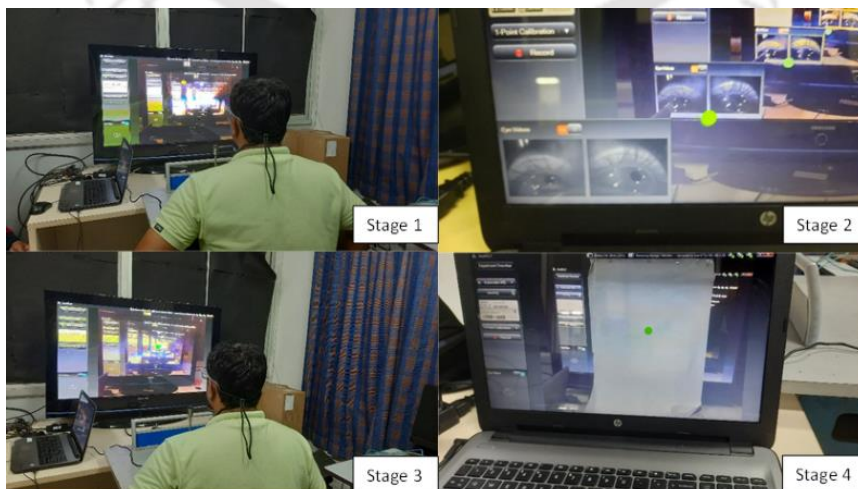


Figure 4. 6. One point calibration in an eye-tracking experiment

The eye-tracking technique involves:

i. Experimenting with eye tracking set-up

The set-up (Figure 4. 7) consists of an eye-tracking glass (ETG) and SMI BeGaze™ eye-tracking analysis software.



Figure 4. 7. Eye tracking set-up box

ii. Data extraction using eye-tracking software

The Areas of Interest (AOI) selected are fixed and static for all the similar signs for all the subjects. There is a variation in AOI for the signs with text and without text, as seen in figures (Figure 4. 8 and Figure 4. 9). For each sign, 10-second gaze data has been extracted. AOI 004 is selected as the external boundary box for the sign under study. AOI 005 is the exact boundary for the sign, and AOI 006 is the boundary for the OSH-related text of the sign (represented by dots in Figure 4.9).

iii. Data formatting and analysis.

For every fourteen (14) signs and forty-one (41) participants in this experiment, forty-six (46) eye-tracking variables were processed, which resulted in 26404 data ($14 \times 46 \times 41$). Descriptive statistics have been conducted, and data checked for its distribution. Subjective data is coded as 0.0 to 1.0 with an interval of 0.1 (11-point scale, ratio data). Normality tests (Kolmogorov-Smirnov and Shapiro-Wilk) have been conducted for subjective and objective data.

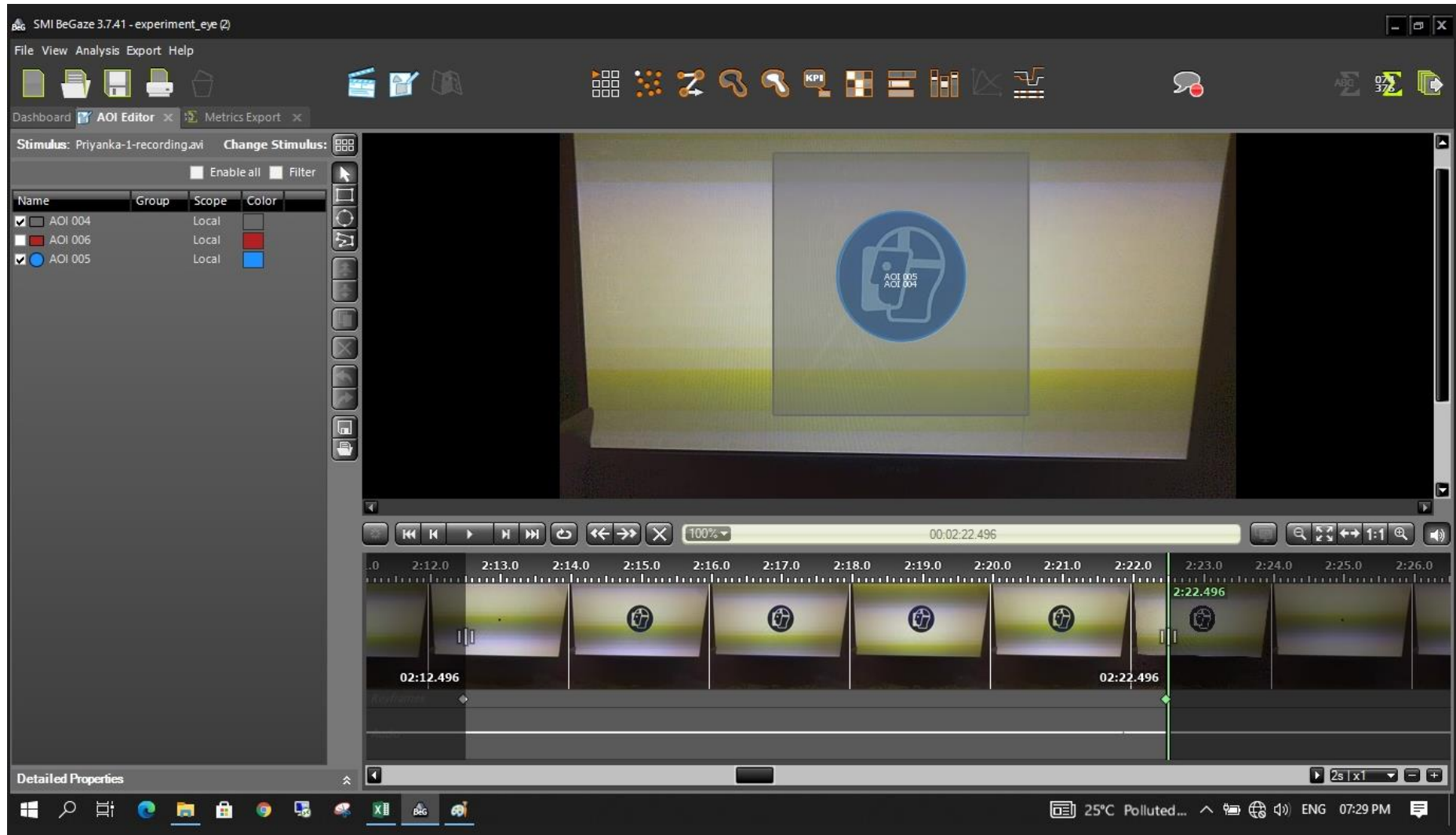


Figure 4. 8. Data extraction for signs without text

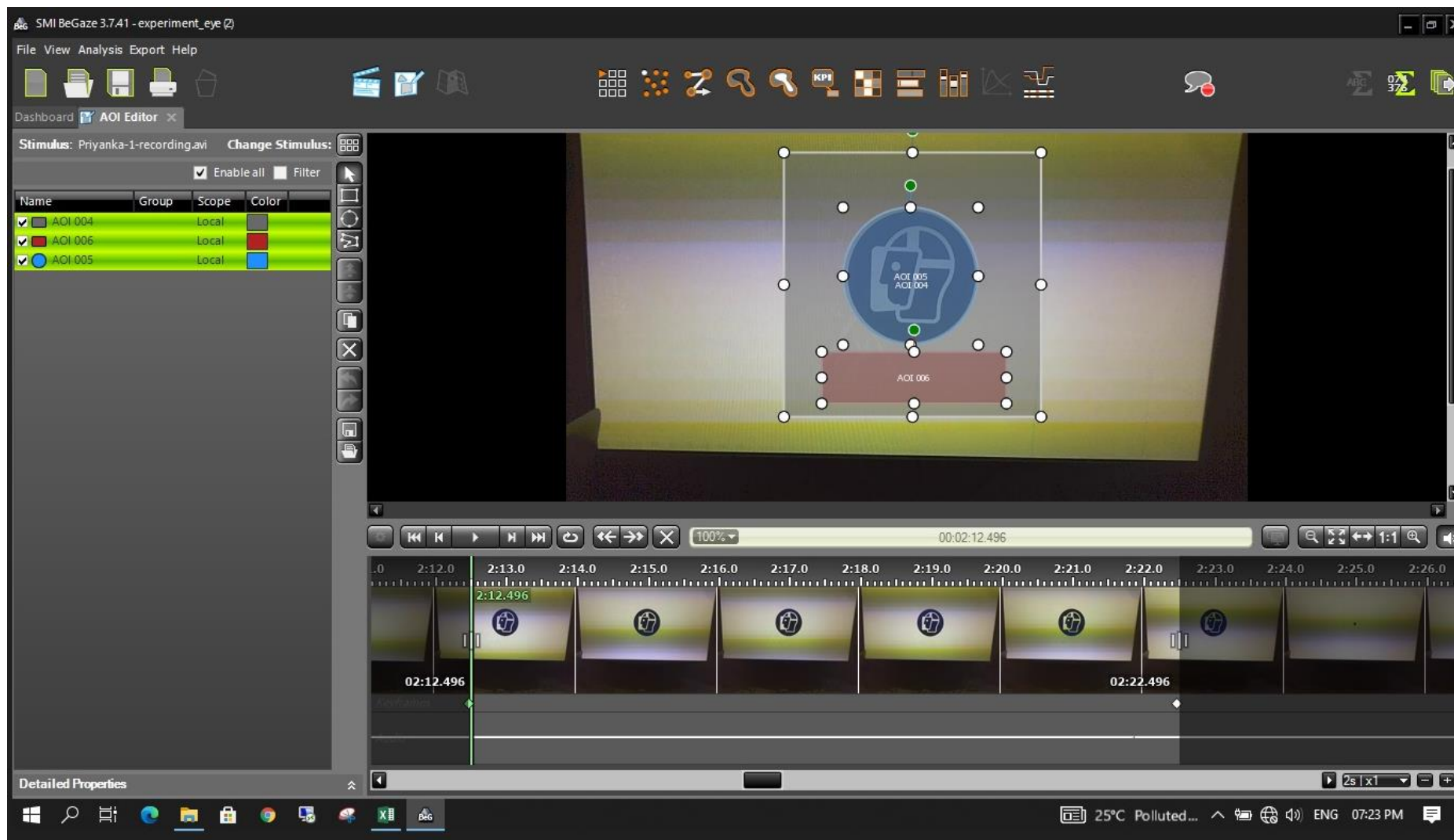


Figure 4. 9. Data extraction for signs with text

4.2.3.3. Statistical tests using data from Method 1 and Method 2

The comprehensibility percentage collected (in method 1) has been used for studying the correlation with the objective measures collected from the eye-tracking experiment (method 2) associated with the signs. Initially, the Pearson plot was used to see their correlation. Later bivariate correlation was used to analyze their relationship. Three correlation models (Pearson's r , Spearman's ρ , and Kendall's Tau) have been used to see which model gives a better correlation. This test helped check the generalizability of the results (Anderson et al., 2012; Price et al., 2017) and find the significant measures. The data is not uniform, and non-parametric tests like Spearman's ρ could have been conducted; however, both parametric and non-parametric have been conducted as both the tests are recommended by many authors as they help to analyze the data in more detail. The intention is to find if there exists any correlation between the two variables. The tests have been performed using IBM SPSS Statistics (Version 25.0). The Kruskal-Wallis test, followed by post hoc analysis, has been performed, which analyses the variability of the comprehensibility percentages among the two groups (categories- comprehensible and incomprehensible) based on ranking the data. The relation between the different significant eye-tracking variables (average pupil diameter right, average pupil diameter left, fixation (or visual intake) count and fixation duration total; the blink count, blink frequency, and blink duration total; saccade count, saccade amplitude average, and saccade duration total, saccade latency average; scan path length) and the comprehensibility score is plotted and studied for eliciting their trend. The overall evaluation model highlighting and listing both the methods and the tests conducted is presented in Figure 4. 10.

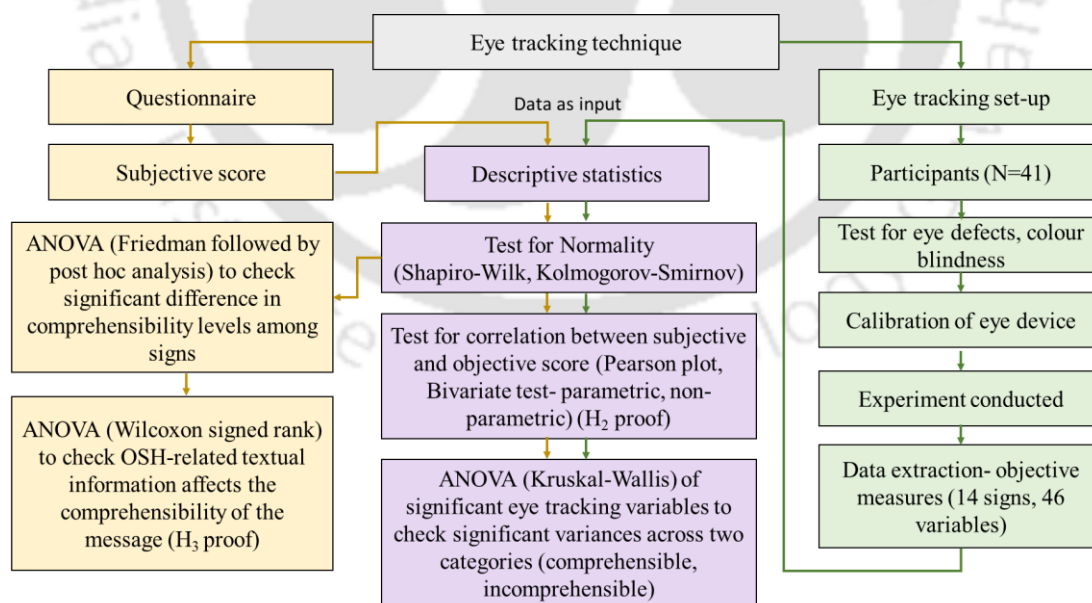


Figure 4. 10: Comprehensibility evaluation model of OSH sign (eye-tracking technique)

4.3. Results and Discussions

Descriptive statistics (mean and standard deviation) for all the subjective (41 data for 14 signs) and objective data (46 measures for every 14 signs and 41 data) have been calculated and are not included in the thesis for brevity.

4.3.1. Test for Normality

Normality tests have been conducted for both the subjective comprehensibility scores and the objective data, and it is found that the data deviates from the normal distribution.

Subjective data: The frequency histogram for all the signs (14) reveals that the data are not normally distributed. Further analysis using normality tests (Kolmogorov-Smirnov and Shapiro-Wilk in Table 4. 1) reveals that the data, i.e., scores of all the 14 signs, deviates from the normal distribution (p-value less than alpha (0.05)). The test fails to accept that sample data are normally distributed (Shapiro & Wilk, 1965). Getting truly, normally distributed data in the real world is hard. Moreover, the normality of data does not cause significant issues with a sample size greater than 25 or 30 (Hogg et al., 1977).

Table 4. 1. Test for normality for subjective variable- comprehensibility scores of 14 signs (eye-tracking technique)

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
DNWSH_score	.173	41	.003	.846	41	.000
NAFUP_score	.152	41	.018	.948	41	.062
WAFS_score	.264	41	.000	.814	41	.000
COSACOEOT_score	.145	41	.029	.925	41	.010
UBC_score	.159	41	.011	.936	41	.023
WRP_score	.188	41	.001	.912	41	.004
DNT_score	.192	41	.001	.893	41	.001
DNWSH_T_score	.301	41	.000	.782	41	.000
NAFUP_T_score	.235	41	.000	.773	41	.000
WAFS_T_score	.266	41	.000	.740	41	.000
COSACOEOT_T_score	.186	41	.001	.836	41	.000
UBC_T_score	.240	41	.000	.828	41	.000
WRP_T_score	.239	41	.000	.795	41	.000
DNT_T_score	.244	41	.000	.780	41	.000

a. Lilliefors Significance Correction

Objective data: A similar normality test has been conducted with objective data (46 measures for every 14 signs) collected from the eye-tracking experiment. The result of the Shapiro-

Wilk test for all the signs (14 signs) has been documented in this report. Only a few measures have shown normal distribution. The highlighted box in the table (Table 4. 2, Table 4. 3, Table 4. 4, and Table 4. 5) reflects the significance value for which the data distribution follows a normal distribution curve. The other values reflect the significant value that deviates from a normal distribution (sig. value < alpha where alpha is 0.05).



Table 4. 2. Normality test result for the eye-tracking variables (N=46) for the 7 signs

Normality test result for the eye-tracking variables (N=46) for 7 signs with 'df=41'

Eye-tracking variables	COSACOEOT				COSACOEOT-T				DNT				DNT-T				DNWSH				DNWSH-T				NAFUP			
	Kolmogorov-Smirnova		Shapiro-Wilk		Kolmogorov-Smirnova		Shapiro-Wilk		Kolmogorov-Smirnova		Shapiro-Wilk		Kolmogorov-Smirnova		Shapiro-Wilk		Kolmogorov-Smirnova		Shapiro-Wilk		Kolmogorov-Smirnova		Shapiro-Wilk		Kolmogorov-Smirnova		Shapiro-Wilk	
	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.
	AvgpupilDR	0.194	0.000	0.876	0.000	0.171	0.004	0.858	0.000	0.174	0.003	0.835	0.000	0.163	0.008	0.884	0.001	0.205	0.000	0.843	0.000	0.145	0.030	0.901	0.002	0.185	0.001	0.866
AvgpupilDL	0.140	0.043	0.866	0.000	0.205	0.000	0.824	0.000	0.165	0.007	0.805	0.000	0.162	0.008	0.865	0.000	0.182	0.002	0.874	0.000	0.161	0.009	0.886	0.001	0.141	0.040	0.838	0.000
vis_intk_cnt	0.132	0.071	0.963	0.200	0.105	.200*	0.978	0.618	0.125	0.107	0.958	0.135	0.087	.200*	0.987	0.916	0.075	.200*	0.979	0.639	0.115	0.196	0.981	0.730	0.124	0.112	0.964	0.215
vis_intk_freq	0.132	0.071	0.963	0.200	0.105	.200*	0.978	0.618	0.125	0.107	0.958	0.135	0.087	.200*	0.987	0.916	0.075	.200*	0.979	0.639	0.115	0.196	0.981	0.730	0.124	0.112	0.964	0.215
Vis_intk_durtot	0.131	0.074	0.817	0.000	0.118	0.168	0.951	0.079	0.239	0.000	0.799	0.000	0.162	0.008	0.868	0.000	0.262	0.000	0.721	0.000	0.162	0.009	0.898	0.001	0.225	0.000	0.819	0.000
Vis_intk_durAvg	0.362	0.000	0.318	0.000	0.188	0.001	0.890	0.001	0.215	0.000	0.701	0.000	0.231	0.000	0.652	0.000	0.292	0.000	0.465	0.000	0.243	0.000	0.765	0.000	0.294	0.000	0.526	0.000
Vis_intk_durMax	0.260	0.000	0.616	0.000	0.229	0.000	0.727	0.000	0.169	0.005	0.815	0.000	0.175	0.003	0.729	0.000	0.233	0.000	0.638	0.000	0.205	0.000	0.716	0.000	0.269	0.000	0.616	0.000
Vis_intk_durMin	0.477	0.000	0.169	0.000	0.274	0.000	0.864	0.000	0.435	0.000	0.367	0.000	0.222	0.000	0.876	0.000	0.416	0.000	0.262	0.000	0.253	0.000	0.763	0.000	0.324	0.000	0.461	0.000
Sac_cnt	0.092	.200*	0.982	0.739	0.112	.200*	0.986	0.886	0.077	.200*	0.983	0.795	0.082	.200*	0.965	0.229	0.108	.200*	0.966	0.251	0.092	.200*	0.978	0.597	0.090	.200*	0.975	0.489
Sac_freq	0.092	.200*	0.982	0.739	0.112	.200*	0.986	0.886	0.077	.200*	0.983	0.795	0.082	.200*	0.965	0.229	0.108	.200*	0.966	0.251	0.092	.200*	0.978	0.597	0.090	.200*	0.975	0.489
Sac_durtot	0.072	.200*	0.985	0.846	0.111	.200*	0.900	0.002	0.088	.200*	0.965	0.227	0.114	.200*	0.949	0.065	0.099	.200*	0.969	0.317	0.128	0.090	0.878	0.000	0.092	.200*	0.969	0.315
Sac_durAvg	0.159	0.011	0.885	0.001	0.230	0.000	0.599	0.000	0.270	0.000	0.644	0.000	0.306	0.000	0.419	0.000	0.196	0.000	0.685	0.000	0.303	0.000	0.571	0.000	0.157	0.012	0.875	0.000
Sac_durMax	0.175	0.003	0.924	0.010	0.175	0.003	0.804	0.000	0.268	0.000	0.527	0.000	0.203	0.000	0.726	0.000	0.209	0.000	0.847	0.000	0.277	0.000	0.625	0.000	0.284	0.000	0.701	0.000
Sac_durMin	0.433	0.000	0.489	0.000	0.496	0.000	0.440	0.000	0.379	0.000	0.425	0.000	0.497	0.000	0.287	0.000	0.396	0.000	0.643	0.000	0.434	0.000	0.583	0.000	0.502	0.000	0.403	0.000
Sac_ampltot	0.214	0.000	0.853	0.000	0.285	0.000	0.605	0.000	0.404	0.000	0.222	0.000	0.177	0.002	0.892	0.001	0.148	0.025	0.833	0.000	0.448	0.000	0.208	0.000	0.126	0.102	0.877	0.000
Sac_amplAvg	0.216	0.000	0.729	0.000	0.351	0.000	0.551	0.000	0.458	0.000	0.190	0.000	0.206	0.000	0.826	0.000	0.250	0.000	0.686	0.000	0.441	0.000	0.201	0.000	0.211	0.000	0.632	0.000
Sac_amplMax	0.347	0.000	0.514	0.000	0.408	0.000	0.243	0.000	0.450	0.000	0.175	0.000	0.386	0.000	0.574	0.000	0.357	0.000	0.395	0.000	0.410	0.000	0.203	0.000	0.455	0.000	0.321	0.000
Sac_amplMin	0.147	0.026	0.960	0.160	0.168	0.005	0.931	0.016	0.235	0.000	0.762	0.000	0.138	0.047	0.958	0.138	0.244	0.000	0.763	0.000	0.117	0.174	0.956	0.110	0.170	0.004	0.874	0.000
Sac_veltot	0.240	0.000	0.554	0.000	0.377	0.000	0.358	0.000	0.398	0.000	0.259	0.000	0.268	0.000	0.565	0.000	0.362	0.000	0.363	0.000	0.375	0.000	0.335	0.000	0.331	0.000	0.344	0.000

Sac_velAvg	0.321	0.000	0.594	0.000	0.376	0.000	0.340	0.000	0.401	0.000	0.313	0.000	0.390	0.000	0.240	0.000	0.460	0.000	0.220	0.000	0.392	0.000	0.299	0.000	0.364	0.000	0.304	0.000
Sac_velMax	0.353	0.000	0.546	0.000	0.382	0.000	0.414	0.000	0.383	0.000	0.392	0.000	0.356	0.000	0.367	0.000	0.423	0.000	0.252	0.000	0.329	0.000	0.466	0.000	0.462	0.000	0.235	0.000
Sac_velMin	0.182	0.002	0.867	0.000	0.113	.200*	0.949	0.064	0.513	0.000	0.164	0.000	0.185	0.001	0.880	0.000	0.285	0.000	0.491	0.000	0.172	0.004	0.895	0.001	0.213	0.000	0.755	0.000
Sac_latAvg	0.188	0.001	0.892	0.001	0.161	0.009	0.875	0.000	0.254	0.000	0.470	0.000	0.125	0.104	0.832	0.000	0.209	0.000	0.798	0.000	0.219	0.000	0.870	0.000	0.233	0.000	0.692	0.000
Blink_cnt	0.289	0.000	0.795	0.000	0.273	0.000	0.735	0.000	0.297	0.000	0.730	0.000	0.189	0.001	0.872	0.000	0.247	0.000	0.798	0.000	0.180	0.002	0.869	0.000	0.237	0.000	0.773	0.000
Blink_freq	0.289	0.000	0.795	0.000	0.273	0.000	0.735	0.000	0.297	0.000	0.730	0.000	0.189	0.001	0.872	0.000	0.247	0.000	0.798	0.000	0.180	0.002	0.869	0.000	0.237	0.000	0.773	0.000
Blink_dur	0.255	0.000	0.689	0.000	0.280	0.000	0.704	0.000	0.292	0.000	0.596	0.000	0.269	0.000	0.603	0.000	0.292	0.000	0.583	0.000	0.232	0.000	0.747	0.000	0.265	0.000	0.657	0.000
Blink_durAvg	0.241	0.000	0.857	0.000	0.223	0.000	0.839	0.000	0.249	0.000	0.803	0.000	0.176	0.003	0.736	0.000	0.202	0.000	0.722	0.000	0.198	0.000	0.772	0.000	0.256	0.000	0.832	0.000
Blink_durMax	0.190	0.001	0.756	0.000	0.223	0.000	0.736	0.000	0.253	0.000	0.666	0.000	0.202	0.000	0.730	0.000	0.316	0.000	0.456	0.000	0.269	0.000	0.597	0.000	0.211	0.000	0.761	0.000
Blink_durMin	0.250	0.000	0.847	0.000	0.214	0.000	0.809	0.000	0.254	0.000	0.753	0.000	0.173	0.004	0.801	0.000	0.242	0.000	0.868	0.000	0.220	0.000	0.881	0.000	0.285	0.000	0.794	0.000
Scanpath_len	0.230	0.000	0.701	0.000	0.167	0.006	0.780	0.000	0.455	0.000	0.204	0.000	0.390	0.000	0.317	0.000	0.181	0.002	0.724	0.000	0.426	0.000	0.250	0.000	0.200	0.000	0.675	0.000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 4. 3. Normality test result for the eye-tracking variables (N=46) for the remaining 7 signs

Normality test result for the eye-tracking variables (N=46) for the remaining 7 signs with 'df=41'

Eye-tracking variables	NAFUP-T				UBC				UBC-T				WAFS				WAFS-T				WRP				WRP-T			
	Kolmogorov-Smirnova		Shapiro-Wilk		Kolmogorov-Smirnova		Shapiro-Wilk		Kolmogorov-Smirnova		Shapiro-Wilk		Kolmogorov-Smirnova		Shapiro-Wilk		Kolmogorov-Smirnova		Shapiro-Wilk		Kolmogorov-Smirnova		Shapiro-Wilk		Kolmogorov-Smirnova		Shapiro-Wilk	
	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.
	AvgpupilDR	0.157	0.012	0.882	0.001	0.186	0.001	0.852	0.000	0.136	0.055	0.898	0.001	0.195	0.000	0.857	0.000	0.168	0.005	0.854	0.000	0.192	0.001	0.802	0.000	0.174	0.003	0.864
AvgpupilDL	0.194	0.000	0.873	0.000	0.211	0.000	0.818	0.000	0.143	0.035	0.867	0.000	0.125	0.106	0.847	0.000	0.184	0.001	0.807	0.000	0.184	0.001	0.769	0.000	0.228	0.000	0.808	0.000
vis_intk_cnt	0.135	0.058	0.969	0.329	0.088	.200*	0.985	0.860	0.142	0.038	0.919	0.006	0.131	0.073	0.965	0.243	0.196	0.000	0.903	0.002	0.106	.200*	0.946	0.050	0.107	.200*	0.957	0.119
vis_intk_freq	0.135	0.058	0.969	0.329	0.088	.200*	0.985	0.860	0.142	0.038	0.919	0.006	0.131	0.073	0.965	0.243	0.196	0.000	0.903	0.002	0.106	.200*	0.946	0.050	0.107	.200*	0.957	0.119
Vis_intk_durtot	0.160	0.010	0.929	0.013	0.212	0.000	0.839	0.000	0.218	0.000	0.876	0.000	0.188	0.001	0.846	0.000	0.252	0.000	0.796	0.000	0.156	0.014	0.919	0.006	0.288	0.000	0.535	0.000

Vis_intk_durAvg	0.262	0.000	0.702	0.000	0.337	0.000	0.361	0.000	0.272	0.000	0.690	0.000	0.229	0.000	0.696	0.000	0.361	0.000	0.460	0.000	0.195	0.000	0.873	0.000	0.272	0.000	0.677	0.000
Vis_intk_durMax	0.226	0.000	0.809	0.000	0.139	0.046	0.801	0.000	0.218	0.000	0.700	0.000	0.197	0.000	0.799	0.000	0.259	0.000	0.752	0.000	0.216	0.000	0.848	0.000	0.285	0.000	0.495	0.000
Vis_intk_durMin	0.286	0.000	0.813	0.000	0.513	0.000	0.165	0.000	0.300	0.000	0.666	0.000	0.176	0.003	0.861	0.000	0.376	0.000	0.302	0.000	0.229	0.000	0.827	0.000	0.233	0.000	0.586	0.000
Sac_cnt	0.099	.200*	0.963	0.201	0.127	0.094	0.962	0.186	0.135	0.058	0.956	0.111	0.098	.200*	0.975	0.480	0.157	0.013	0.939	0.029	0.140	0.041	0.946	0.050	0.087	.200*	0.978	0.582
Sac_freq	0.099	.200*	0.963	0.201	0.127	0.094	0.962	0.186	0.135	0.058	0.956	0.111	0.098	.200*	0.975	0.480	0.157	0.013	0.939	0.029	0.140	0.041	0.946	0.050	0.087	.200*	0.978	0.582
Sac_durtot	0.109	.200*	0.974	0.460	0.132	0.070	0.936	0.022	0.080	.200*	0.990	0.971	0.090	.200*	0.974	0.454	0.104	.200*	0.975	0.503	0.123	0.121	0.954	0.099	0.084	.200*	0.971	0.369
Sac_durAvg	0.229	0.000	0.576	0.000	0.276	0.000	0.631	0.000	0.211	0.000	0.647	0.000	0.288	0.000	0.465	0.000	0.222	0.000	0.789	0.000	0.264	0.000	0.564	0.000	0.264	0.000	0.612	0.000
Sac_durMax	0.177	0.002	0.778	0.000	0.229	0.000	0.787	0.000	0.155	0.015	0.921	0.008	0.197	0.000	0.791	0.000	0.159	0.011	0.932	0.016	0.277	0.000	0.651	0.000	0.164	0.007	0.818	0.000
Sac_durMin	0.455	0.000	0.550	0.000	0.419	0.000	0.539	0.000	0.506	0.000	0.361	0.000	0.510	0.000	0.253	0.000	0.426	0.000	0.535	0.000	0.502	0.000	0.402	0.000	0.494	0.000	0.436	0.000
Sac_ampltot	0.305	0.000	0.478	0.000	0.237	0.000	0.477	0.000	0.101	.200*	0.962	0.192	0.476	0.000	0.235	0.000	0.124	0.117	0.944	0.044	0.454	0.000	0.253	0.000	0.183	0.001	0.770	0.000
Sac_amplAvg	0.415	0.000	0.264	0.000	0.349	0.000	0.380	0.000	0.159	0.011	0.918	0.006	0.481	0.000	0.185	0.000	0.161	0.009	0.930	0.015	0.431	0.000	0.203	0.000	0.290	0.000	0.779	0.000
Sac_amplMax	0.455	0.000	0.198	0.000	0.403	0.000	0.394	0.000	0.355	0.000	0.609	0.000	0.469	0.000	0.213	0.000	0.394	0.000	0.603	0.000	0.432	0.000	0.192	0.000	0.380	0.000	0.652	0.000
Sac_amplMin	0.129	0.083	0.963	0.201	0.258	0.000	0.802	0.000	0.184	0.001	0.796	0.000	0.128	0.087	0.973	0.425	0.272	0.000	0.758	0.000	0.183	0.001	0.944	0.044	0.202	0.000	0.783	0.000
Sac_veltot	0.276	0.000	0.478	0.000	0.357	0.000	0.275	0.000	0.244	0.000	0.507	0.000	0.399	0.000	0.246	0.000	0.264	0.000	0.586	0.000	0.426	0.000	0.225	0.000	0.103	.200*	0.944	0.043
Sac_velAvg	0.401	0.000	0.272	0.000	0.384	0.000	0.242	0.000	0.446	0.000	0.211	0.000	0.417	0.000	0.194	0.000	0.388	0.000	0.273	0.000	0.447	0.000	0.216	0.000	0.255	0.000	0.572	0.000
Sac_velMax	0.339	0.000	0.431	0.000	0.385	0.000	0.347	0.000	0.338	0.000	0.405	0.000	0.413	0.000	0.369	0.000	0.272	0.000	0.517	0.000	0.407	0.000	0.266	0.000	0.338	0.000	0.569	0.000
Sac_velMin	0.139	0.046	0.942	0.036	0.139	0.045	0.866	0.000	0.364	0.000	0.326	0.000	0.182	0.001	0.663	0.000	0.205	0.000	0.800	0.000	0.163	0.008	0.860	0.000	0.215	0.000	0.805	0.000
Sac_latAvg	0.161	0.009	0.870	0.000	0.279	0.000	0.709	0.000	0.257	0.000	0.648	0.000	0.296	0.000	0.528	0.000	0.229	0.000	0.760	0.000	0.240	0.000	0.639	0.000	0.305	0.000	0.586	0.000
Blink_cnt	0.304	0.000	0.777	0.000	0.249	0.000	0.746	0.000	0.239	0.000	0.828	0.000	0.244	0.000	0.779	0.000	0.322	0.000	0.748	0.000	0.315	0.000	0.701	0.000	0.280	0.000	0.743	0.000
Blink_freq	0.304	0.000	0.777	0.000	0.249	0.000	0.746	0.000	0.239	0.000	0.828	0.000	0.244	0.000	0.779	0.000	0.322	0.000	0.748	0.000	0.315	0.000	0.701	0.000	0.280	0.000	0.743	0.000
Blink_dur	0.343	0.000	0.403	0.000	0.306	0.000	0.570	0.000	0.289	0.000	0.679	0.000	0.285	0.000	0.650	0.000	0.286	0.000	0.616	0.000	0.298	0.000	0.546	0.000	0.309	0.000	0.547	0.000
Blink_durAvg	0.279	0.000	0.555	0.000	0.334	0.000	0.435	0.000	0.301	0.000	0.420	0.000	0.191	0.001	0.765	0.000	0.255	0.000	0.612	0.000	0.161	0.009	0.818	0.000	0.348	0.000	0.346	0.000
Blink_durMax	0.387	0.000	0.272	0.000	0.351	0.000	0.445	0.000	0.348	0.000	0.534	0.000	0.243	0.000	0.624	0.000	0.290	0.000	0.555	0.000	0.286	0.000	0.550	0.000	0.368	0.000	0.316	0.000
Blink_durMin	0.182	0.002	0.784	0.000	0.390	0.000	0.301	0.000	0.313	0.000	0.410	0.000	0.249	0.000	0.848	0.000	0.287	0.000	0.456	0.000	0.196	0.000	0.892	0.001	0.267	0.000	0.759	0.000
Scanpath_len	0.118	0.167	0.954	0.098	0.463	0.000	0.202	0.000	0.258	0.000	0.514	0.000	0.477	0.000	0.219	0.000	0.202	0.000	0.653	0.000	0.414	0.000	0.244	0.000	0.143	0.034	0.935	0.022

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 4. 4. Normality test result for the eye-tracking variables (N=46) considering AOI for 7 signs

Normality test result for the eye-tracking variables (N=46) considering AOI for 7 signs with ‘df=41’

Eye-tracking variables	COSACOEOT				COSACOEOT-T				DNT				DNT-T				DNWSH				DNWSH-T				NAFUP			
	Kolmogorov-				Kolmogorov-		Shapiro-		Kolmogorov-				Kolmogorov-		Shapiro-		Kolmogorov-				Kolmogorov-		Shapiro-		Kolmogorov-		Shapiro-	
	Smirnova		Shapiro-Wilk		Smirnova		Wilk		Smirnova		Shapiro-Wilk		Smirnova		Shapiro-Wilk		Smirnova		Shapiro-Wilk		Smirnova		Shapiro-Wilk		Smirnova		Shapiro-Wilk	
	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.
AOI_size004First_vis_intkdur	0.379	0.000	0.255	0.000	0.298	0.000	0.567	0.000	0.331	0.000	0.533	0.000	0.334	0.000	0.487	0.000	0.280	0.000	0.455	0.000	0.272	0.000	0.628	0.000	0.208	0.000	0.790	0.000
AOI_size004_Revisits	0.532	0.000	0.287	0.000	0.516	0.000	0.329	0.000	0.530	0.000	0.288	0.000	0.480	0.000	0.452	0.000	0.540	0.000	0.226	0.000	0.522	0.000	0.344	0.000	0.538	0.000	0.144	0.000
AOI_size004_Vis_intkCnt	0.142	0.036	0.956	0.115	0.112	.200*	0.973	0.425	0.119	0.150	0.958	0.131	0.079	.200*	0.983	0.778	0.095	.200*	0.976	0.537	0.108	.200*	0.980	0.685	0.131	0.073	0.958	0.134
AOI_size004_Vis_intkTime	0.130	0.078	0.830	0.000	0.107	.200*	0.956	0.115	0.238	0.000	0.803	0.000	0.157	0.012	0.858	0.000	0.261	0.000	0.723	0.000	0.145	0.030	0.910	0.003	0.215	0.000	0.784	0.000
AOI_size004_Avg_vis_intkdur	0.363	0.000	0.317	0.000	0.176	0.003	0.900	0.002	0.215	0.000	0.700	0.000	0.230	0.000	0.658	0.000	0.288	0.000	0.467	0.000	0.243	0.000	0.764	0.000	0.251	0.000	0.733	0.000
AOI_size005First_vis_intkdur	0.380	0.000	0.253	0.000	0.236	0.000	0.709	0.000	0.331	0.000	0.533	0.000	0.273	0.000	0.626	0.000	0.280	0.000	0.455	0.000	0.353	0.000	0.413	0.000	0.208	0.000	0.792	0.000
AOI_size005_Revisits	0.479	0.000	0.505	0.000	0.294	0.000	0.826	0.000	0.517	0.000	0.331	0.000	0.207	0.000	0.877	0.000	0.513	0.000	0.389	0.000	0.162	0.009	0.906	0.003	0.522	0.000	0.344	0.000
AOI_size005_Vis_intkCnt	0.147	0.026	0.949	0.067	0.127	0.092	0.944	0.042	0.112	.200*	0.961	0.175	0.118	0.162	0.976	0.528	0.114	.200*	0.969	0.315	0.118	0.160	0.961	0.172	0.142	0.038	0.951	0.074
AOI_size005_Vis_intkTime	0.172	0.004	0.852	0.000	0.107	.200*	0.954	0.099	0.236	0.000	0.807	0.000	0.126	0.097	0.937	0.024	0.258	0.000	0.743	0.000	0.100	.200*	0.970	0.350	0.225	0.000	0.805	0.000
AOI_size005_Avg_vis_intkdur	0.360	0.000	0.320	0.000	0.241	0.000	0.656	0.000	0.215	0.000	0.700	0.000	0.243	0.000	0.645	0.000	0.288	0.000	0.469	0.000	0.313	0.000	0.384	0.000	0.256	0.000	0.721	0.000
AOI_size006First_vis_intkdur	0.504	0.000	0.286	0.000	0.276	0.000	0.431	0.000	0.538	0.000	0.230	0.000	0.249	0.000	0.691	0.000	0.504	0.000	0.196	0.000	0.289	0.000	0.684	0.000				
AOI_size006_Revisits	0.530	0.000	0.206	0.000	0.201	0.000	0.882	0.001	0.538	0.000	0.144	0.000	0.287	0.000	0.767	0.000	0.538	0.000	0.144	0.000	0.265	0.000	0.794	0.000				
AOI_size006_Vis_intkCnt	0.507	0.000	0.299	0.000	0.110	.200*	0.961	0.174	0.523	0.000	0.179	0.000	0.225	0.000	0.853	0.000	0.501	0.000	0.185	0.000	0.184	0.001	0.880	0.000				
AOI_size006_Vis_intkTime	0.496	0.000	0.246	0.000	0.146	0.029	0.936	0.023	0.524	0.000	0.183	0.000	0.238	0.000	0.758	0.000	0.516	0.000	0.240	0.000	0.159	0.010	0.928	0.012				
AOI_size006_Avg_vis_intkdur	0.504	0.000	0.288	0.000	0.182	0.001	0.821	0.000	0.540	0.000	0.229	0.000	0.170	0.004	0.825	0.000	0.505	0.000	0.202	0.000	0.258	0.000	0.740	0.000				
TimeToFirstAppear_allAOI	0.127	0.092	0.952	0.080	0.122	0.133	0.953	0.092	0.110	.200*	0.919	0.006	0.088	.200*	0.920	0.007	0.122	0.129	0.883	0.001	0.127	0.093	0.950	0.072	0.120	0.146	0.934	0.020

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 4. 5. Normality test result for the eye-tracking variables (N=46) considering AOI for the remaining 7 signs

Normality test result for the eye-tracking variables (N=46) considering AOI for the remaining 7 signs with 'df=41'

Eye-tracking AOI variables	NAFUP-T				UBC				UBC-T				WAFS				WAFS-T				WRP				WRP-T			
	Kolmogorov-				Kolmogorov-		Shapiro-		Kolmogorov-				Kolmogorov-				Kolmogorov-		Shapiro-		Kolmogorov-				Kolmogorov-		Shapiro-	
	Smirnova	Shapiro-Wilk	Smirnova	Wilk	Smirnova	Shapiro-Wilk	Smirnova	Shapiro-Wilk	Smirnova	Shapiro-Wilk	Smirnova	Shapiro-Wilk	Smirnova	Shapiro-Wilk	Smirnova	Shapiro-Wilk	Smirnova	Shapiro-Wilk	Smirnova	Shapiro-Wilk	Smirnova	Shapiro-Wilk	Smirnova	Shapiro-Wilk	Smirnova	Shapiro-Wilk		
	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.	Stat	Sig.		
AOI_size004First_vis_intkdur	0.198	0.000	0.710	0.000	0.397	0.000	0.312	0.000	0.195	0.000	0.710	0.000	0.253	0.000	0.747	0.000	0.237	0.000	0.632	0.000	0.278	0.000	0.635	0.000	0.298	0.000	0.511	0.000
AOI_size004_Revisits	0.517	0.000	0.331	0.000	0.524	0.000	0.185	0.000	0.496	0.000	0.344	0.000	0.530	0.000	0.206	0.000	0.538	0.000	0.144	0.000	0.538	0.000	0.144	0.000	0.512	0.000	0.311	0.000
AOI_size004_Vis_intkCnt	0.124	0.113	0.969	0.320	0.102	.200*	0.982	0.743	0.159	0.011	0.922	0.008	0.138	0.047	0.962	0.183	0.199	0.000	0.903	0.002	0.103	.200*	0.944	0.042	0.137	0.051	0.934	0.019
AOI_size004_Vis_intkTime	0.151	0.019	0.933	0.018	0.221	0.000	0.824	0.000	0.235	0.000	0.879	0.000	0.203	0.000	0.821	0.000	0.253	0.000	0.795	0.000	0.157	0.012	0.909	0.003	0.264	0.000	0.608	0.000
AOI_size004_Avg_vis_intkdur	0.251	0.000	0.720	0.000	0.319	0.000	0.363	0.000	0.273	0.000	0.688	0.000	0.229	0.000	0.697	0.000	0.361	0.000	0.460	0.000	0.196	0.000	0.872	0.000	0.269	0.000	0.716	0.000
AOI_size005First_vis_intkdur	0.217	0.000	0.759	0.000	0.396	0.000	0.311	0.000	0.291	0.000	0.528	0.000	0.248	0.000	0.765	0.000	0.238	0.000	0.743	0.000	0.278	0.000	0.635	0.000	0.292	0.000	0.628	0.000
AOI_size005_Revisits	0.287	0.000	0.852	0.000	0.456	0.000	0.451	0.000	0.198	0.000	0.900	0.002	0.433	0.000	0.524	0.000	0.221	0.000	0.854	0.000	0.465	0.000	0.484	0.000	0.231	0.000	0.870	0.000
AOI_size005_Vis_intkCnt	0.091	.200*	0.976	0.540	0.098	.200*	0.975	0.511	0.169	0.005	0.954	0.098	0.125	0.108	0.952	0.083	0.150	0.021	0.910	0.003	0.130	0.080	0.928	0.012	0.137	0.052	0.956	0.116
AOI_size005_Vis_intkTime	0.100	.200*	0.987	0.904	0.207	0.000	0.850	0.000	0.106	.200*	0.958	0.136	0.200	0.000	0.825	0.000	0.114	.200*	0.943	0.040	0.190	0.001	0.870	0.000	0.160	0.010	0.901	0.002
AOI_size005_Avg_vis_intkdur	0.242	0.000	0.777	0.000	0.316	0.000	0.365	0.000	0.277	0.000	0.564	0.000	0.236	0.000	0.718	0.000	0.360	0.000	0.480	0.000	0.194	0.000	0.878	0.000	0.335	0.000	0.661	0.000
AOI_size006First_vis_intkdur	0.178	0.002	0.884	0.001					0.189	0.001	0.788	0.000	0.538	0.000	0.144	0.000	0.244	0.000	0.698	0.000	0.520	0.000	0.259	0.000	0.191	0.001	0.640	0.000
AOI_size006_Revisits	0.239	0.000	0.835	0.000					0.225	0.000	0.827	0.000					0.294	0.000	0.771	0.000	0.538	0.000	0.144	0.000	0.307	0.000	0.750	0.000
AOI_size006_Vis_intkCnt	0.171	0.004	0.912	0.004					0.216	0.000	0.822	0.000	0.538	0.000	0.144	0.000	0.147	0.025	0.912	0.004	0.511	0.000	0.223	0.000	0.144	0.031	0.931	0.015
AOI_size006_Vis_intkTime	0.219	0.000	0.782	0.000					0.199	0.000	0.741	0.000	0.538	0.000	0.144	0.000	0.228	0.000	0.665	0.000	0.507	0.000	0.210	0.000	0.279	0.000	0.516	0.000
AOI_size006_Avg_vis_intkdur	0.242	0.000	0.614	0.000					0.342	0.000	0.445	0.000	0.538	0.000	0.144	0.000	0.266	0.000	0.626	0.000	0.523	0.000	0.266	0.000	0.219	0.000	0.681	0.000
TimeToFirstAppear_allAOI	0.144	0.032	0.929	0.014	0.098	.200*	0.955	0.106	0.108	.200*	0.960	0.155	0.102	.200*	0.969	0.317	0.085	.200*	0.955	0.108	0.106	.200*	0.954	0.100	0.106	.200*	0.912	0.004

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

4.3.2. Comprehensibility percentage from subjective data and the result interpretation for objective data

The comprehensibility percentage is calculated from subjective data, i.e., scores provided by the subjects, obtained for every 14 signs to study the correlation between the subjective score and objective variables. The 11-point rating scale has been converted to a binary category. The total comprehensibility percentage is calculated and the signs that are more than or equal to 67% (Berrio et al., 2022; ISO 3864-1, 2011) comprehensible by the target population are considered acceptable (Table 4. 6). This means that a minimum of 67% target population found the signs comprehensible. This comprehensibility percentage table (Table 4. 6) reveals that signs with text are more comprehensible than signs without text. This hypothesis is further strengthened using ANOVA and hypothesis tests (sections 4.3.3 and 4.3.4).

By viewing the Comprehensibility Percentage (Table 4. 6), it is possible to determine which sign is comprehensible and which is not, and they can be arranged according to the increasing/ decreasing order of comprehensibility. Considering the criteria for the minimum threshold to be 67% (ISO 9186), all the signs are comprehensible except COSACOEOT, DNT, and WRP.

Table 4.6 shows that COSACOEOT is less comprehensible than DNT which is in turn less comprehensible than WRP and it follows. DNWSH_T is the most comprehensible sign found in this study. It is to be noted that COSACOEOT_T, UBC_T, and WRP_T show the same comprehensibility percentage in this study. Similarly, NAFUP-T and WAFS_T show the same comprehensibility percentage.

Table 4. 6. Comprehensibility percentage based on subjective data (eye-tracking technique)

Sign label	Comprehensibility Percentage [%]	Sign label	Comprehensibility Percentage [%]
COSACOEOT	43.90244	COSACOEOT_T	90.2439
DNT	63.41463	DNT_T	92.68293
DNWSH	95.12195	DNWSH_T	100
NAFUP	73.17073	NAFUP_T	97.56098
UBC	75.60976	UBC_T	90.2439
WAFS	90.24390244	WAFS_T	97.56098
WRP	65.85366	WRP_T	90.2439

These comprehensibility percentages have been used in the Pearson plot and bivariate correlation analysis (test for correlation between subjective score and eye-tracking objective metrics).

4.3.2.1. Result Analysis using Correlation Plot

The relationship between the subjective score and the objective measure has been studied using the correlation plot (Pearson plot) followed by bivariate correlation analysis.

A hypothesis has been proposed:

Hypothesis 2 (H₂): There is a significant correlation between the subjective response of the comprehensibility and the objective measures (eye tracking variables: average pupil diameter, visual intake count, visual intake duration, blink count, blink duration, blink frequency, saccade count, saccade amplitude, saccade latency, saccade duration, scan path length).

Test of hypothesis: To test the hypothesis, a correlation (Pearson) plot followed by bivariate correlation analysis (Pearson's r , Spearman's ρ , and Kendall's tau- refer to Appendix G) has been conducted to examine the existence of a relationship between the two variables-scores provided by the participants and the eye tracking variables. The Pearson plot and the analysis results prove that there is definitely a significant relationship between the subjective response of the comprehensibility and the objective measures (eye tracking variables: average pupil diameter, visual intake count, visual intake duration, blink count, blink duration, blink frequency, saccade count, saccade amplitude, saccade latency, saccade duration, scan path length), evident from their trend analysis in correlation plot (R^2 values), and bivariate correlation analyses (evident from p values), thus helping to reject the null hypothesis and proving the alternate hypothesis (proved). The detailed discussion of the Pearson plot continues, followed by the bivariate correlation result and discussion.

Pearson plot visualization has been created to see the correlation or relationship between the two variables. Eye tracking data has been plotted against the comprehensibility percentage (levels), and trends have been analyzed. Considering the average pupil diameter (right and left) (Figure 4. 12 and Figure 4. 11), it is seen that as the comprehensibility percentage (%) increases, there is a decrease in the average pupil diameter (Avg Pupil Diameter Right and Avg Pupil Diameter Left). This means that as the comprehensibility of signs increases (transition from less comprehensible signs to more comprehensible signs), the average pupil diameter tends to decrease. Incomprehensible signs elicited larger average pupil diameter, which decreased as the sign became comprehensible. The result is in line with the study of Siswandari and Xiong (Siswandari & Xiong, 2015).

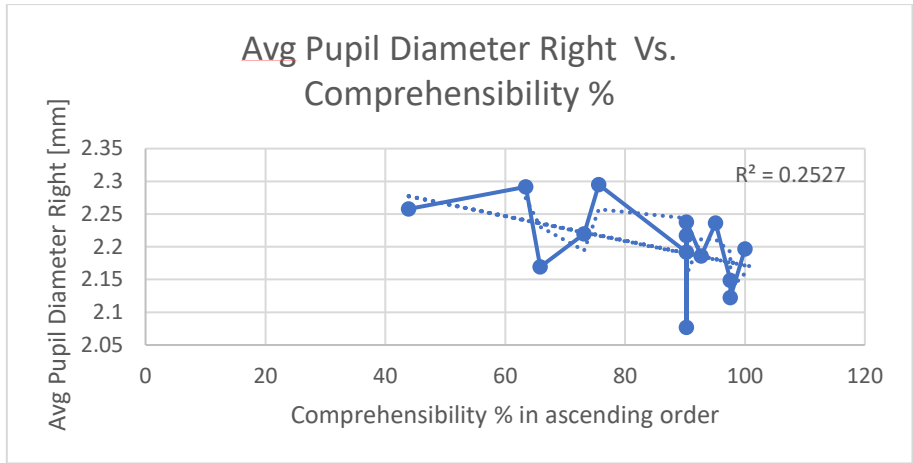


Figure 4. 11. Plot showing a relationship between Avg Pupil Diameter Right and comprehensibility percentage

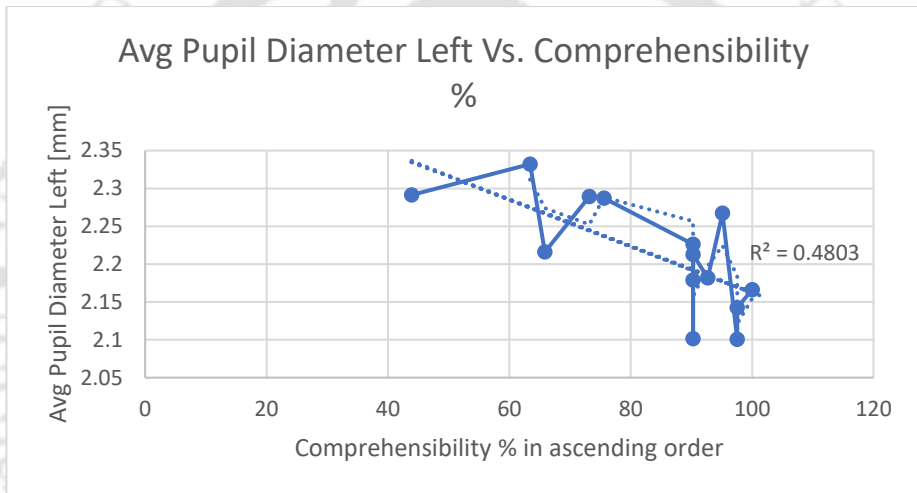


Figure 4. 12. Plot showing a relationship between Avg Pupil Diameter Left and comprehensibility percentage

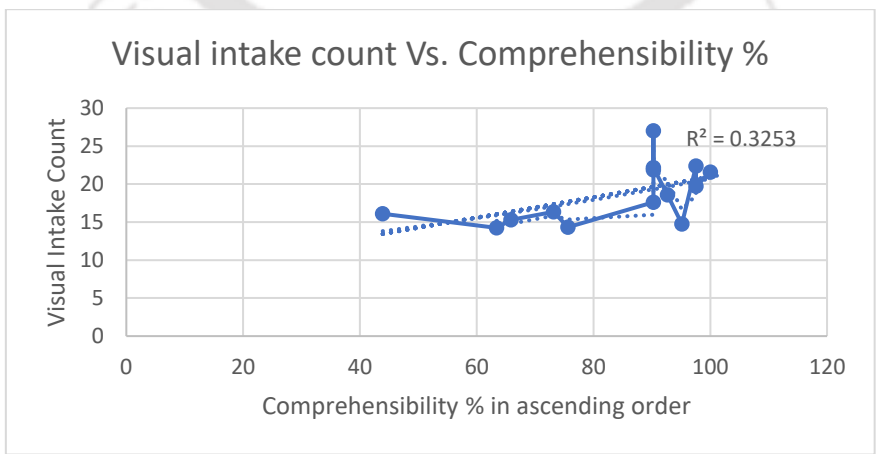


Figure 4. 13. Plot showing a relationship between visual intake count and comprehensibility percentage

Considering fixation (or visual intake) count, and fixation duration, it is seen that fixation count (attention) increases with the increased comprehensibility percentage (Figure 4. 13). As the fixation count increases for a particular AOI, the duration for each fixation decreases. Similarly, an increase in fixation count gave rise to an increase in saccade too. Hence, an increased comprehension level decreases the fixation duration (visual intake duration) (Figure 4. 14).

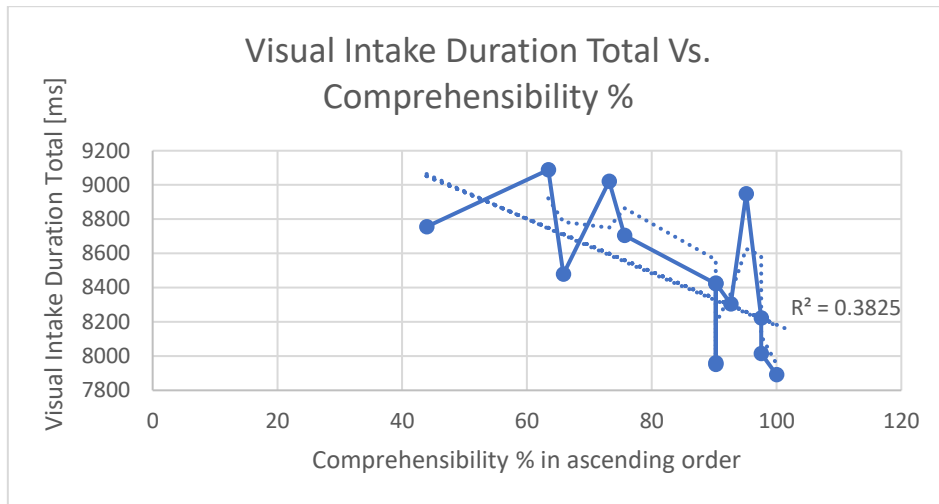


Figure 4. 14. Plot showing a relationship between visual intake duration and comprehensibility percentage

Considering the blink count, blink frequency, and blink duration total, the trend line shows that these measures increase as the comprehensibility percentage increases (Figure 4.15, Figure 4.16, and Figure 4.17). Participants tend to perform less number of blinks while being exposed to a hard-to-comprehend sign since they had to concentrate and focus their attention on the displayed sign to guess the actual meaning of the sign correctly. This result is also in line with the study of Siswandari and Xiong (Siswandari & Xiong, 2015).

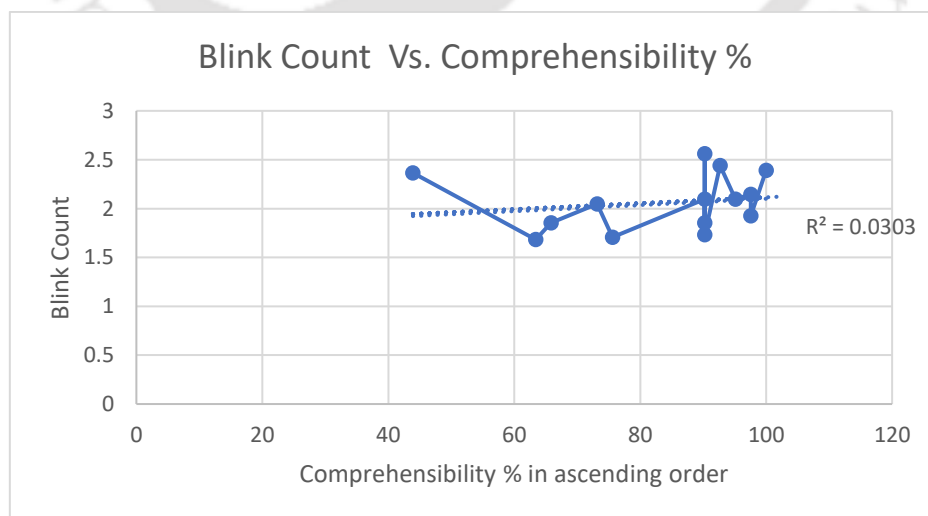


Figure 4. 15. Plot showing a relationship between blink count and comprehensibility percentage

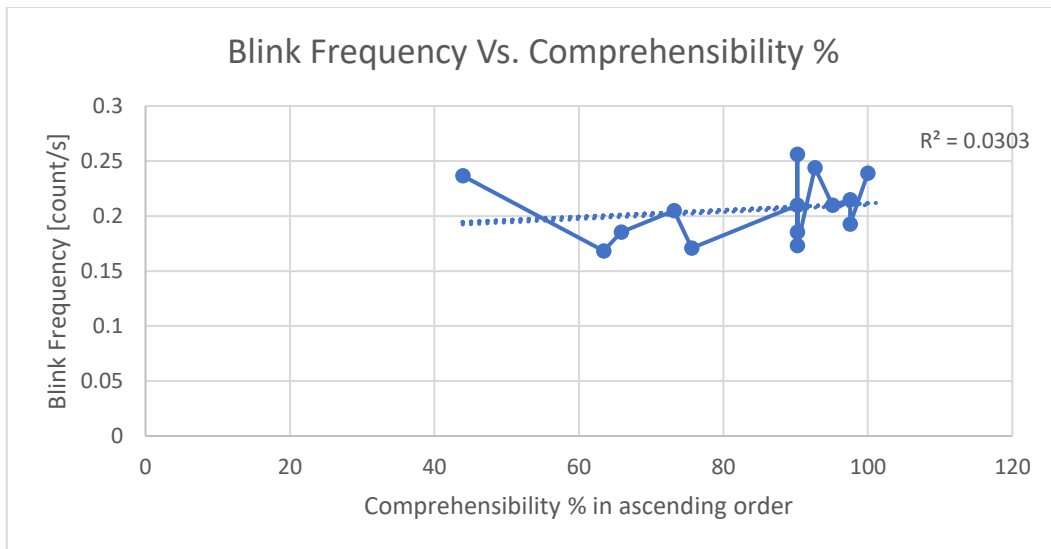


Figure 4. 16. Plot showing a relationship between blink frequency and comprehensibility percentage

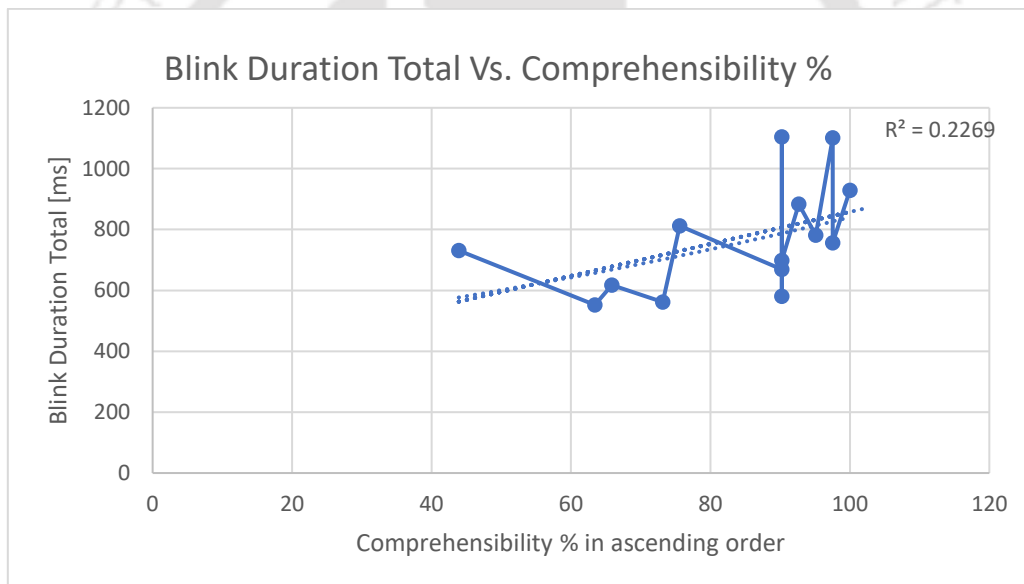


Figure 4. 17. Plot showing a relationship between blink duration total and comprehensibility percentage

Considering the saccade count, saccade amplitude average, and saccade duration total, it is seen that as the comprehension of signs increases, the saccade also increases (Figure 4.18, Figure 4.19, and Figure 4.20). The more comprehensible the sign, the more the saccade count, saccade amplitude, and the saccade duration total. The result is in line with the study of the Krieber groups (Krieber et al., 2016).

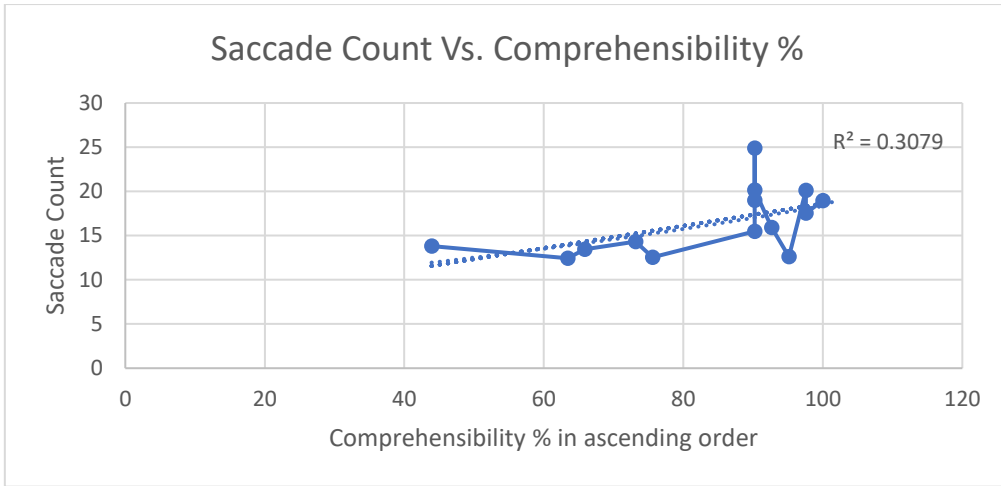


Figure 4. 18. Plot showing a relationship between saccade count and comprehensibility percentage

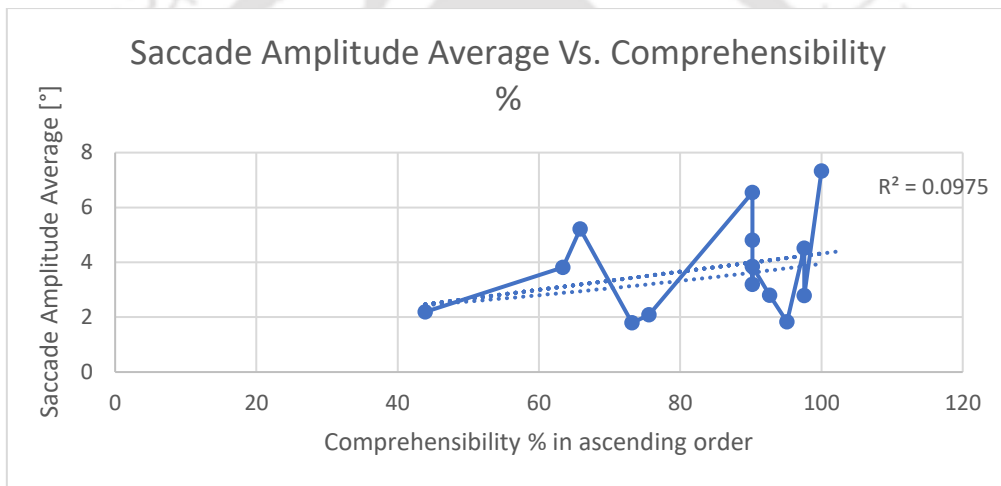


Figure 4. 19. Plot showing a relationship between saccade amplitude average and comprehensibility percentage

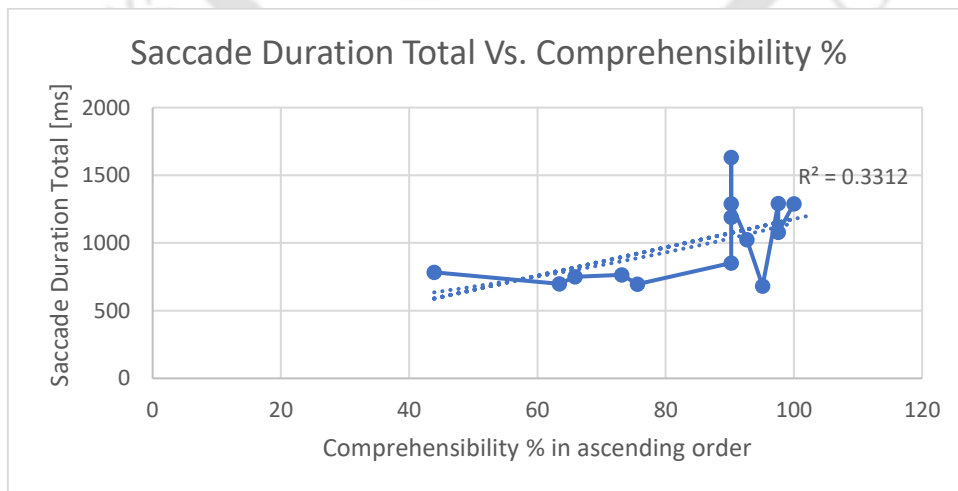


Figure 4. 20. Plot showing a relationship between saccade duration total and comprehensibility percentage

The saccade latency average is less when the sign is comprehensible. *Latency* is the time taken from the appearance of a target to the beginning of a saccade in response to that target. As the sign’s comprehensibility increases, there is a tendency to decrease the saccade latency average (Figure 4.21). As the sign is comprehensible, the participants lose interest, and it jumps to another AOI.

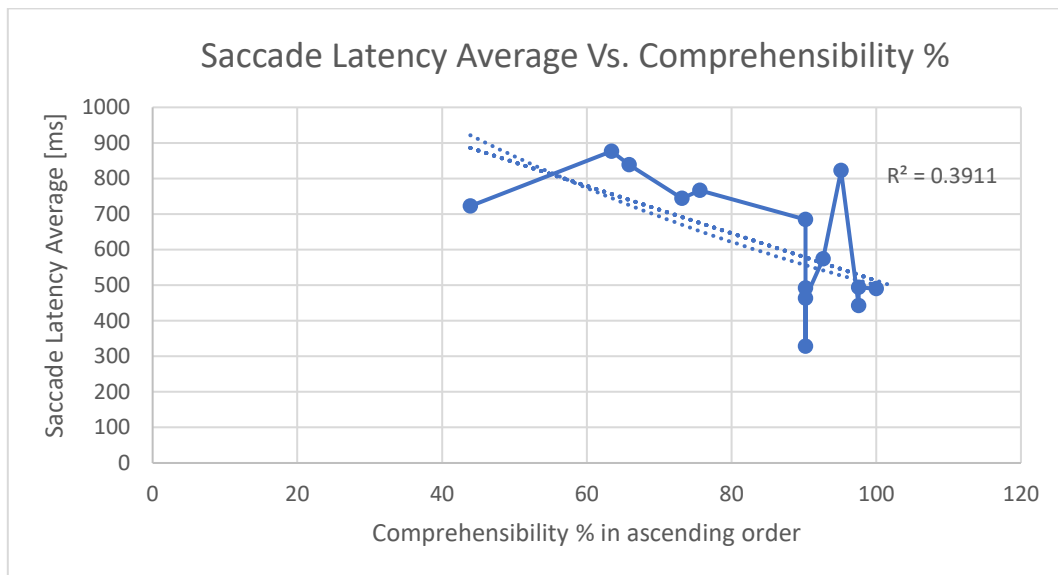


Figure 4. 21. Plot showing a relationship between saccade latency average and comprehensibility percentage

The scan path length tends to increase as the comprehension of signs increases (Figure 4. 22).

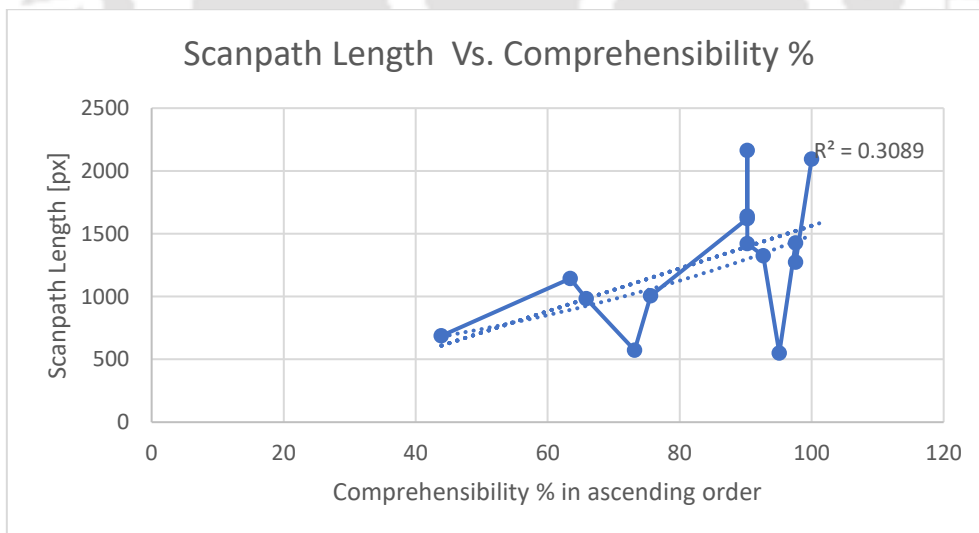


Figure 4. 22. Plot showing a relationship between scan path length and comprehensibility percentage

For all the AOIs, the “Time to First Appearance [ms], Revisits” shows a positive trend with an increased comprehensibility percentage (%). The Average Visual Intake Duration [ms] shows a negative trend with increased comprehensibility percentage (%) for AOI 004 and

AOI 005, in line with the study of Siswandari (Siswandari & Xiong, 2015). However, it shows a positive trend for AOI 006 in line with Hsu groups (Y.-L. Hsu et al., 2013). The increase in Average Visual Intake (fixation) Duration with increased comprehensibility percentage might be due to the text processing time the participants use while trying to comprehend the message.

Conversely, as the Average Visual Intake (fixation) Duration increases, the comprehensibility of that particular text also increases. Hence, the experimental results recommend increasing the average fixation duration to increase any sign's comprehensibility. Simply put, people should spend time viewing the sign to understand or comprehend that sign. This result aligns with the study of Hsu groups (Hsu et al., 2013). The trends for the other variables can be found in Table 4.7.

From the study, it is revealed that blink count, blink frequency, and blink duration; Saccade count, saccade amplitude, saccade amplitude average, and saccade duration; Scan path length; and Revisits increase with the increased sign comprehensibility. Average pupil diameter, Fixation duration, and saccade latency average decrease with increased comprehensibility. Hence, it is concluded that sign comprehensibility has significant effects on real-time nonintrusive eye movements. These findings demonstrate the potential to integrate physiological measures from eye movements with existing evaluation methods to assess OSH signs' comprehensibility. Hence the null hypothesis (*There is no correlation between the subjective response of the comprehensibility and the objective measures (eye tracking variables: average pupil diameter, visual intake count, visual intake duration, blink count, blink duration, blink frequency, saccade count, saccade amplitude, saccade latency, saccade duration, scan path length)*) can be rejected.

Table 4. 7. Trend analysis of eye tracking data

For all AOI	AOI_004 Outer Box					AOI_005 Image Circle					AOI_006 Text Box				
Time to First Appearance [ms]	First Visual Intake Duration [ms]	Revisits	Visual Intake Count	Visual Intake Time [ms]	Average Visual Intake Duration [ms]	First Visual Intake Duration [ms]	Revisits	Visual Intake Count	Visual Intake Time [ms]	Average Visual Intake Duration [ms]	First Visual Intake Duration [ms]	Revisits	Visual Intake Count	Visual Intake Time [ms]	Average Visual Intake Duration [ms]
✓		✓	✓				✓				✓	✓	✓	✓	✓

Note: ✓ signifies the trend increase with increased comprehensibility Percentage [%]

A blank box signifies a negative trend

4.3.2.2. Result Analysis using Bivariate Correlation

A hypothesis has been proposed in section 4.3.2.1, which is repeated below:

Hypothesis H₂: There is a significant correlation between the subjective response of the comprehensibility and the objective measures (eye tracking variables: average pupil diameter, visual intake count, visual intake duration, blink count, blink duration, blink frequency, saccade count, saccade amplitude, saccade latency, saccade duration, scan path length).

Test of hypothesis: Bivariate correlation analysis (Pearson's r , Spearman's ρ , and Kendall's tau- refer to Appendix G) has been conducted to examine the existence of a relationship between the two variables- scores provided by the participants and the eye tracking variables. The analyses prove that there is definitely a significant correlation between the subjective response of the comprehensibility evaluation and the objective measures (eye tracking variables: average pupil diameter, visual intake count, visual intake duration, blink count, blink duration, blink frequency, saccade count, saccade amplitude, saccade latency, saccade duration, scan path length), evident from their p values, thus helping to reject the null hypothesis and proving the alternative hypothesis (proved).

Discussion regarding all the signs with their eye-tracking measures could not be included in the thesis for brevity. The detailed discussion about the bivariate correlation of a few relevant significant signs is discussed here. The bivariate correlation analysis shows that for the sign1 (DNWSH), the objective measure- Saccade velocity Minimum (Sac_velMin) ($^{\circ}/s$) is significant with a p (sig) value of 0.048 (sig.<0.05) where p is 0.05. The strength of the relationship is medium (Cohen), and the value of Spearman's rho coefficient is -.311. The negative sign (-) before the number signifies a negative correlation between the two variables. The result can be interpreted as Spearman's rho correlation coefficient has been conducted to evaluate the null hypothesis that there is no relationship between the 'Saccade velocity Minimum' and the 'Comprehensibility score' (for $N=41$). Preliminary analysis shows that there are violations of normality (shown by Kolmogorov-Smirnov and Shapiro-Wilk Test). There is significant evidence to reject the null hypothesis and conclude that there is a medium, negative association between Saccade velocity Minimum ($M= 18.3366$, $SD= 15.65835$) and Comprehensibility score ($M= 0.8146$, $SD= 0.19693$), $r(41) = -.311$, $p<.05$. So, it can be seen that lower levels of Saccade velocity Minimum are associated with higher levels of Comprehensibility score.

Similarly, 46 variables have been studied for 14 signs, of which 31 are significant. Of these 31 variables, 16 are significant and have a fair correlation (medium) with the comprehensibility score. The rest 15 variables are found to be significant; however, their direction of correlation depends on the comprehensibility of the signs under study.

Variables, for example, Sac_velMin (for signs DNWSH, WAFS, DNT), Blink_durMax (for signs WAFS, DNT_T), Sac_velAvg (for signs DNT, WAFS_T), Blink_cnt (for signs NAFUP_T, DNT_T), Blink_freq (for signs NAFUP_T, DNT_T), Blink_dur (for signs NAFUP_T, DNT_T), Scanpath_len (for signs COSACOEOT, WAFS_T), Sac_amplMin (for

signs NAFUP_T, COSACOEOT_T, UBC_T) and AOI_size004_Revisits (for signs COSACOEOT, WAFS-T) have been found to have a medium positive correlation with comprehensibility score. Again Vis_intk_durMin (for signs NAFUP, WRP_T), Vis_intk_durtot (for signs WAFS, WRP), Vis_intk_durMax (for signs COSACOEOT, WRP), Vis_intk_durAvg (for signs NAFUP, NAFUP_T), AOI_size004_Vis_intkTime (for signs WAFS, WRP, WRP_T), AOI_size005_Vis_intkCnt (for signs WAFS, WRP, COSACOEOT), and AOI_size004First_vis_intkdur (for signs WRP, DNWSH-T, NAFUP-T) have been found to have a medium negative correlation with comprehensibility score. The correlation study results using three methods (Pearson's r, Spearman's, and Kendall's tau) are vast. The significant results are tabulated in Appendix G. The analyses for the correlation similarity between the eye-movement variables and comprehensibility score for different signs have been highlighted with similar colors in Figure 4. 23, Figure 4. 24, and Figure 4. 25, respectively. The text, which is colored similarly but unhighlighted, is found to be a significant feature. However, their direction of correlation depends on the comprehensibility of the signs under study. The original test result of significant eye-tracking variables for the significant signs found by applying bivariate correlation between subjective score and the eye-tracking variables can be found in Appendix G.



									COSAC			
									OEOT_	UBC_		
DNWSH	NAFUP	WAFS	COSACOET	UBC WRP	DNT	DNWSH-T	NAFUP_T	WAFS_T	T	T	WRP_T	DNT_T
Sac_vel Min (sp)- M*	Vis_intk_dur Min (p)+M*; (k)+M**	Vis_intk_durt ot (p)+M*	vis_intk_cnt (p)-M*; (k)- S*; (sp)-M*	Vis_intk_d urMax (p)+M*	Sac_dur Avg (p)- M*	Sac_dur Max (k)+S*	Blink_cnt (p)-M*	Sac_vel Avg (p)- M*			Vis_intk durMin (k)+M*; (sp)+M**	Blink_cnt (p)- M*; (k)-M**; (sp)-M**
	Vis_intk_dur Avg (k)+S*	Sac_cnt (p)+M*; (k)+S*; (sp)+M*	vis_intk_fre q (p)-M*; (k)-S*; (sp)- M*	Blink_dur Min (p)+M**; (k)+M**; (sp)+M**;	Sac_dur Min (p)- M*		Blink_freq (p)-M*	Scanpat h_len (p)-M*				Blink_freq (p)- M*; (k)-M**; (sp)-M**
	Sac_durMax (k)-S*	Sac_freq (p)+M*; (k)+S*; (sp)+M*	Vis_intk_dur Max (p)+M*	Vis_intk_d urtot (k)+S*; (sp)+M*	Sac_velA vg (p)- M**		Blink_dur (p)-M*	Sac_am plMax (k)+S*; (sp)+M *; ,				Blink_dur (k)- M*; (sp)-M*
		Sac_velMin (p)-M*	Sac_cnt (p)- M*; (sp)- M**	Blink_durA vg (k)+S*	Sac_vel Max (p)- M*		Vis_intk_du rAvg (p)+S*; (sp)+M*		(sp): Spearman's rho; (p): Pearson and (k): Kendall's Tau_b correlation +: positive, -: negative S: Small correlation, M: Medium correlation strength and, L: large correlation strength Here Alpha is 0.05 is represented by * and 0.01 is represented by **			Blink_durMax (sp)-M*
		Blink_durAvg (p)-M**; (k)- M**; (sp)- M**	Sac_freq (p)-M*; (sp)- M**		Sac_vel Min (p)- M*		Sac_durAvg (k)+S*; (sp)+M*					
		Blink_durMa x (p)-M**; (k)-S*; (sp)- M**	Sac_durtot (p)-M*; (k)- M**; (sp)- M**				Sac_durMa x (k)+S*; (sp)+M*					

Figure 4. 23. Abstract view of analysis for event statistics – correlation of objective measures with subjective scores (eye tracking technique)

DNWSH NAFUP	WAFS	COSACOEOT	UBC	WRP	DNT	DNWSH-T	NAFUP_T	WAFS_T	COSACOEOT_T	UBC_T	WRP_T	DNT_T
	Blink_durMin (p)-M**; (k)-S*; (sp)-M**	Sac_ampl tot (p)- M*					Sac_velMa x (k)+S*; (sp)+M*					
	vis_intk_cnt (k)+S*; (sp)+M*	Sac_latAv g (p)+M*					Sac_ampl Min (k)- S*;		Sac_amplMin (p)- M*; (k)-S*; (sp)- M*;	Sac_ampl Min (k)- S*;		
	vis_intk_freq (k)+S*; (sp)+M*	Scanpath _len (p)- M**; (k)- M**; (sp)-M**;										
	Sac_ampltot (k)+S*; (sp)+M*	Sac_ampl tot (k)- M**; (sp)-M**;					(sp): Spearman's rho; (p): Pearson and (k): Kendall's Tau_b correlation +: positive, -: negative S: Small correlation, M: Medium correlation strength and, L: large correlation strength Here Alpha is 0.05 is represented by * and 0.01 is represented by **					
	Sac_amplMin (k)+S*; (sp)+M*											
	Sac_veltot (k)+S*; (sp)+M*											
	Scanpath_len (k)+S*											
	Sac_amplAvg (sp)+M*											

Figure 4. 24. Abstract view of Analysis for event statistics – correlation of objective measures with subjective scores continued (eye-tracking technique)

DNWSH	NAFUP	WAFS	COSACOEOT	UBC	WRP	DNT	DNWSH-T	NAFUP_T	WAFS_T	COSACOEOT_T	UBC_T	WRP_T	DNT_T
AOI_size006_Vis_intkTime (p)-M*	AOI_size004_Vis_intkCnt (p)+M*; (k)+S*; (sp)+M*	AOI_size004_Revisits (p)-M*; (k)-M*; (sp)-M*	AOI_size004_Revisits (p)-M*; (k)-M*; (sp)-M*		AOI_size004_First_vis_intkdur (p)+M*		AOI_size004_First_vis_intkdur (k)+S*; (sp)+M*	AOI_size004_4First_vis_intkdur (k)+S*; (sp)+M*	AOI_size004_Revisits (p)-M*	AOI_size006_Vis_intkTime (k)+S*; (sp)+M*		AOI_size004_Vis_intkTime (k)+S*; (sp)+M*	AOI_size006_Avg_vis_intkdur (k)+S*
	AOI_size004_Vis_intkTime (p)+M*	AOI_size004_Vis_intkCnt (p)-M*; (k)-S*; (sp)-M*			AOI_size005_First_vis_intkdur (p)+M*			AOI_size004_4_Avg_vis_intkdur (sp)+M*		AOI_size006_Vis_intkCnt (sp)+M*			
	AOI_size005_Vis_intkCnt (k)+S*; (sp)+M*	AOI_size005_Vis_intkCnt (p)-M*			TimeToFirstAppear_allAOI (p)-M*								
					AOI_size004_Vis_intkTime (k)+S*; (sp)+M*								
					AOI_size005_Vis_intkCnt (sp)-M*								

Figure 4. 25. Abstract view of correlation and significance analysis for AOI statistics (eye-tracking technique)

In the above figures (Figure 4. 23, Figure 4. 24, and Figure 4.25), a few measures/ features are seen with colored text but not highlighted. The plus (+) or minus sign (-) shows the direction of the correlation between a measure and comprehensibility score (C_Score) depending on whether a sign is comprehensible or incomprehensible. For example, consider the objective measure, saccade amplitude total "Sac_ampltot" for the sign COSACOEOT and WAFS. COSACOEOT, one of the least comprehensible signs, shows a negative correlation with the comprehensibility score evident from the direction of correlation. Similarly, let us consider the sign WAFS. The "Sac_ampltot" shows a positive correlation direction with a comprehensibility score as the sign is comprehensible. For the sign WAFS, the result is interpreted as Kendall's Tau_b correlation has been conducted to evaluate the null hypothesis that there is no relationship between the 'Saccade amplitude total' and the 'Comprehensibility score' (N=41). Preliminary analysis shows that there are violations of normality (shown by Kolmogorov-Smirnov and Sapiro-Wilk Test). There is significant evidence to reject the null hypothesis and conclude that there is a small, positive association between Saccade amplitude total (M= 87.4, SD= 317.55309) and Comprehensibility score (M= 0.7805, SD= 0.24106), $r(41) = .238^*$, $p < .05$. Higher levels of Saccade amplitude total are associated with higher levels of Comprehensibility score. The same conclusion is drawn from Spearman's correlation analysis. This reveals that for a comprehensible sign like WAFS, there seems to be a positive correlation between Sac_ampltot and C_Score.

For the sign COSACOEOT, Kendall's Tau_b correlation has been conducted to evaluate the null hypothesis that no relationship exists between the 'Saccade amplitude total' and the 'Comprehensibility score' (N=41). Preliminary analysis shows that there are violations of normality (Kolmogorov-Smirnov and Sapiro-Wilk Test). There is significant evidence to reject the null hypothesis and conclude that there is a medium, negative association between Saccade amplitude total (M= 30.6976, SD= 25.88928) and Comprehensibility score (M= 0.3805, SD= 0.29514), $r(41) = -.357^{**}$, $p < .01$. Lower levels of Saccade amplitude total are associated with higher levels of Comprehensibility score. The same conclusion is drawn from the result of Spearman's correlation analysis. The result reveals that for a comprehensible sign like COSACOEOT, there seems to be a negative correlation between Sac_ampltot and C_Score.

The outcome of the significant parameters of event statistics in eye tracking and their relative direction is shown below.

- Eye-tracking measures negatively correlated with the comprehensibility score (Medium):
- * Sac_velMin (DNWSH (M= 18.3; SD= 15.6), WAFS (M= 17.11; SD= 6.35), DNT (M= 44.7; SD= 182))
- * Blink_durMax (WAFS (M= 231.7; SD= 332.4), DNT_T (M= 308.8; SD= 369.5))
- * Sac_velAvg (DNT (M= 114; SD= 309), WAFS_T (M= 75.5; SD= 140))
- * Blink_cnt (NAFUP_T (M=2.14; SD=2.60), DNT_T (M= 2.43; SD= 2.39))
- * Blink_freq (NAFUP_T (M=0.24; SD=0.26), DNT_T (M=0.24; SD=0.23))
- * Blink_dur (NAFUP_T (M=1100; SD=2725), DNT_T (M=883.9; SD=1437))

- * Scanpath_len (COSACOEOT (M= 686.3; SD= 729.3), WAFS_T (M= 1274; SD=1144))
- * Sac_amplMin (NAFUP_T (M= 0.60; SD=0.33), COSACOEOT_T (M= 0.52; SD= 0.27), UBC_T (M= 0.59; SD= 0.47))
- Eye-tracking measures positively correlated with the comprehensibility score (Medium):
- * Vis_intk_durMin (NAFUP (M= 176; SD=171.7), WRP_T (M=97.78; SD=40.56))
- * Vis_intk_durtot (WAFS (M= 8422 ; SD=1679), WRP (M= 8479; SD= 1509))
- * Vis_intk_durMax (COSACOEOT (M=2547; SD=2873), WRP (M=2619; SD= 2065))
- Eye-tracking measures positively correlated with comprehensibility score (Small, Medium):
- * Vis_intk_durAvg (NAFUP (M=767; SD= 812), NAFUP_T (M=419; SD=246))

The outcome of the significant parameters of AOI statistics in eye tracking and their relative direction is as follows.

- Eye-tracking measures negatively correlated with the comprehensibility score (Medium)
- * AOI_size004_Vis_intkTime (WAFS (M=8383; SD=1791), WRP (M=8469; SD= 1536), WRP_T (M= 8124; SD= 3319))
- * AOI_size005_Vis_intkCnt (WAFS (M=16.97; SD= 7.14), WRP (M=14.68; SD= 7.82), COSACOEOT (M= 15.41; SD=7.49))
- * AOI_size004First_vis_intkdur (WRP (M=1038; SD= 1495), DNWSH-T (M=260.5; SD= 283.5), NAFUP-T (M= 230.5; SD=171.4))
- Eye-tracking measures positively correlated with the comprehensibility score (Medium)
- * AOI_size004_Revisits (COSACOEOT (M= 0.12; SD=0.45), WAFS-T (M= 0.02; SD= 0.15))

Table 4.6 reports the comprehensibility percentage of the participants based on subjective data. The table analysis shows that signs with text performed better than signs without text. The hypothesis is supported by a Wilcoxon signed rank (non-parametric) test (section 4.3.4) to check whether there is any truth in the result.

4.3.3. Analysis of variance (ANOVA)

ANOVA has been conducted to study variances of comprehensibility between all the sign pairs viewed by the same group of participants. All the signs (subjective data) are tested using

Friedman’s ANOVA followed by post hoc analysis. Friedman is a non-parametric test where each observation's median and interquartile range are reported instead of means and standard deviations. The result of Friedman's ANOVA test has been tabulated (Table 4. 8 and Table 4. 9). Friedman’s ANOVA calculates overall statistical significance (significant differences) for all signs together and not as pairs. So, for our 14 signs, it will give a p-value for the significant difference between all the signs together and not specify which two groups are significant and which are not. Friedman, followed by post hoc analysis, has been conducted to find which two signs are significantly different among themselves.

Table 4. 8. Rank of the signs generated by Friedman's ANOVA (subjective data- eye-tracking technique)

Ranks	
Sign_Techniques	Mean Rank
DNWSH_score	8.45
NAFUP_score	5.44
WAFS_score	8.13
COSACOEOT_score	3.09
UBC_score	5.15
WRP_score	5.22
DNT_score	5.65
DNWSH_T_score	10.16
NAFUP_T_score	9.65
WAFS_T_score	10.09
COSACOEOT_T_score	7.51
UBC_T_score	8.43
WRP_T_score	8.73
DNT_T_score	9.32

Table 4. 9. Test statistics of Friedman's ANOVA (subjective data- eye-tracking technique)

Test Statistics ^(a)	
N	41
Chi-Square	168.919
df	13
Asymp. Sig.	.000

a. Friedman Test

The test calculated the p-value (Table 4. 9), which is less than 0.05 and is significant here. The result shows significant differences between all the signs considered for the study. A post hoc analysis (Figure 4. 26) following the Friedman test helps to identify the significant and non-significant pairs. The result (Figures 4.27 and 4.28) shows that thirty-two (32) pairs

are significantly different, and the pairs are highlighted in yellow (p-values- Adj. Sig) in the figures.

Overall significance shows that there is a significant difference between the signs. Pairwise analysis shows that the comprehensibility level of COSACOEOT significantly differs from DNWSH_T, WAFS_T, NAFUP_T, DNT_T, WRP_T, UBC_T (Adj. sig .000). Similarly, the pairwise significant differences for all the signs can be seen from the Figures 4.27 and 4.28).

Hypothesis Test Summary

Null Hypothesis	Test	Sig.	Decision
1 The distributions of DNWSH_score, NAFUP_score, WAFS_score, COSACOEOT_score, UBC_score, WRP_score, DNT_score, DNWSH_T_score, NAFUP_T_score, WAFS_T_score, COSACOEOT_T_score, UBC_T_score, WRP_T_score and DNT_T_score are the same.	Related-Samples Friedman's Two-Way Analysis of Variance by Ranks	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 4. 26. Post hoc analysis (subjective data- eye-tracking technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
COSACOEOT_score-DNWSH_T_score	-7.073	.924	-7.655	.000	.000
COSACOEOT_score-WAFS_T_score	-7.000	.924	-7.576	.000	.000
COSACOEOT_score-NAFUP_T_score	-6.561	.924	-7.101	.000	.000
COSACOEOT_score-DNT_T_score	-6.232	.924	-6.745	.000	.000
COSACOEOT_score-WRP_T_score	-5.646	.924	-6.111	.000	.000
COSACOEOT_score-UBC_T_score	-5.341	.924	-5.781	.000	.000
UBC_score-DNWSH_T_score	-5.012	.924	-5.425	.000	.000
WRP_score-DNWSH_T_score	-4.939	.924	-5.346	.000	.000
UBC_score-WAFS_T_score	-4.939	.924	-5.346	.000	.000
WRP_score-WAFS_T_score	-4.866	.924	-5.266	.000	.000
NAFUP_score-DNWSH_T_score	-4.720	.924	-5.108	.000	.000
NAFUP_score-WAFS_T_score	-4.646	.924	-5.029	.000	.000
DNT_score-DNWSH_T_score	-4.512	.924	-4.884	.000	.000
UBC_score-NAFUP_T_score	-4.500	.924	-4.870	.000	.000
DNT_score-WAFS_T_score	-4.439	.924	-4.804	.000	.000

Figure 4. 27. Post hoc analysis result showing significantly different pairs in sequence-1 (subjective data-eye-tracking technique)

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
WRP_score-NAFUP_T_score	-4.427	.924	-4.791	.000	.000
COSACOEOT_score-COSACOEOT_T_score	-4.427	.924	-4.791	.000	.000
NAFUP_score-NAFUP_T_score	-4.207	.924	-4.554	.000	.000
UBC_score-DNT_T_score	-4.171	.924	-4.514	.000	.001
WRP_score-DNT_T_score	-4.098	.924	-4.435	.000	.001
DNT_score-NAFUP_T_score	-4.000	.924	-4.329	.000	.001
NAFUP_score-DNT_T_score	-3.878	.924	-4.197	.000	.002
DNT_score-DNT_T_score	-3.671	.924	-3.973	.000	.006
UBC_score-WRP_T_score	-3.585	.924	-3.881	.000	.009
WRP_score-WRP_T_score	-3.512	.924	-3.801	.000	.013
NAFUP_score-WRP_T_score	-3.293	.924	-3.564	.000	.033
UBC_score-UBC_T_score	-3.280	.924	-3.551	.000	.035
WRP_score-UBC_T_score	-3.207	.924	-3.471	.001	.047
WRP_score-DNWSH_score	3.232	.924	3.498	.000	.043
UBC_score-DNWSH_score	3.305	.924	3.577	.000	.032
COSACOEOT_score-WAFS_score	5.049	.924	5.464	.000	.000
COSACOEOT_score-DNWSH_score	5.366	.924	5.808	.000	.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .05. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Figure 4. 28. Post hoc analysis result showing significantly different pairs in sequence-2 (subjective data-eye-tracking technique)

4.3.4. OSH sign with OSH-related textual information affects the understanding or comprehensibility of the messages intended to convey (eye-tracking technique)

The analysis of Table 4. 6 shows that signs with text performed better than signs without text. A hypothesis test- Wilcoxon signed-rank (non-parametric test) further checks whether there is any truth in the result.

Null Hypothesis (H_0): The level of comprehensibility of the OSH sign is not significantly different when the sign is presented with or without textual information.

Alternate Hypothesis (H_3): The level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information.

Test of Hypothesis: To test the hypothesis, a Wilcoxon signed rank test has been conducted, and the result is shown in the following tables. Descriptive statistics (Table 4. 10) shows the mean of all the signs, indicating the significant difference between the signs with and without textual information. The mean of the signs with text is more than the mean of the signs without text. This difference is significant, and the result is the same as expected from this technique. The minimum and maximum values signify the minimum and maximum scores assigned for that sign under study.

Table 4. 10. Descriptive statistics for the subjective comprehensibility score in the Wilcoxon Signed Rank test (eye-tracking technique)

Descriptive Statistics					
	N	Mean	Std. Deviation	Minimum	Maximum
COSACOEOT_score	41	.3805	.29514	.00	1.00
COSACOEOT_T_score	41	.7537	.24093	.00	1.00
DNT_score	41	.6122	.27404	.00	1.00
DNT_T_score	41	.8317	.20669	.20	1.00
DNWSH_score	41	.8146	.19693	.20	1.00
DNWSH_T_score	41	.9024	.12142	.60	1.00
NAFUP_score	41	.6195	.25221	.00	1.00
NAFUP_T_score	41	.8585	.16275	.20	1.00
UBC_score	41	.6220	.25051	.10	1.00
UBC_T_score	41	.8000	.21564	.20	1.00
WAFS_score	41	.7805	.24106	.00	1.00
WAFS_T_score	41	.8756	.16550	.20	1.00
WRP_score	41	.5878	.31559	.00	1.00
WRP_T_score	41	.8171	.20965	.20	1.00

Considering the sign pair COSACOEOT and COSACOEOT_T (Table 4.11), the result shows that out of 41 participants, 2 of them gave less rank to COSACOEOT_T, where 32 of them gave more rank to COSECOEOT_T (That is 32 cases, the COSECOEOT_T is ranked more than COSECOEOT). For all the sign pairs, it is seen that the number of ‘positive’ ranks is higher than the number of ‘negative’ ranks, and the sum of ranks for all the ‘positive’ ranks is much higher than the sum of ranks of all the ‘negative’ ranks. This result matches the expectation, and these results support our hypothesis.

Table 4. 11. Wilcoxon Signed Rank test ranks for the sign pairs: without and with text (subjective data-eye-tracking technique)

S.No	Sign	Negative Ranks			Positive Ranks			Ties
		N	Mean Rank	Sum of Ranks	N	Mean Rank	Sum of Ranks	
1	COSACOEOT_score - COSACOEOT_T_score	2 ^a	4.00	8.00	32 ^b	18.34	587.00	7 ^c
2	DNT_T_score - DNT_score	4 ^a	9.75	39.00	29 ^b	18.00	522.00	8 ^c
3	DNWSH_T_score - DNWSH_score	6 ^a	5.50	33.00	17 ^b	14.29	243.00	18 ^c
4	NAFUP_T_score - NAFUP_score	5 ^a	10.20	51.00	30 ^b	19.30	579.00	6 ^c
5	UBC_T_score - UBC_score	4 ^a	11.63	46.50	27 ^b	16.65	449.50	10 ^c
6	WAFS_T_score - WAFS_score	3 ^a	5.50	16.50	16 ^b	10.84	173.50	22 ^c
7	WRP_T_score - WRP_score	2 ^a	7.00	14.00	30 ^b	17.13	514.00	9 ^c

- a. COSACOEOT_T_score < COSACOEOT_score
- b. COSACOEOT_T_score > COSACOEOT_score
- c. COSACOEOT_T_score = COSACOEOT_score

- a. DNT_T_score < DNT_score
- b. DNT_T_score > DNT_score
- c. DNT_T_score = DNT_score

- a. DNWSH_T_score < DNWSH_score
- b. DNWSH_T_score > DNWSH_score
- c. DNWSH_T_score = DNWSH_score

- a. NAFUP_T_score < NAFUP_score
- b. NAFUP_T_score > NAFUP_score
- c. NAFUP_T_score = NAFUP_score

- a. WRP_T_score < WRP_score
- b. WRP_T_score > WRP_score
- c. WRP_T_score = WRP_score

- a. UBC_T_score < UBC_score
- b. UBC_T_score > UBC_score
- c. UBC_T_score = UBC_score

- a. WAFS_T_score < WAFS_score
- b. WAFS_T_score > WAFS_score
- c. WAFS_T_score = WAFS_score

To further strengthen our assumption, we calculated the effect size using the Z value (Table 4. 10). Table 4. 11 shows that the p-value (significance value) is less than 0.05, which is statistically significant; hence we can reject the null hypothesis (H₃) that OSH sign with OSH-related textual information does not affect the understanding or comprehensibility of the messages intended to convey.

The effect size is calculated using the following formula,

$$r = \frac{Z}{\sqrt{N}} \quad (4.1)$$

where Z is the z-score of the Wilcoxon test, and N is the size of the study (number of total observations). The effect size (z-score) is calculated for all seven (7) pairs of signs- with and without text (Table 4. 11), and they reveal that the effect size is significant for all the pairs, which is excellent for our study.

Table 4. 12. Test statistics for Wilcoxon Signed Rank test for the sign pairs: without and with text (eye-tracking technique)

Test Statistics ^(a)							
	COSACOEOT_T_score - - COSACOEOT_score	DNT_T_score - DNT_score	DNWSH_T_score - - DNWSH_score	NAFUP_T_score - NAFUP_score	UBC_T_score - UBC_score	WAFS_T_score - WAFS_score	WRP_T_score - WRP_score
Z	-4.960 ^b	-4.340 ^b	-3.232 ^b	-4.344 ^b	-3.976 ^b	-3.215 ^b	-4.716 ^b
Asymp. Sig. (2-tailed)	.000	.000	.001	.000	.000	.001	.000

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Table 4. 13. Effect Size for Wilcoxon Signed Rank test for the sign pairs: without and with text (eye-tracking technique)

	COSACOEOT_T_score - COSACOEOT_score	DNT_T_score - DNT_score	DNWSH_T_score - - DNWSH_score	NAFUP_T_score - NAFUP_score	UBC_T_score - UBC_score	WAFS_T_score - WAFS_score	WRP_T_score - WRP_score
Z	-4.96	-4.34	-3.232	-4.344	-3.976	-3.215	-4.716
N	82	82	82	82	82	82	82
Root N	9.055385138	9.055385138	9.055385138	9.055385138	9.055385138	9.055385138	9.055385138
r	-0.547740369	-0.479272823	-0.356914692	-0.479714549	-0.439075748	-0.355037356	-0.520795077

The result can be interpreted as:

- * Wilcoxon signed-rank test reveals that comprehensibility scores for the OSH sign with textual information (COSACOEOT_T) are found to be significantly higher compared to the sign without textual information (COSACOEOT) (M= .3805; SD=.29514) compared to sign with text with (M= .7537; SD= .24093), $z = -4.960$, $p = .000$, with large effect size, $r = -0.547740369$.
- * Wilcoxon signed-rank test reveals that comprehensibility scores for OSH signs with textual information (DNT_T) are found to be significantly higher compared to the signs without textual information (DNT) with (M= .6122; SD= .27404) compared to signs with text (M= .8317; SD= .20669), $z = -4.340$, $p = .000$, with large effect size, $r = -0.479272823$.
- * Wilcoxon signed-rank test reveals that comprehensibility scores for OSH signs with textual information (DNWSH_T) are found to be significantly higher compared to the signs without textual information (DNWSH) with (M= .8146; SD= .19693) compared to signs with text (M= .9024; SD= .12142), $z = -3.232$, $p = .000$, with large effect size, $r = -0.356914692$.
- * Wilcoxon signed-rank test reveals that comprehensibility scores for OSH signs with textual information (NAFUP_T) are found to be significantly higher compared to the signs without textual information (NAFUP) with (M= .6195; SD= .25221) compared to signs with text (M= .8585; SD= .16275), $z = -4.344$, $p = .000$, with large effect size, $r = -0.479714549$.
- * Wilcoxon signed-rank test reveals that comprehensibility scores for OSH signs with textual information (UBC_T) are found to be significantly higher compared to the signs without textual information (UBC) with (M= .6220; SD= .25051) compared to signs with text (M= .8000; SD= .21564), $z = -3.976$, $p = .000$, with large effect size, $r = -0.439075748$.
- * Wilcoxon signed-rank test reveals that comprehensibility scores for OSH signs with textual information (WAFS_T) are found to be significantly higher compared to the signs without textual information (WAFS) with (M= .7805; SD= .24106) compared to signs with text (M= .8756; SD= .16550), $z = -3.215$, $p = .000$, with large effect size, $r = -0.355037356$.
- * Wilcoxon signed-rank test reveals that comprehensibility scores for OSH signs with textual information (WRP_T) are found to be significantly higher compared to the signs without textual information (WRP) with (M= .5878; SD= .31559) compared to signs with text (M= .8171; SD= .20965), $z = -4.716$, $p = .000$, with large effect size, $r = -0.520795077$.
- * Hence it could be concluded that the level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information.

4.3.5. ANOVA of objective data for the comprehensible and incomprehensible signs

Considering the signs NAFUP, UBC, DNWSH_T, WAFS_T, NAFUP_T, DNWSH, DNT_T, WRP_T, UBC_T, COSACOEOT_T, and WAFS as comprehensible (based on 67% Gold standard by ISO); and COSACOEOT, DNT, and WRP, as non-comprehensible, two groups of signs has been formed for analysis of variances for the objective measures collected from the eye-tracking experiment.

Kruskal-Wallis test has been performed, which calculates significant differences between the two groups based on ranking the data. A Post hoc test performs multiple comparisons between all combinations of groups. It is found that the eye-tracking variables such as *Vis_intk_durAvg*, *Blink_durAvg*, *AOI004Firstvis_intk_dur*, *AOI004Avg_vis_intk_dur*, *AOI005Firstvis_intk_dur*, and *AOI005Avg_vis_intk_dur* are not same across categories of comprehensibility levels (comprehensible and incomprehensible signs), supported by the p-value which is less than 0.05 which is significant.

The descriptive statistics (Table 4.14), ranks (Table 4.15), test statistics (Table 4.16) effect size (Table 4.17), and hypothesis summary (Figure 4.29) for the above-selected features are provided for brevity.

Table 4.14. Descriptive statistics for the significant objective measures-Kruskal-Wallis test (eye-tracking technique)

Descriptive Statistics					
	N	Mean	Std. Deviation	Minimum	Maximum
Vis_intk_durAvg	14	675.4709	264.79518	315.87	1159.20
Blink_durAvg	14	624.1547	175.00701	328.27	876.47
AOI004Firstvis_intk_dur	14	465.8387	267.84957	222.00	1038.59
AOI004Avg_vis_intk_dur	14	668.4202	262.57897	318.20	1158.28
AOI005Firstvis_intk_dur	14	526.3577	241.59053	239.03	1039.00
AOI005Avg_vis_intk_dur	14	745.8697	205.21861	532.49	1164.04
Comprehensibility_levels	14	.79	.426	0	1

Table 4.15. Rank for the significant objective measures-Kruskal-Wallis test (eye-tracking technique)

Ranks			
	Comprehensibility_levels	N	Mean Rank
Vis_intk_durAvg	Incomprehensible	3	12.00
	Comprehensible	11	6.27
	Total	14	
Blink_durAvg	Incomprehensible	3	12.00
	Comprehensible	11	6.27
	Total	14	

AOI004Firstvis_intk _dur	Incomprehensible	3	13.00
	Comprehensible	11	6.00
	Total	14	
AOI004Avg_vis_intk _dur	Incomprehensible	3	12.33
	Comprehensible	11	6.18
	Total	14	
AOI005Firstvis_intk _dur	Incomprehensible	3	12.67
	Comprehensible	11	6.09
	Total	14	
AOI005Avg_vis_intk _dur	Incomprehensible	3	12.00
	Comprehensible	11	6.27
	Total	14	

Table 4. 16. Test statistics for the significant objective measures-Kruskal-Wallis test (eye-tracking technique)

Test Statistics^(a,b)

	AOI004Firstvis_i Vis_intk_durAvg	AOI004Avg_vis_i Blink_durAvg	AOI005Firstvis_i ntk_dur	AOI005Avg_vis_i ntk_dur	AOI005Firstvis_i ntk_dur	AOI005Avg_vis_i ntk_dur
Kruskal-Wallis H	4.418	4.418	6.600	5.097	5.824	4.418
Df	1	1	1	1	1	1
Asymp. Sig.	.036	.036	.010	.024	.016	.036

a. Kruskal Wallis Test

b. Grouping Variable: Comprehensibility_levels

Table 4. 17. Effect Size for Kruskal-Wallis test for the variation of objective measures across categories

	AOI004Firstvis_ Vis_intk_durAvg	AOI004Avg_vis_ Blink_durAvg	AOI005Firstvis_ intk_dur	AOI005Avg_vis_ _intk_dur	AOI005Firstvis_ _intk_dur	AOI005Avg_vis_ _intk_dur
Kruskal-Wallis H	4.418	4.418	6.600	5.097	5.824	4.418
K (nos of groups)	2	2	2	2	2	2
n (nos of obs)	82	82	82	82	82	82
η^2 (Eta squared) = (H - k + 1)/(n - k)	0.042725	0.042725	0.07	0.051213	0.0603	0.042725
Percentage of Variance= η^2 *100	4.2725	4.2725	7	5.12125	6.03	4.2725

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
4	The distribution of Vis_inrk_durAvg is the same across categories of Comprehensibility_levels.	Independent-Samples Kruskal-Wallis Test	.036	Reject the null hypothesis.
4	The distribution of Blink_durAvg is the same across categories of Comprehensibility_levels.	Independent-Samples Kruskal-Wallis Test	.036	Reject the null hypothesis.
1	The distribution of AOI004Firstvis_inrk_dur is the same across categories of Comprehensibility_levels.	Independent-Samples Kruskal-Wallis Test	.010	Reject the null hypothesis.
5	The distribution of AOI004Avg_vis_inrk_dur is the same across categories of Comprehensibility_levels.	Independent-Samples Kruskal-Wallis Test	.024	Reject the null hypothesis.
1	The distribution of AOI005Firstvis_inrk_dur is the same across categories of Comprehensibility_levels.	Independent-Samples Kruskal-Wallis Test	.016	Reject the null hypothesis.
5	The distribution of AOI005Avg_vis_inrk_dur is the same across categories of Comprehensibility_levels.	Independent-Samples Kruskal-Wallis Test	.036	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 4. 29. Hypothesis test summary for the significant objective measures-Kruskal-Wallis test (eye-tracking technique)

4.4. Conclusion


Literature reveals numerous eye-tracking experiments; however, there is a lacuna in the study of sign comprehensibility and its correlation with the physiological measures of human eyes. A study has been conducted to capture real-time physiological data of eye movements during the comprehension of fourteen (14) OSH signs by forty-one (41) participants. Semantic transparency is significant in this study and is maintained in the experiment by involving novice students. The philosophy is that if novices can comprehend any sign, the professionals who work in organizations can easily do so. The comprehensibility percentage has been calculated using subjective data. Considering the comprehensibility level, a vital component in studying the association between variables, this has been thoroughly assessed based on the 67% threshold (ISO standard). The subject-based data has been used to check the significant difference in comprehensibility levels among signs. Variability is analyzed using Friedman’s test followed by post hoc analysis, and it is found that forty-nine (49) pairs are significantly different from each other. The same data has been used for testing whether OSH sign with OSH-related textual information affects the understanding or comprehensibility of the messages intended to convey, and it is found to be true for all the seven pairs of signs. The evaluation uses the Wilcoxon signed rank test supported by effect size, thus supporting our hypothesis.

The data collected by both methods (subject-based score and eye-tracking) are tested for its normality. The data is non-uniform. A correlation plot was used for the correlation study, and it elicited that the average pupil diameter, fixation duration (visual intake duration), and

saccade latency average decreased with the increased comprehensibility of the signs. Similarly, blink count, blink frequency, and blink duration; saccade count, saccade amplitude, saccade amplitude average, and saccade duration; scan path length; revisits increase with the increased sign comprehensibility. Another bivariate correlation analysis considering both parametric and non-parametric tests has been conducted to check the correlation between the subjective and the objective data to get a clear idea of their relation, and parametric tests are found to give good results in case of data being slightly skewed.

Forty-six (46) significant physiological factors have been found in the bivariate correlation study, out of which five (5) are positively correlated with the comprehensibility score (level), and eleven (11) variables are negatively correlated. The correlation study reveals significant variables/ measures from eye-tracking experiments, and this technique is successful in finding a relation between subjective and objective data.

Two categories/ groups of comprehensibility level (comprehensible/ incomprehensible) have been formed from the subjects-based score. ANOVA (Kruskal-Wallis) test has been performed, which calculates significant differences for the significant objective measures between the two groups of signs based on ranking the data. A Post hoc test performs multiple comparisons between all combinations of groups. It is found that the eye-tracking variables such as Vis_intk_durAvg, Blink_durAvg, AOI004Firstvis_intk_dur, AOI004Avg_vis_intk_dur, AOI005Firstvis_intk_dur, and AOI005Avg_vis_intk_dur are not same across categories of comprehensibility levels (comprehensible and incomprehensible signs), supported by the p-value which is less than 0.05 and the significant effect size. These outcomes exhibit the potential to integrate physiological measures from eye movements with existing evaluation methods in assessing the comprehensibility of OSH signs.

The logo of the Indian Institute of Technology Guwahati is a circular emblem. It features a central stylized figure with three rounded, bulbous shapes protruding from its body, resembling a traditional Indian deity or a symbolic figure. The figure is rendered in a light gray color. Surrounding the figure is a circular border containing text in both Hindi and English. The Hindi text at the top reads "भारतीय प्रौद्योगिकी संस्थान गुवाहाटी" and the English text at the bottom reads "Indian Institute of Technology Guwahati".

5

Comparative study between the techniques- its merits and demerits

Chapter 5: Comparative study between the techniques- its merits and demerits

Abstract

This chapter compares the four techniques studied in the previous chapters, i.e., subject-based, expert-based, computational, and eye-tracking techniques. Their comparison is discussed in terms of similarity, dissimilarity, advantages, and disadvantages. Such a comparison might assist the researchers or experts in selecting one or more sign comprehensibility evaluation techniques that will better fit their study or organizations. Finally, there is a discussion and proof of validation of all four techniques discussed in the study. The algorithm used in the computational model is further validated by a small comparative study with other related computational models, and the chapter ends with a conclusion.

5.1. Introduction

Various studies have been found regarding the comprehensibility of safety signs, traffic signs, and other visualizations (refer to Chapter 1, sub-sections 1.2, 1.3, 1.4, and 1.5). Those studies lack the inclusion of critical discussions regarding various techniques, and their advantages, drawbacks, and pitfalls simultaneously. Only one paper by Arcia and groups (Arcia et al., 2019) discussed variations of the ISO technique deployed in a health-related study, successes and lessons learned, and recommendations for future studies. The present thesis discussed the implementation of four techniques, among which two are novel, like computational technique and eye-tracking techniques in the OSH context (comprehensibility evaluation of OSH signs), and these techniques were never compared before with the other two techniques, like, subject-based and expert-based techniques. This chapter details the similarities, dissimilarities, advantages, and disadvantages of the techniques discussed throughout the thesis. Discussing the advantages and pitfalls will allow industry experts or researchers to choose a technique or more than one technique depending on their context, environment, and existing technology. The validation of the techniques helps to validate the output. The bivariate correlation method is one way of validating the techniques (Belouafa et al., 2017; Deepthy et al., 2022) among many methods used, and this method has been used here to validate the techniques.

5.2. Methodology

Analysis with a correlation (Pearson) plot followed by bivariate analysis has been done for the first three techniques (subject-based, expert-based, and computational) based on the comprehensibility scores extracted from them. For bivariate analysis of the three techniques, three correlation analysis results for Pearson's r , Kendall's tau, and Spearman's rho have been considered. The data distribution deviated from a true normal distribution (refer to 2.3.1.2, 2.3.2.3, 3.3.3); however, parametric tests like Pearson have also been considered here besides the nonparametric tests because Pearson's correlation coefficient have recently proved to offer a substantial improvement in statistical power even for distributions with

moderate skewness or excess kurtosis (Chok, 2010). Descriptive statistics have been calculated for all the signs, followed by bivariate analysis for all three techniques. The two statistical analyses by Pearson plot and bivariate correlation strengthen the proof of validation and validate the techniques/ models used in the study.

Next, a discussion on the correlation study of the subject-based and eye-tracking techniques has been provided to see the cognitive effect of the signs on the human eye. This study helps to identify the significant features of eye movement and its change with the change in comprehensibility levels. This study validates the eye-tracking experiment by validating the results, which align with the other Psychophysiological studies involving eye-tracking devices (refer to section 1.3.3). The eye-tracking device used here needs no output validation as the device has already been tested by its manufacturer, and various research has already been done using it (Blignaut et al., 2014; Salvucci & Goldberg, 2000). A small separate study has been conducted for the computational technique DAN-based model with related other models (pre-trained), and results have been tabulated.

5.3. Results and Discussions

This section discusses the results from different statistical tests used to compare the techniques implemented for research.

5.3.1. Correlation (Pearson) plot of the techniques

Initially, a total of forty-two (42) correlation plots (scatter plots) were plotted for all fourteen (14) signs with the three techniques (subject-based, expert-based, and computational). It is found that they have a fair, positive strength of correlation between each pair of techniques (subject-based and expert-based; subject-based and computational; expert-based and computational). The result validates the three techniques used for the research work. Only plots for one sign (COSACOEOT) have been shown here. The rest of the plots for the rest of the signs could not be included in this thesis for brevity. Considering COSACOEOT, the figure reveals that there is a fair correlation between the subject-based technique and expert-based technique (Figure 5.1) and the subject-based technique and computational technique (Figure 5.2), evident from the squared correlation coefficient (R^2) value. However, their correlation coefficient is much less as compared to the expert-based technique versus the computational technique (Figure 5.3). The reason is that the subject-based technique extracted the score from the participants/ subjects who are novices and are not experts in rating/ scoring. Though training had been imparted to them before the experiment, they still could not perform well in the scoring method. They are found to score more when they wrote incorrectly or comprehended the sign incorrectly, as evidenced by their written response and the score they provided.

Similarly, few participants are seen to provide minimum scores when they wrote correct/ exact answers. This result is evident from the score of the novices, which is why the expert-based technique has been considered. The experts are experienced in this OSH domain and scoring procedure. They are well-trained, and the experts agree upon a protocol for the

scoring procedure before the actual trial. They are less ambiguous in marking/ rating the responses. Their scoring is believed to be more consistent compared to the subject-based technique. This inconsistency in the rating is the reason behind the lower correlation coefficient for the techniques involving subject score; however, the two techniques overall scored well. Their justified correlation strength validates the two techniques used in the study (Figure 5.1).

The expert-based technique versus the computational technique is found to be highly correlated, as evident from (Figure 5.3). The reason behind such a high correlation coefficient can be stated that the computational model/ technique has been trained with real-time expert data from experts (humans) using the STS benchmark and SNLI corpus. The transfer learning model helps to transfer the knowledge to the computational model, which easily compares the written text (response of the participants/ subjects) and the standard interpretation of the signs (provided by the experts in organizations/ ISO-ANSI standards) and provides value which defines the semantic similarity between the two texts. This is precisely the way an expert rates a response. The computational model/ technique simulates the judgment like experts and scores the responses. However, there is every possibility that human experts can commit unintentional errors while scoring; and provide biased judgment due to a preoccupied mindset, repetitive evaluation of the innumerable answer scripts (survey), environment, mindset, notions, persuasion, etc., which is uncontrollable. However, in the case of computational techniques, the computer can provide consistent scores irrespective of the situation. It will not get biased due to external conditions/ environment.

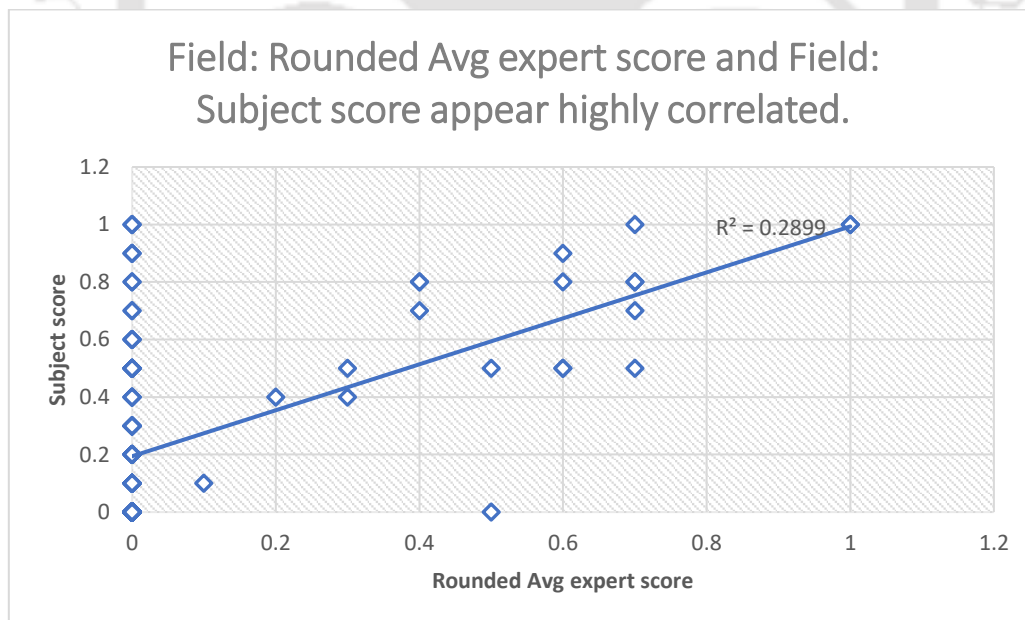


Figure 5. 1. Correlation plot of subject-based technique versus expert-based technique for COSACOEOT

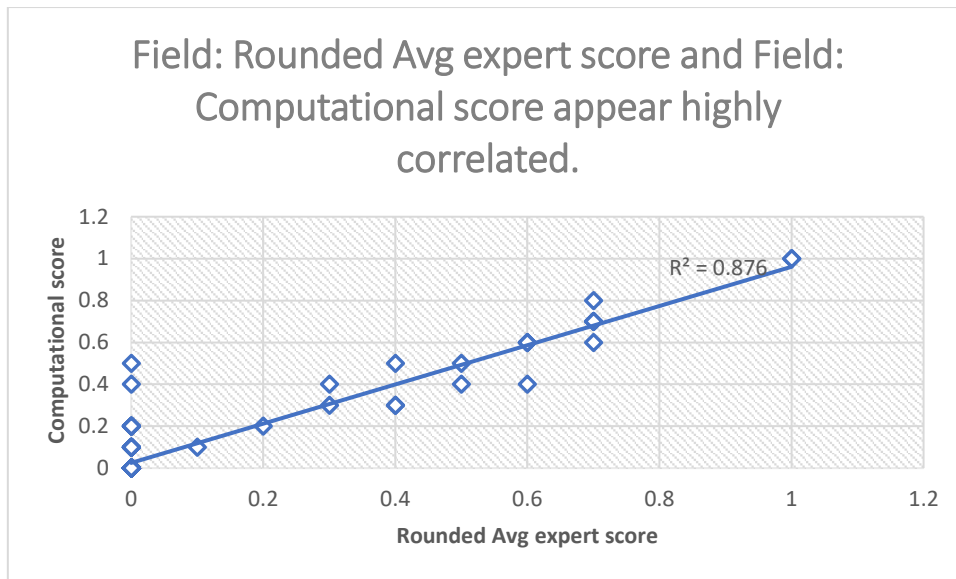


Figure 5. 2. Correlation plot of expert-based technique versus computational technique for COSACOEOT

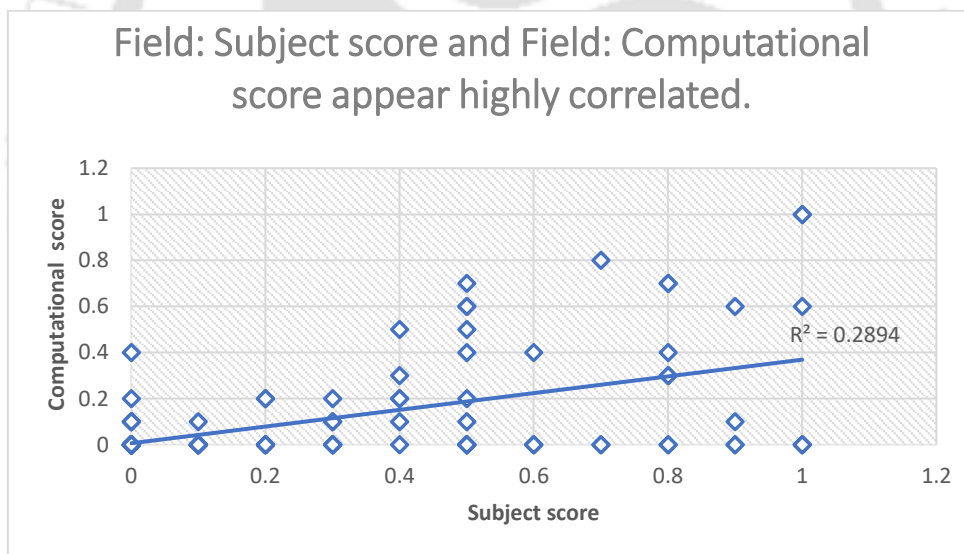


Figure 5. 3. Correlation plot of subject-based technique versus computational technique for COSACOEOT

The result discussed here is only for COSACOEOT; however, the same discussion applies to all the signs under study. The discussion for all the signs has not been included in the thesis for brevity. All the signs have shown a similar correlation trend between the techniques discussed for COSACOEOT. The expert-based technique versus computational technique shows a high correlation compared to the subject-based technique versus computational technique; and subject-based technique versus expert-based technique, respectively.

The above three techniques discussed have a disadvantage in that the techniques are dependent on subjective data. The participants may be unable to write correctly (as it was

subconscious or forgotten) what they say and think. They have to rely on memory for the experiment. To get unobtrusive, real-time data, eye-tracking analysis can lead to discoveries that would be considerably more difficult to uncover with other evaluation techniques. Though eye-tracking cannot replace the other techniques studied here, such a technique can be implemented when there is a need for objective data. Eye tracking is a costly method, so prior research about its feasibility must be conducted before directly delving into this technique. Besides, any comprehensibility scores could not be generated using this technique. First, the comprehensibility score is extracted via a questionnaire (verbal is also possible), and then an eye-tracking experiment is conducted. Later the comprehensibility score is studied concerning the eye tracking features (46 significant eye features in this study). So for the eye-tracking technique implemented here, both the subjective data and the objective data have been used for drawing a correlation (initial scatter plot followed by bivariate correlation) between the two variables (comprehensibility score and the eye-tracking features/ measures) (refer to Figure 4. 11- Figure 4. 22). Hence, the eye-tracking study involved here exploits both the features of the subjective as well as the objective data extraction technique to see how objective measures like fixations, saccades and blinks correlate with the subjective score.

5.3.2. Bivariate correlation of the techniques

i. Bivariate correlation for three techniques- subject-based, expert-based, and computational technique

Initially, descriptive statistics have been computed for the three techniques- subject-based, expert-based, and computational technique (computational being synonymously used as Machine-based - short form is mac) (refer to Table 5.1). The descriptive statistics show the measures of central tendency (mean value considered in this study) and measures of variability (standard deviation considered) in this study. The mean and the standard deviation are found to differ very little when the expert-based technique is compared with the computational technique. The reason behind the same is that the computational technique has been designed based on the expert-based model and keeping in mind the expert's views. Similarly, there is a gap in the mean and the standard deviation when the expert-based technique/ computational technique is compared with the subject-based technique. The reason behind the same is that the subject-based technique is based on the scores/ ratings provided by employees who are novices in the rating procedure. Though training before the actual experiment has been conducted on how to rate the sign considering the subject-based rating technique (considering signs that are not used in the study), however, with one day of training, expertise in rating cannot be expected from the users.

To further confirm the result, i.e., the differences between the techniques (subject-based, expert-based, and computational-based), the Bivariate correlation tests have been conducted with comprehensibility scores of all the signs generated from three techniques (subject-based, expert-based, and computational-based). Both parametric (Pearson's r) and non-parametric (Kendall's τ_b and Spearman's ρ) tests have been performed. It is seen that the parametric test gave a better result than the non-parametric test (Table 5. 2, and Table 5. 3). The result obtained is significant considering the p-value to be less than 0.01 and 0.05. The higher the

correlation coefficient (concerning and comparing the expert-based technique), the better the model. Pearson's bivariate correlation coefficient shows a high positive linear relationship between two techniques (expert-based and computational, *exp_mac*: $r = .936$ in the case of COSACOEOT) that is significantly different from zero ($p < 0.001$). Similar results have been found for other signs too. The Pearson correlation for the expert-based and subject-based are found to have a low positive linear relationship (correlation) compared to expert-based and computational). Similarly, the Pearson correlation for the subject-based and machine-based are found to have a low positive correlation compared to expert-based and computational. Similar relationships have been found when all three comprehensibility evaluation techniques have been compared using Kendall's Tau_b and Spearman's Rho bivariate correlation techniques respectively. The bivariate results of all 14 signs for the three techniques using parametric and non-parametric methods have been tabulated in (Table 5. 2 and Table 5. 3). All the results tabulated were found to be significant, with the exception in the case of NAFUP. The p-value in the case of 'sub_mac' (Pearson r; Kendall's tau_b; and Spearman's rho) are found to be greater than 0.01 and 0.05, and the correlation coefficient is not significant. However, this result does not affect the outcome as it is already seen that the subject-based technique involves subjects who are novices in the scoring procedure (for details, refer to 5.3.1). The same result is obvious for other combinations 'sub_mac' involving subject-based technique.

ii. *Bivariate correlation for eye-tracking experiment (subjective data versus objective measures)*

A bivariate correlation test has been conducted for the eye-tracking technique also. The objective measures have been tested for their correlation with the subjective comprehensive score (within-group experiment) using the same parametric (Pearson r) and non-parametric (Kendall's tau_b and Spearman's rho) tests. The outcome of the significant parameters of event statistics is as follows: *Sac_velMin* (DNWSH WAFS, DNT), *Blink_durMax* (WAFS, DNT_T), *Sac_velAvg* (DNT, WAFS_T), *Blink_cnt* (NAFUP_T, DNT_T), *Blink_freq* (NAFUP_T, DNT_T), *Blink_dur* (NAFUP_T, DNT_T), *Scanpath_len* (COSACOEOT, WAFS_T), *Sac_amplMin* (NAFUP_T, COSACOEOT_T, UBC_T) have a negative correlation with comprehensibility score. The parameters *Vis_intk_durMin* (NAFUP, WRP_T), *Vis_intk_durtot* (WAFS, WRP), and *Vis_intk_durMax* (COSACOEOT, WRP) have a positive correlation with comprehensibility score. *Vis_intk_durAvg* (NAFUP, NAFUP_T) positively correlates with the comprehensibility score. The outcome of the significant parameters of AOI statistics can be depicted as *AOI_size004_Vis_intkTime* (WAFS, WRP, WRP_T), *AOI_size005_Vis_intkCnt* (WAFS, WRP, COSACOEOT), *AOI_size004First_vis_intkdur* (WRP, DNWSH_T, NAFUP_T) negatively correlates with comprehensibility score. Similarly, *AOI_size004_Revisits* (COSACOEOT, WAFS_T) correlate positively with the comprehensibility score (for details, please refer to 4.3.2.2 and Appendix G). The results align with the previous literature, thus validating the technique.

The SMI eye-tracking device used for the eye-tracking experiment (fourth technique) needs no separate validation as the Senso motoric instruments claim to be drift-free and high accuracy has been obtained by calibrating the device each time the participants were recorded (Blignaut et al., 2014). Data loss may occur during blink (Niehorster et al., 2020); however,

Sign	Mean	Std. Deviation	Sign	Mean	Std. Deviation
COSACOEOT_mac	.1000	.21386	NAFUP_T_mac	0.4591	0.32844
COSACOEOT_exp	.0803	.21380	NAFUP_T_exp	0.5000	0.37632
COSACOEOT_sub	.2584	.31731	NAFUP_T_sub	0.6255	0.34471
COSACOEOT_T_mac	0.4409	0.36653	UBC_mac	0.2759	0.15460
COSACOEOT_T_exp	0.4693	0.39660	UBC_exp	0.3124	0.16603
COSACOEOT_T_sub	0.4401	0.36491	UBC_sub	0.5956	0.32985
DNT_mac	0.5942	0.30432	UBC_T_mac	0.5387	0.33721
DNT_exp	0.6263	0.41926	UBC_T_exp	0.5949	0.34520
DNT_sub	0.5737	0.34220	UBC_T_sub	0.6066	0.34390
DNT_T_mac	0.7876	0.28139	WAFS_mac	0.6774	0.24404
DNT_T_exp	0.8438	0.32807	WAFS_exp	0.7212	0.28657
DNT_T_sub	0.7007	0.32897	WAFS_sub	0.7445	0.27757
DNWSH_mac	0.4146	0.19909	WAFS_T_mac	0.8387	0.19298
DNWSH_exp	0.5007	0.27507	WAFS_T_exp	0.8978	0.20017
DNWSH_sub	0.6766	0.32883	WAFS_T_sub	0.8701	0.20771
DNWSH_T_mac	0.6234	0.26988	WRP_mac	0.3080	0.23641
DNWSH_T_exp	0.7015	0.31436	WRP_exp	0.2839	0.27766
DNWSH_T_sub	0.7723	0.31006	WRP_sub	0.4000	0.35314
NAFUP_mac	0.1708	0.15912	WRP_T_mac	0.5679	0.33889
NAFUP_exp	0.1730	0.19982	WRP_T_exp	0.5591	0.37974
NAFUP_sub	0.6496	0.31181	WRP_T_sub	0.5818	0.36162

Here 'mac' signifies machine/ computational; 'exp' signifies expert-based; and 'sub' signifies a subject-based technique

Table 5. 1. Descriptive statistics of all the signs with three different techniques (subject-based, expert-based, and computational technique)

the experiment has been done cautiously such that no such loss has been incurred, evident from the eye data. A minor movement during speaking may result in a significant deviation in the gaze position signal (Niehorster et al., 2020); hence subjective data was collected in a separate experiment to study its relation with the eye tracking data. The participant was not disturbed by being forced to write or speak while conducting the eye-tracking experiment.



Table 5. 2. Bivariate correlation of the seven signs with three different techniques (subject-based, expert-based, and computational technique)

		Pearson Correlation			Kendall's tau_b			Spearman's rho		
		exp_sub	exp_mac	sub_mac	exp_sub	exp_mac	sub_mac	exp_sub	exp_mac	sub_mac
COSACOEOT	Correlation	.538**	.936**	.546**	.414**	.742**	.413**	.465**	.768**	.469**
	Sig. (2-tailed)	.000	.000	.000	0.000	0.000	0.000	0.000	0.000	0.000
COSACOEOT_T	Correlation	.650**	.924**	.684**	.511**	.868**	.510**	.638**	.937**	.651**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DNT	Correlation	.592**	.844**	.464**	.460**	.724**	.339**	.558**	.848**	.433**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DNT_T	Correlation	.606**	.842**	.541**	.456**	.667**	.332**	.530**	.732**	.407**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DNWSH	Correlation	.538**	.840**	.471**	.327**	.726**	.293**	.403**	.840**	.354**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DNWSH_T	Correlation	.636**	.906**	.553**	.399**	.788**	.350**	.467**	.893**	.429**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NAFUP	Correlation	.222**	.835**	0.157	.132*	.498**	0.061	0.160	.555**	0.072
	Sig. (2-tailed)	0.009	0.000	0.067	0.050	0.000	0.374	0.062	0.000	0.402

Note: **. Correlation is significant at the 0.01 level (2-tailed).

Table 5. 3. Bivariate correlation of the remaining seven signs with three different techniques (subject-based, expert-based, and computational technique)

		Pearson Correlation			Kendall's tau_b			Spearman's rho		
		exp_sub	exp_mac	sub_mac	exp_sub	exp_mac	sub_mac	exp_sub	exp_mac	sub_mac
NAFUP_T	Correlation	.617**	.935**	.567**	.478**	.847**	.421**	.598**	.923**	.547**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UBC	Correlation	.423**	.885**	.356**	.318**	.704**	.256**	.379**	.754**	.314**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UBC_T	Correlation	.682**	.959**	.680**	.530**	.919**	.531**	.639**	.970**	.641**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WAFS	Correlation	.521**	.921**	.543**	.249**	.798**	.293**	.307**	.873**	.358**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WAFS_T	Correlation	.582**	.906**	.545**	.173*	.713**	.141*	.196*	.767**	.169*
	Sig. (2-tailed)	0.000	0.000	0.000	0.022	0.000	0.049	0.022	0.000	0.049
WRP	Correlation	.736**	.953**	.730**	.651**	.878**	.628**	.765**	.946**	.763**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WRP_T	Correlation	.752**	.979**	.718**	.566**	.939**	.543**	.690**	.977**	.672**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

** . Correlation is significant at the 0.01 level (2-tailed). * . Correlation is significant at the 0.05 level (2-tailed).

5.3.3. Comparison of techniques considering different parameters

A comparison study has been conducted based on a few selected parameters, and the results are tabulated in Table 5.4. The table is self-explanatory, and no separate discussion is required. The parameters are based on all the techniques implemented in this Ph.D. work from Chapter 2 to Chapter 4. Finally, the comprehensible and incomprehensible signs have been identified. Comprehensible signs have been considered for use, and incomprehensible signs have been recommended for redesign and reevaluation.

To discuss the first three techniques, subject-based, expert-based, and computational-based techniques, subjective data has been considered.

The first method considers the 11-point rating score, second and the third methods consider the descriptive text. The first method collects data from subjects who are novices in scoring and hence the data may be biased, reflected by the correlation and bivariate correlation study and the comprehensibility result in Chapter 2. The subject-based technique sometimes does not need experts and the organizations can just directly distribute the standard questionnaire and get the users' response. Hence, this can be sometimes cost-effective for a small organization, however, the responses may not be so reliable and validated when experts are not involved. Hence, for proper evaluation, experts must be involved in this technique. Hence, the technique thought to be cost-effective becomes more costlier than the other two techniques because training needs to be provided to the subjects for rating, and the experts also need to be involved.

The expert-based technique is less biased than the subject-based technique as the experts are well-trained in evaluating the signs. The subjects need not be trained rigorously as the experts rated/ scored the descriptive responses collected from users. Hence, data may be reliable.

The existing manual procedure (subject or expert-based technique) may lead to inconsistency in evaluating the responses and agreement on the category of responses due to individual biases and self-doubt of the experts. Peer review might provide acumen in this context, but often the ideas seem to diverge instead of converging, thus worsening the situation. The inconsistency might be due to their previous experience, age, context, or persuasion. Whatever the reason, this is unintentional and leads to the wrong utilization of valuable human resources. Organizations can better utilize their experts in designing and developing the signs rather than engaging their expert brains for the same repeated and tedious evaluation task. An approach has been attempted to cut off one among the numerous repeated jobs from the schedule, i.e., scoring survey responses, and prevent the OSH sign designers/ experts from spending time marking (scoring) the response sheet collected from the surveys. The comprehensibility evaluations for numerous signs can be conducted parallelly, therefore increasing the number of new signs and meeting the demand of organization(s) within a short span, which is something unimaginable previously. There is a requirement for consistent and accurate predictions of comprehensibility (unambiguous results), which is lacking in an earlier body of knowledge due to the biases of the human-based subjective evaluation. Hence, the computational model has been created with the collaboration of human experts which

makes it flexible to adapt to work in various contexts and environments. This technique can generate accurate and unambiguous scores without human intervention within a few seconds. The computational model decides the scores and generates the comprehensibility levels. This technique is cost and time-effective which can generate innumerable results within fractions of seconds for numerous signs simultaneously. The comprehensible and incomprehensible signs identified have been recorded in Table 5.4.

The eye-tracking technique is costly, delicate, and complex in comparison to the other three techniques. The experiment design needed a within-group study. A separate subject-based study and a separate eye-tracking study have been conducted. The details of the experiment designs for both techniques have been elaborated in Chapter 4. Designing an eye-tracking experiment needs patience and knowledge about the experimental setup and how to use it. Initially, calibration of the eye-tracking device plays a very important part in the experiment. The machine has been calibrated each time, the experiment has been conducted. The participants must be initially checked for their eye defects or color blindness before conducting any experiment. Convincing users that the experiment is safe is very difficult. The experimenter must be knowledgeable in using the eye-tracking device and analyzing the output. The comprehensible and incomprehensible signs identified from the subject-based technique have been recorded in Table 5.4. It is to be noted that the eye-tracking objective data collected does not measure the comprehensibility level directly, but the data has been used to study the correlation between the comprehensibility levels calculated from the subject-based technique and the objective data. The correlation study assisted in understanding the variability of the objective data across comprehensibility levels (Comprehensibility levels are of two types - comprehensible and incomprehensible groups).

Table 5. 4. Comparison of the techniques using various parameters

S.No	Parameter	Subject-based	Expert-based	Computational	Eye-tracking
1	Data type	Subjective	Subjective	Subjective	Subjective and Objective
2	Data Source	Participant's score (0.0 - 1.0)	Participants textual response	Participants textual response	Participant's score and eye recording
3	No. of Participants	137	137	137	41
4	Participant type	Working professionals (Novices in terms of OSH)	Working professionals (Novices in terms of OSH)	Working professionals (Novices in terms of OSH)	Novice students- IIT Guwahati (Semantic transparency)
5	Instrument	Questionnaire	Questionnaire	Questionnaire	Eye tracking device: eye-tracking glass (ETG) and questionnaire
6	Experimental set-up	Computer	Computer	Computer	Computer and eye tracking device
7	Softwares/ Analysis	Nil/ IBM SPSS Statistics (Version 25.0).	Nil/ IBM SPSS Statistics (Version 25.0).	Python/ IBM SPSS Statistics (Version 25.0).	Nil/ BeGazeTM, IBM SPSS Statistics (Version 25.0)
8	Ease of handling experiment	Difficult (subjects need adequate training and trained experts must be available)	Easy (subjects do not need much training and trained experts are a must)	Easy (subjects do not need much training, and experts need not be involved in decision-making)	Very difficult (Eye device needs a good amount of calibration, and subjects need much training, experts must have good knowledge about the device and the analysis of eye data)
9	Decision maker: Sign comprehensible or incomprehensible	Experts	Experts	Computer	Experts
10	Comprehensible signs	UBC_T, DNT, NAFUP_T, NAFUP, DNWSH,	DNWSH_T, DNT_T,	DNWSH_T, DNT_T,	NAFUP, UBC, WAFS, DNWSH, COSACOEOT_T, WRP_T, UBC_T,

		DNT_T, DNWSH_T, WAFS, WAFS_T	WAFS, WAFS_T	WAFS, WAFS_T	DNT_T, NAFUP_T, WAFS_T, DNWSH_T
11	Incomprehensible signs	COSACOEOT, WRP, COSACOEOT_T, UBC, WRP_T	NAFUP, COSACOEOT, UBC, WRP, NAFUP_T, DNWSH, COSACOEOT_T, DNT, UBC_T, WRP_T	NAFUP, COSACOEOT, UBC, WRP, DNWSH, NAFUP_T, COSACOEOT_T, UBC_T, DNT, WRP_T	COSACOEOT, DNT, WRP
12	Most comprehensible sign (in %)	WAFS_T (94.89%)	WAFS_T (96.35%)	WAFS_T (94.16%)	DNWSH_T (100%) from subjective data
13	Least comprehensible sign (in %)	COSACOEOT (28.46%)	NAFUP (6.56%)	NAFUP (5.83%)	COSACOEOT (43.90%) from subjective data

5.3.3.1. Comparison of DAN-based computational model with related computational models

A comparison study has been conducted between the DAN-based model and the other related pre-trained general-purpose sentence transformer models.

The models are- all-MiniLM-L6-v2, all-mpnet-base-v2, multi-qa-distilbert-cos-v1, msmarco-distilbert-base-v2, and average_word_embeddings_glove.6B.300d (as shown in Table 5.5). The correlation coefficient score for each training algorithm/ model is displayed in the green rows. Consider DAN based model: The standard message for the sign is “Do not Walk or stand here” and the five participants’ responses with their correlation scores (comprehensibility score) compared to the standard message are: “do not stand, 0.58,” “don't step on the floor, 0.45”, “Do not step, 0.51”, “Shoes not allowed, 0.22”, and “Shoes or man not allowed, 0.39”. The correlation score for the other models can be seen in the same Table (Table 5. 5).

For each model, the computation time is shown in the table with its CPU/GPU specification for executing the algorithm (second column, white background). The training dataset and the dimension of the mapped vector space have been displayed in the third column with a white background. The models have been taken from <https://huggingface.co/sentence-transformers> and tested with the OSH data.

It is found that the DAN-based model performed as per the related models in terms of generating the scores in relation to the human experts. This further validates the computational model implemented in this research. Comparison with other related models (Table 5.5) also helped to prove that the pre-trained model does not suffer from the 'overfitting' of the data during training.

Table 5. 5. Comparative study between the proposed MMH model- based on DAN and the other similar kinds of models

Standard OSH Message: Do not Walk or stand here					
Training model: DAN-based model (applied model)		Computation time on Python 3 Google Compute Engine backend (GPU): 1.139 (secs)		Training Dataset: STS dataset. Dimension: 512-dimensional dense vector space	
Participant’s responses:	do not stand	don't step on the floor	Do not step	Shoes not allowed	Shoes or man not allowed
Correlation scores:	0.58	0.45	0.51	0.22	0.39
Training model: Sentence-transformers/all-MiniLM-L6-v2		Computation time on Intel Xeon 3rd Gen Scalable CPU: 0.020 (secs)		Training Dataset: SNLI, ELI5, TRIVIA_QA Dimension: 384-dimensional dense vector space	

Correlation scores:	0.71	0.58	0.54	0.28	0.33
Training model: Sentence-transformers/all-mpnet-base-v2		Computation time on Intel Xeon 3rd Gen Scalable CPU: 0.062 (secs)		Training Dataset: SNLI, ELI5, TRIVIA_QA	Dimension: 768-dimensional dense vector space
Correlation scores:	0.77	0.57	0.69	0.45	0.41
Training model: sentence-transformers/multi-qa-distilbert-cos-v1		Computation time on Intel Xeon 3rd Gen Scalable CPU: 0.039 (secs)		Training Dataset: Trivia_qa, eli5, yahoo_answers_topics	Dimension: 768-dimensional dense vector space
Correlation scores:	0.74	0.64	0.60	0.24	0.31
Training model: sentence-transformers/msmarco-distilbert-base-v2		Computation time on Intel Xeon 3rd Gen Scalable CPU: 0.039 (secs)		Training Dataset: STS dataset, NLI	Dimension: 768-dimensional dense vector space
Correlation scores:	0.78	0.56	0.64	0.44	0.45
Training model: sentence-transformers/average_word_embeddings_glove.6B.300d		Computation time on Intel Xeon 3rd Gen Scalable CPU: 0.003 (secs)		Training Dataset: Wikipedia 2014	Dimension: 300-dimensional dense vector space
Correlation scores:	0.81	0.53	0.46	0.51	0.59

5.3.4. Validation of techniques

The correlation study for all the techniques helped validate the techniques and, thus, validate the outputs. The correlation (Pearson) plot (refer to 5.3.1) and the bivariate correlation study (both parametric and non-parametric- refer to Appendix G, 4.3.2.1, 4.3.2.2, and 5.3.2 (Table 5. 2)) helped to validate all three techniques (subject-based, expert-based and computational techniques) used in the study. The strength of the correlation between all the techniques is significant, with p-values less than 0.001 and 0.05, respectively. Besides these proofs, the computational model is further validated with the other related models (Table 5. 5). The eye-tracking study is validated by the outcome/ results (refer to Figure 4. 11- Figure 4. 22), which are in line with the other Psychophysiological studies conducted involving eye-tracking devices (Y.-L. Hsu et al., 2013; Siswandari & Xiong, 2015) (refer to section 1.3.3). Further

discussion is provided in 5.3.1 and 5.3.2. Besides, the eye-tracking device has already been tested by its manufacturer, and various research has already been done using this device (Blignaut et al., 2014; Salvucci & Goldberg, 2000).

5.4. Conclusion

This chapter did a comparative study of all four techniques implemented for the thesis work. For the first three techniques, the scatter/ Pearson correlation plot reveals a linear correlation between the scores generated by all the techniques. The bivariate analysis further assists in evaluating the strength of the techniques implemented and validating the techniques. Both parametric (Pearson's r) and non-parametric (Kendall's τ_b and Spearman's ρ) tests have been performed, and it is seen that the parametric test performed better than the non-parametric test. The result is significant, considering $p < 0.001$ and $p < 0.05$. The correlation coefficient between all the techniques (R-value for subject-based technique versus expert-based technique, expert-based technique versus computational technique, subject-based technique versus computational technique) shows a significant correlation with each other. This analysis compares all the techniques and validates all three techniques used in the study. A separate comparative study has been done for the DAN model with other related models, and DAN generated results in line with the other computational models. The eye-tracking result generated aligns with the previous Psychophysiological studies; hence, eye-tracking results are validated. The common comprehensible signs found after evaluating with all the evaluation techniques are DNWSH_T, DNT_T, WAFS, and WAFS_T. The common incomprehensible signs found are COSACOEOT and WRP. These are the common results (output: comprehensible/ incomprehensible signs) found for all the techniques, thus validating the techniques and the outputs from the study.



Proof of concept

Chapter 6: Proof of concept

Abstract

This is a complementary chapter that discusses the redesign of one of the incomprehensible signs COSACOEOT generated from the research (Chapter 2 to Chapter 4). COSACOEOT is the outcome of the comprehensibility evaluation study, and it has been redesigned into two new variants. The COSACOEOT undergoes various phases of the sign design/ redesign life cycle proposed here. A ranking-based comprehensibility study, followed by a discussion with experts/ designers was conducted to generate the most comprehensible variant(s), a complement to the previous four techniques. Two final variants (signs): COSACOE and COSIT were finally designed for use.

6.1. Introduction

The thesis is related to evaluation techniques of comprehensibility (understandability) of the OSH signs. After implementing the evaluation techniques, the less comprehensible signs have been identified, and the poorly designed sign has been redesigned as a proof of concept and tested further before it can be made available for use in industries. It is to be noted that both the outcome of the study, COSACOEOT, and the two redesigned variants are the proof of concept. The re-designed variants can also be evaluated by the previous four discussed techniques which is out of the scope of the thesis. Additional evaluation by ranking method, a complement to the studied techniques is considered here where more than one variant needs to be selected. This chapter discusses the design life cycle of two new variants from the existing design COSACOEOT.

6.2. Sign design/ redesign life cycle

The comprehensibility evaluation using four techniques found COSACOEOT as one of the most incomprehensible and ambiguous OSH signs. Hence, it is recommended to redesign the sign (Gaver et al., 2003). The intention is to redesign the sign so that it is comprehensible to novices; hence novices have been involved during the conceptualization phase (semantic transparency). The theory behind applying semantic transparency is that if novices successfully comprehend the message the sign conveys, then the industry/ organization personnel can easily comprehend it. The design/ redesign phase incepts with ideation, simultaneously followed by conceptualization; both go hand-in-hand. For ideation, '*Brain dumping*' has been adopted. 'Brain dumping' is like brainstorming; however, it is done individually. After brain dumping, '*Brainwalking*' has been followed. Brainwalking is like brain-dumping; members walk about the room, adding to others' ideas.

6.2.1. Initial conceptualization

“Design conceptualization is the procedure for generating ideas for an optimum solution to the design problem” (Elmogahzy, 2019; Martin et al., 2012). The initial conceptualization of a sign starts with understanding the context of use, the intended user, the severity of the context, the intended response from the user, the content for the pictogram, the text to be embedded, and many hidden factors. The ideation for the optimal solution to the design problem generates concepts that may exist, be new, or be a combination of both. For a sign to be comprehensible, it is always better to prioritize a concept related to a previous sign or concept that everybody knows. A drastic or radical change in a variant is not acceptable, and it is always better to make incremental changes (Mugge & Dahl, 2013), reducing learning costs and effort (cognitive effort). Thus, an incremental change has been adopted here for the sign redesign.

Novice users are involved throughout the conceptualization phase, different from the participants involved in the final variants’ evaluation. A few concept designs are displayed in Figure 6. 1.



Figure 6. 1. Conceptualization for a redesign of COSACOEOT (sign)

6.2.2. Design of variants

After the conceptualization, five (5) incremental variants have been designed.

Solo effort (Pedgley, 2007) followed by *brainstorming* has been mainly used while designing the variants besides one-to-one discussions with few participants. The design of the variants has been divided into two phases, namely *Phase 1* and *Phase 2*.

Phase 1: The first four variants (Variant 1- Variant 4) are designed initially (Figure 6. 2). In the initial phase of the design, they have been discarded. *The solo effort* followed by the *brainstorming* with the participants deciphered that the sign “Cough or sneeze at the crease of the elbow or in a tissue (COSACOEOT)” is two concepts in one message (sign) and should not be communicated via a single sign. The two concepts made the sign ambiguous. The ambiguity is eliminated here by splitting the message concept into two sub concepts, “Cough and sneeze at the crease of the elbow” and “Cough and sneeze into a tissue.”

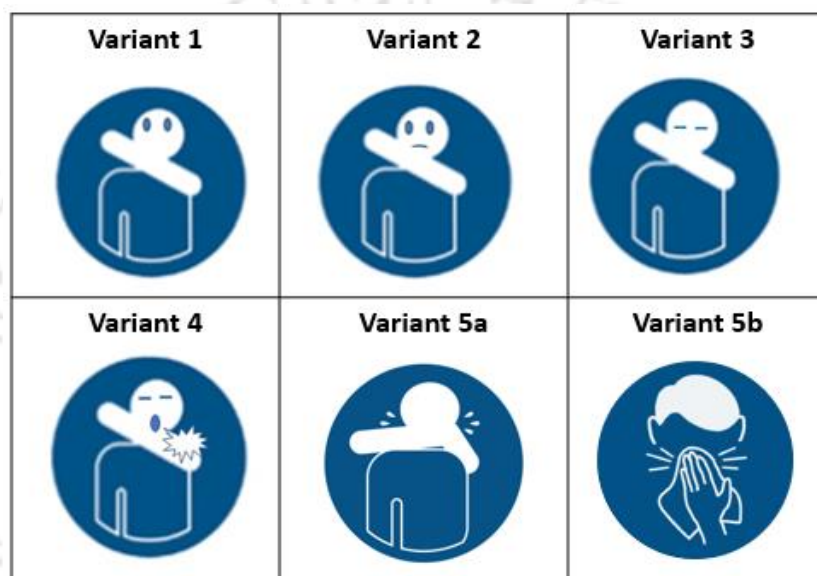


Figure 6. 2. Design of variants of COSACOEOT (Phase 1 and Phase 2)

Phase 2: Variant 5a is designed for “Cough or sneeze at the crease of the elbow- COSACOE,” and variant 5b is for “Cough or sneeze into a tissue- COSIT” (refer to variants 5a and 5b in Figure 6.2. Further research on these two variants reveals that variant 5b is not according to the sign standard (ISO). The variant 5b does not also conform with the 5a sign (variant). Again, variant 5b undergoes through redesign phase, and four incremental variants of 5b have been redesigned (5b.a., 5b.b., 5b.c., 5b.d.), respectively (Figure 6. 3. Redesigned variants of Sign 5b (Sign: Cough or sneeze in a tissue- COSIT)



Figure 6. 3. Redesigned variants of Sign 5b (Sign: Cough or sneeze in a tissue- COSIT)

6.2.3. Sign Design Evaluation Phase

An initial informal evaluation and group discussion were conducted for the initial designs. Once the design variants are finalized, they are ready for actual evaluation (re-iteration of the discussed one or triangulation of techniques or ranking as done here as the variants are four, before they can be declared as standardized and ready to be used.

6.2.3.1. Participants

The participants in the experiment are students and research scholars from the Indian Institute of Technology, Guwahati, India, who are novices in the Occupational Safety and Health (OSH) domain and have little or no knowledge about the OSH signs and symbols. Eighty-nine (N= 89) participants participated in the experiment, of which eighty (N=80) participants' data was taken. The remaining nine (N=9) data were corrupted and not considered for the study. The participants volunteered for the study with no history of color deficiency, severe eye conditions, or visual impairment (self-reported). They were selected based on purposive sampling with a mean age of 25.18 and a standard deviation of 7.41. They had no eye defect (tested by Snell's chart) or color blindness (checked using Ishihara Test for Colour Blindness). The participants read the instructions for the experiment before the commencement of the experiment. Participants gave consent before the beginning of the trial. The experiment has been conducted following ethical guidelines and principles.

6.2.3.2. Stimuli

Four variants (Figure 6. 3) have been used as stimuli for the experiment. The questionnaire was divided into two parts. The first part enquired about the demographic information of the participants. The second part asked the participants to rank the variants of the signs according to their perceived level of comprehensibility.

6.2.3.3. Research Design

Once the sign(s) have been designed, they can go for the evaluation phase with any or all or more than one of the four (4) techniques discussed in chapters 2, 3, and 4. The decision depends on the designers or OSH experts if they feel it is necessary to go for a strict evaluation once again. Otherwise, they can use other simple evaluation techniques like the ranking method implemented in this work. The present design has four variants, and knowing which variant to select is essential. Hence, a ranking method (Eygue et al., 2020; La Fata et al., 2021) has been used for the experiment. Rank 1 signifies the highest comprehensibility, followed by 2, 3, and 4 in the decreasing order of the comprehensibility level. The participants were asked to rank each variant and not any variant with the same rank. Records with similar ranks were discarded and not considered further in the study. The sign with the highest rank has been considered for use. Before the release, the sign can go through a final discussion, or suggestions are welcome for further improvement before the final release, which has been considered here, keeping in mind the commitment to deliver an appropriate

sign. The novices and the designers were involved in an informal discussion before the final release. The users and the designers were encouraged to comment on the final variant. A few iterations and slight modifications have been done with the final version before its release. A sign design/ redesign life cycle (Figure 6. 4) has been proposed here, keeping in mind the users (a user-centered design (UCD) approach (Vigoroso et al., 2020)).



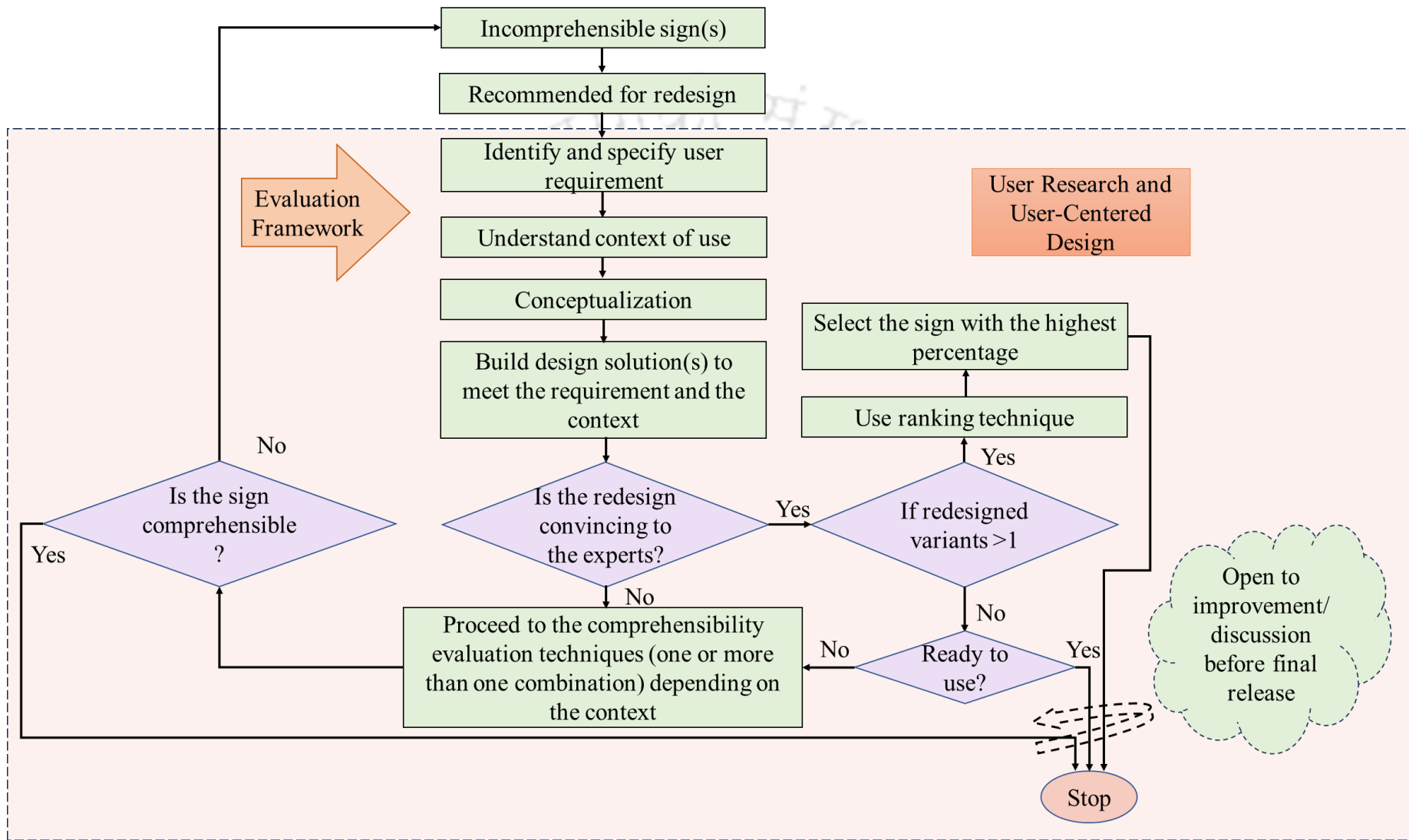


Figure 6. 4. UCD-based sign design/ redesign cycle

6.2.3.4. Result Analysis and Discussion

The result generated from the ranking method elicited that Variant '5b.d' scored maximum rank (rank 1). Thirty (30) participants gave rank 1 to variant '5b.d'. Similarly, 28 participants gave rank 2 to Variant '5b.c'. 26 gave rank 3 to variant '5b.b', and 26 gave rank 4 to '5b.a' and '5b.b'. Variant '5b.d' scored the maximum of all the four variants concerned in the study, variant 5b.d is maximumly comprehensible, and variants '5b.a' and '5b.b' scored a maximum for rank 4. The result of the study shows that Variants '5a' and '5b' got the lowest rank with the highest frequency. Considering rank 1 (one) as the highest comprehensible, the decreasing order of the frequency of participants who found variant '5b.d' as the most comprehensible can be portrayed as Variant 5b.d> Variant 5b.a> Variant 5b.c > Variant 5b.b. This variant 5b.d has been shown to design experts and novice users whether there is a chance of any improvement. There is a slight modification; the final variant is shown in Figure 6. 6.

Table 6. 1. Result of the ranking technique applied for sign 5 variants (Sign: Cough or Sneeze into a tissue)

Rank	Variant 5b. a	Variant 5b. b	Variant 5b.c	Variant 5b. d
1	20	12	16	30
2	15	26	28	14
3	19	26	21	15
4	26	26	15	21



Figure 6. 5. Variant 5b.d with maximum comprehensibility



Figure 6. 6. Final Variant 6 after slight modification in the variant 5b.d

6.3. Conclusion

This chapter discusses the redesign of one of the most incomprehensible signs studied here (COSACOEOT). The chapter discusses the design methods followed during the redesign of the sign (under study). The life cycle of the sign design/ redesign is proposed here. The sign undergoes various phases of redesign/ design procedure. The open discussion with the designers/ experts after the ranking experiment with the designed variants assisted in molding the design further which was found to be more comprehensible and acceptable by the participants. An attempt has been made to provide a clear picture of the tacit stages of the sign design/ redesign cycle for better comprehensibility of sign(s). The designed signs COSACOE and COSIT are found to be comprehensible and ready for use.



7

**Discussion: Key findings,
contribution of the thesis,
limitations, and future scope**

Chapter 7: Discussion: Key findings, contribution of the thesis, limitations, and future scope

Abstract

This chapter discusses the overall thesis. Observations and results of the analysis from Chapter 2 to “Chapter 6 have been briefly elucidated in the present chapter with key findings of the present thesis. This explanation includes how research questions have been answered and hypotheses tested. There are recommendations from the current thesis work and a small discussion regarding the implications of the findings. This chapter also discusses the novelty of the research and details the objective fulfillment and contributions of the present research to the knowledge base, methodology, design research, and society. In the end, there are discussions regarding the limitations of the study, future scope, and the overall conclusion of the thesis.

7.1. Discussion

The thesis discusses four comprehensibility evaluation techniques, namely subject-based, expert-based, computational, and eye-tracking techniques, and studies the correlation between the subject-based and expert-based techniques, subject-based and computational techniques, expert-based and computational techniques, and subject-based and eye-tracking techniques. The thesis attempts to study the techniques, highlight the advantages and disadvantages of the techniques used in this research, and design evaluation techniques based on the disadvantages of the previous techniques. Previous studies did not include much on the study of techniques, their correlation with each other, and their comparison, to mention only a few (Arcia et al., 2019). The present thesis shows the correlation of one technique with another technique and the correlation between the subjective score and the objective eye movement variables.

The present study discusses the evaluation techniques based on the ISO standard 67% threshold. ‘Comprehensibility frequency’ in this study measures the total number of respondents who correctly interpreted the sign. The comprehensibility percentage (C_p) is calculated by adding the total number of participants who understood the sign (Comprehensibility frequency), dividing by the total number of participants who participated in the experiment, and multiplying by 100. The C_p equal to or exceeding 67% (ISO standard) is considered comprehensible, and the remaining is incomprehensible.

The thesis is related to evaluation techniques of comprehensibility (understandability) of the OSH signs, not design components. However, knowledge of the design principles is significant here. After implementing the evaluation techniques, the less comprehensible signs have been identified, and the poorly designed sign has been redesigned as a proof of concept and tested further with a ranking method that adds complement to the studied four techniques and helps in finding the most comprehensible sign among many similar kinds of variants (discussed in Chapter 6) before it can be made available for use in industries. However, discussion of this method is out of the scope of the thesis.

The present chapter discusses the key findings/ outcomes of the present research discussed throughout the previous chapters, fulfillment of the objectives, testing of hypothesis, implications of the research, key contributions, recommendations from the study, limitations, and future scope.

7.1.1. Key findings of the present thesis

The salient findings of the present thesis generated from the working chapters (Chapter 2 to Chapter 6) are listed below.

- Four techniques were designed, each an advanced version of the other. Similarly, each technique has its drawbacks and advantages in some aspects.

7.1.1.1. Key Findings from Chapter 1

- Knowledge about various existing studies, the theoretical and conceptual foundation for sign evaluation, and techniques related to the comprehensibility evaluation of OSH signs from the literature.
- Research gap, research questions, aim, and objectives.

7.1.1.2. Key findings from Chapter 2

- The *subject-based technique* is the simplest among all the techniques. The experiment is easy to design; however, proper training needs to be imparted to the participants about the scoring/ rating method before the onset of the experiment. One expert is enough for this experiment.
- Data deviated from a normal distribution (i.e., non-parametric data).
- Friedman's ANOVA test, followed by post hoc analysis, reveals that forty-nine (49) pairs are significantly different in terms of comprehensibility level.
- Wilcoxon signed rank test (descriptive statistics, the strength of correlation, and the p-value <0.05) followed by the significant effect size confirm that signs with OSH-related text affect the comprehensibility of OSH Signs.
- Signs like UBC_T, DNT, NAFUP_T, NAFUP, DNWSH, DNT_T, DNWSH_T, WAFS, and WAFS_T are found to be comprehensible, and WAFS-T being the most comprehensible one. They are ready to be used for organizations.
- COSACOEOT, WRP, COSACOEOT_T, UBC, and WRP_T are incomprehensible, and COSACOEOT is found to be the most incomprehensible one.
- The comprehensibility scores are slightly inconsistent as the participants are novices or just exposed to such scoring procedures. They might have given more scores when they

did not comprehend the sign. They might have scored less when they comprehended the sign appropriately.

- The *expert-based technique* is slightly more complex than the subject-based technique. A minimum of two experts were employed, which seems to somewhat avoid the personal biases of a single expert up to a certain extent, which might have destroyed the participants' views.
- The experts must sit for an agreement and decide upon a protocol to check the responses before the commencement of the experiment.
- The trained experts helped to attain high interrater reliability (Cronbach's alpha) for the two experts (alpha- 0.872), which is excellent for our study.
- Cronbach's alpha of 0.872 demonstrates a perfect internal consistency of the scale, the score assigned by the experts, and the internal consistency between the experts. The reliability of the scales and the excellent correlation strength between the data validate the scales and the item statistics for the two experts for all the signs.
- The mean and standard deviation for each sign for the two experts is almost similar, which is the desired result.
- Descriptive statistics characterize the data, followed by the normality test. The data is not normally distributed, which is significant considering the response is from the same participants who scored in the subject-based technique. The data was also not normally distributed in the subject-based technique.
- Friedman's ANOVA test, followed by post hoc analysis, reveals that Sixty-five (65) pairs are significantly different in terms of comprehensibility level.
- Wilcoxon signed rank test (descriptive statistics, the strength of correlation, and the p-value <0.05) followed by the significant effect size confirm that signs with OSH-related text affect the comprehensibility of OSH Signs.
- Signs such as DNWSH_T, DNT_T, WAFS, and WAFS_T are found to be comprehensible, and WAFS-T is the most comprehensible.
- NAFUP, COSACOEOT, UBC, WRP, NAFUP_T, DNWSH, COSACOEOT_T, DNT, UBC_T, and WRP_T are found to be incomprehensible; and NAFUP being the most incomprehensible sign among them.
- The expert-based scores suffer from biases of experts and their pre-notions. Experts are human beings, and they might also commit mistakes. They may get fatigued due to the prolonged evaluation of the numerous surveys and questionnaires besides the other EHS jobs. The evaluation and decision work has to be finished in a limited time when there is a demand for various signs and symbols in organizations. The experts are also limited; they, too, need the training to get expertise in the evaluation procedure.

7.1.1.3. Key findings from Chapter 3

- The computational technique eradicates the scoring anomalies present in the subject and expert-based techniques.
- An automated model designed is seen to produce consistent and unbiased scoring. The model is found to generate scores automatically without committing any mistakes.
- The correlation found between the expert's score and the computational score is very high, and this is the significance of the model. The algorithm seems to perform well on any given safety message.
- The match between the human and the computational score also validates that the model has been trained using natural human language data.
- Descriptive statistics characterize the data, followed by the normality test. The data is not normally distributed, which is significant considering the response is from the same participants who scored in the subject-based technique and wrote in the expert-based technique. The data also deviated from the normal distribution in those techniques.
- Friedman's ANOVA test, followed by post hoc analysis, reveals that Sixty-three (63) pairs are significantly different in terms of comprehensibility level. The result validates the computational technique (model) designed based on human expertise.
- Wilcoxon signed rank test (descriptive statistics, the strength of correlation, and the p-value <0.05) followed by the significant effect size confirm that signs with OSH-related text affect the comprehensibility of OSH Signs.
- Signs such as DNWSH_T, DNT_T, WAFS, and WAFS_T are found to be comprehensible, and WAFS-T is the most comprehensible.
- NAFUP, COSACOEOT, UBC, WRP, DNWSH, NAFUP_T, COSACOEOT_T, UBC_T, DNT, and WRP_T are the incomprehensible signs, and NAFUP is the most incomprehensible.
- This automated model is found to be acting as a supporting tool to the OSH experts by assisting the experts in evaluating the comprehensibility of the signs. The model categorizes the signs as comprehensible or incomprehensible based on the survey/questionnaire.
- The human experts only need to push a few buttons and get the result. This model can evaluate as many signs as the organization wants to evaluate parallelly within seconds without involving the experts. The automated model's current version can be installed on any PC, anywhere, and without complicated hardware with an internet facility. This is indeed a piece of good news for any EHS department in an organization.
- Moreover, this model can be used in any organization with varying severity, context of use, and danger because there is an option where the experts can tune the threshold. This feature makes the model flexible to be adapted to any situation.

- The computational technique is a novel technique for the comprehensibility evaluation of an OSH sign within seconds, involving no human brains during the generation of the scores parallelly for numerous signs; however, the data is subjective, written by the participants/ subjects.
- Often, the subjects are unable to express what they see. Sometimes they are biased due to many unknown reasons. They think something and write something else. The situation applies to the other two techniques (subject-based and expert-based). This information was found when the subjects were interviewed after the experiment.

7.1.1.4. Key findings from Chapter 4

- To avoid biases from subjective data, an eye-tracking device has been exploited to collect objective data.
- A few difficulties were faced during the experiment. Searching for participants has been difficult. The experiment demanded that they do not have any eye-related impairments (color blindness and power).
- A within-group experiment demands the same participants to be contacted after six months. Getting all the same participants again is difficult.
- The time spent during eye calibration before the experiment makes many participants restless. The participants sometimes got frustrated with the calibration process.
- Both subjective and eye-tracking data deviated from the normal distribution.
- Descriptive statistics characterize the data, followed by the normality test. The subjective and eye-tracking data is not normally distributed.
- Friedman's ANOVA test, followed by post hoc analysis, reveals that thirty-two (32) pairs are significantly different in terms of comprehensibility level.
- Wilcoxon signed rank test (descriptive statistics, the strength of correlation, and the p-value <0.05) followed by the significant effect size confirm that signs with OSH-related text affect the comprehensibility of OSH Signs.
- A bivariate correlation study (Pearson, Kendall's tau, and Spearman's rho) between subjective and objective data found thirty-one (31) variables significant (p-value<0.05) out of forty-six (46) variables. Of these 31 variables, 16 are significant and have a fair correlation (medium) with the comprehensibility score. The rest 15 variables are found to be significant.
- Variables, for example, *Sac_velMin* (for signs DNWSH, WAFS, DNT), *Blink_durMax* (for signs WAFS, DNT_T), *Sac_velAvg* (for signs DNT, WAFS_T), *Blink_cnt* (for signs NAFUP_T, DNT_T), *Blink_freq* (for signs NAFUP_T, DNT_T), *Blink_dur* (for signs NAFUP_T, DNT_T), *Scanpath_len* (for signs COSACOEOT, WAFS_T), *Sac_amplMin* (for signs NAFUP_T, COSACOEOT_T, UBC_T) and *AOI_size004_Revisits* (for signs

COSACOEOT, *WAFS-T*) have been found to have a medium positive correlation with comprehensibility score.

- Again *Vis_intk_durMin* (for signs *NAFUP*, *WRP_T*), *Vis_intk_durtot* (for signs *WAFS*, *WRP*), *Vis_intk_durMax* (for signs *COSACOEOT*, *WRP*), *Vis_intk_durAvg* (for signs *NAFUP*, *NAFUP_T*), *AOI_size004_Vis_intkTime* (for signs *WAFS*, *WRP*, *WRP_T*), *AOI_size005_Vis_intkCnt* (for signs *WAFS*, *WRP*, *COSACOEOT*), and *AOI_size004First_vis_intkdur* (for signs *WRP*, *DNWSH-T*, *NAFUP-T*) have been found to have a medium negative correlation with comprehensibility score.
- The Kruskal-Wallis test followed by a Posthoc test found the eye-tracking variables such as *Vis_intk_durAvg*, *Blink_durAvg*, *AOI004Firstvis_intk_dur*, *AOI004Avg_vis_intk_dur*, *AOI005Firstvis_intk_dur*, and *AOI005Avg_vis_intk_dur* varies across categories of comprehensibility-levels (comprehensible and incomprehensible signs), supported by the p-value which is less than 0.05 and is significant.
- The objective measures are significant and can be included in the evaluation strategy of comprehensibility levels of OSH signs.

7.1.1.5. Key findings from Chapter 5

- The Pearson plot and bivariate correlation using three methods (Pearson, Kendall's Tau, and Spearman's Rho) for the three techniques (subject-based technique, expert-based technique, and computational technique) found a good correlation between them (evident from their strength of correlation and the p-value < 0.05).
- A less correlation was found between the subject-based and the expert-based scores, subject-based scores, and computer-generated scores compared to the correlation between the expert-based and computer-generated scores. The reason is that the participants are novices or just exposed to such scoring procedures, and the comprehensibility scores could be slightly inconsistent. They might have given more scores when they did not comprehend the sign. They might have scored less when they comprehended the sign appropriately.
- The computer-generated scores in the computational technique are consistent and unbiased. They are highly correlated with the scores generated by expert-based techniques. This is evident from their correlation study. The computational technique is developed, simulating how human experts score a message.
- Similarly, a bivariate study between the subjective data and the eye-tracking data reveals that: The parameters *Sac_velMin* (*DNWSH WAFS*, *DNT*), *Blink_durMax* (*WAFS*, *DNT_T*), *Sac_velAvg* (*DNT*, *WAFS_T*), *Blink_cnt* (*NAFUP_T*, *DNT_T*), *Blink_freq* (*NAFUP_T*, *DNT_T*), *Blink_dur* (*NAFUP_T*, *DNT_T*), *Scanpath_len* (*COSACOEOT*, *WAFS_T*), and *Sac_amplMin* (*NAFUP_T*, *COSACOEOT_T*, *UBC_T*) have a negative correlation with comprehensibility score.
- Similarly, parameters like *Vis_intk_durMin* (*NAFUP*, *WRP_T*), *Vis_intk_durtot* (*WAFS*, *WRP*), and *Vis_intk_durMax* (*COSACOEOT*, *WRP*) have a positive correlation with comprehensibility score.

- *Vis_intk_durAvg* (NAFUP, NAFUP-T) has a positive correlation with the comprehensibility score.
- The outcome of the significant parameters of AOI statistics is as follows: *AOI_size004_Vis_intkTime* (WAFS, WRP, WRP-T), *AOI_size005_Vis_intkCnt* (WAFS, WRP, COSACOEOT), *AOI_size004First_vis_intkdur* (WRP, DNWSH-T, NAFUP-T) have a negative correlation with comprehensibility score.
- Similarly, *AOI_size004_Revisits* (COSACOEOT, WAFS-T) have a positive correlation with the comprehensibility score.
- The results align with the literature found and thus foolproof the validation of the technique used in this research.
- The comparative study also reveals that DNT-T, WAFS, DNWSH-T, and WAFS-T are comprehensible in all the four techniques applied here. WAFS-T scored the maximum percentage in the first three techniques (subject-based technique, expert-based technique, and computational technique).
- A separate set of participants from the previous three techniques provides the subjective score in the eye-tracking experiment; hence it is normal to get a different result. However, WAFS-T scored second highest in this case which validates the techniques.
- COSACOEOT and WRP are common in the incomprehensible groups found in all the techniques. COSACOEOT is found to be least comprehensible in expert-based and computational evaluation techniques. It ranked second most least comprehensible, scoring 9.48%, which is slightly less incomprehensible than the NAFUP (6.56%) in the case of expert-based technique and (9.48%), which is also slightly less incomprehensible than the NAFUP (5.83%) in the case of computational technique.
- The results align with the previous literature and validate the models/ techniques used in the study.
- It is essential to note that the bivariate correlation pointed out the significant eye-tracking variables, which are the same as found in the correlation plot variables when the comprehensibility threshold has been raised from 67% to 85% in the eye-tracking experiment. The results are not included for brevity.
- A separate small comparative study between the DAN algorithm (computational model) and the other related algorithms (computational models) reveals DAN model works on par with other algorithms discussed. This finding opens a door for further improvement of the proposed (DAN-based model).
- The comparison of the techniques based on parameters will provide new researchers and new OSH specialists/ experts with a preliminary idea about the feasibility of a technique to be implemented in their organization. They may first come up with a report of their organization's feasibility study (economic, technical, and behavioral) before selecting any particular type of evaluation technique or a combination of techniques.

7.1.1.6. Key Findings from Chapter 6

- The study of the four techniques with 14 signs found that COSACOEOT (Cough or sneeze at the crease of the elbow or in tissue) is the most confusing and the least comprehensible.
- COSACOEOT is found to have two messages in one pictogram, which creates ambiguity and goes against the protocol of the sign design. Hence the message has been divided into two parts- i.e., Cough or sneeze at the crease of the elbow, and ii. Cough or sneeze into a tissue.
- For redesigning, incremental designs are found to be the best option considering the ease of understanding and less cognitive load.
- The proof-of-concept design validates the sign design/ redesign life cycle and the evaluation framework based on a User-centred Design concept (UCD).
- The data collected helped in further analysis and proving the hypothesis.
- The evaluation framework developed is the conglomeration of different techniques. The validation of the techniques validates the whole framework generated. Every part of the framework is validated, evident from the validation of each technique.
- The study reveals that the whole or part of the framework can be used to evaluate signs. The triangulation of techniques can also be done, which will work better, as evidenced by the evaluation result from different techniques implemented here. The common comprehensible and incomprehensible signs from different techniques are the framework's outcomes, thus validating the framework.
- The evaluation framework with the UCD-based sign design/ redesign cycle is a novel contribution to the design society and humankind.

7.1.2. Fulfillment of the objectives

Objective 1: To study the existing comprehensibility evaluation techniques for the OSH signs.

The existing evaluation techniques for the comprehensibility of signs related to OSH are studied in detail in Chapter 1. Thus, '*objective 1*' of the research is fulfilled.

Objective 2: To collect data regarding the comprehensibility of the OSH sign through a survey and eye-tracking study.

The online questionnaire-based survey (Chapters 2 and 3), a questionnaire-based online study (Chapter 4), and an eye-tracking experiment (Chapter 4) have been conducted to collect data regarding the comprehensibility of the OSH sign. Initially, an online survey was done for the three techniques (subject-based, expert-based, and computational technique), where 137 data

were collected for various participants working for an organization in various states of India. The sample size has been decided using Cochran's formula, which is calculated to be a minimum of 133 for the present study. All participants were healthy and white-collar employees of a food manufacturing company with a mean age of 33.85 and a standard deviation of 7.2 (detailed in Chapters 2 and 3). For the eye-tracking experiment, a separate set of 41 samples was collected, conducting a questionnaire-based study and an eye-tracking experiment. Subjective and objective data related to the comprehensibility study of the OSH sign (detailed in Chapter 4) have been collected. Thus, 'objective 2' of the research is fulfilled.

Objective 3: To study the level of comprehensibility of the existing few OSH sign(s) by subject-based and expert-based techniques based on the data collected from the survey.

The subject-based technique has been implemented using seven pairs of signs (7 pairs with and without OSH text). The comprehensibility level of the signs has been studied. The signs in the increasing order of comprehensibility are as follows: COSACOEOT < WRP < COSACOEOT_T < UBC < WRP_T < UBC_T < DNT < NAFUP_T < NAFUP < DNWSH < DNT_T < DNWSH_T < WAFS < WAFS_T. The signs have been categorized under two levels- comprehensible and incomprehensible. The study reveals that COSACOEOT, WRP, COSACOEOT_T, and UBC are incomprehensible signs. WRP_T, UBC_T, DNT, NAFUP_T, NAFUP, DNWSH, DNT_T, DNWSH_T, WAFS, and WAFS_T are comprehensible signs. More details about data collection can be found in Chapter 2.

The expert-based technique has been implemented using the same sign pairs. The comprehensibility level of the signs has been studied. The signs in the increasing order of comprehensibility are as follows: NAFUP < COSACOEOT < UBC < WRP < DNWSH < NAFUP_T < COSACOEOT_T < DNT < UBC_T < WRP_T < DNWSH_T < DNT_T < WAFS < WAFS_T. More sequences of comprehensibility (order) can be found in Chapter 2, where the comprehensibility level is the same for more than one sign. The signs have been categorized under two levels- comprehensible and incomprehensible. The expert-based evaluation found DNWSH_T, DNT_T, WAFS, and WAFS_T as comprehensible with, WAFS_T as the most comprehensible similar to the subject-based technique. The rest of the signs NAFUP, COSACOEOT, UBC, WRP, DNWSH, NAFUP_T, COSACOEOT_T, DNT, UBC_T, and WRP_T fall under the category of incomprehensible signs.

A separate Wilcoxon signed rank test followed by effect size calculation has been shown for both techniques to show that OSH sign with OSH-related textual information affects the understanding or comprehensibility of the messages intended to convey (detailed in Chapter 2). Thus 'objective 3' is fulfilled.

Objective 4: To study the level of comprehensibility of the existing few OSH signs (s) by a computational technique based on the data collected from surveys.

The computational technique has been implemented using seven pairs of signs (seven pairs with and without text). The data has been collected from the survey, and details can be found in Chapter 3. The comprehensibility level of the signs has been studied. The signs in the increasing order of comprehensibility are as follows: NAFUP < COSACOEOT < UBC <

WRP< DNWSH< NAFUP_T< COSACOEOT_T< UBC_T< DNT< WRP_T< DNWSH_T< DNT_T< WAFS< WAFS_T. The signs have been categorized under two levels- comprehensible and incomprehensible. The study reveals that COSACOEOT, COSACOEOT_T, DNT, DNWSH, NAFUP, NAFUP_T, UBC, UBC_T, WRP, and WRP_T are incomprehensible signs. DNWSH_T, DNT_T, WAFS, and WAFS_T fall into the comprehensible level. A separate Wilcoxon signed rank test followed by effect size calculation shows that OSH signs with OSH-related textual information affect the understanding or comprehensibility of the messages intended to be conveyed (detailed in Chapter 3). Thus 'objective 4' is fulfilled.

Objective 5: To study the level of effectiveness of comprehensibility of the existing few OSH signs (s) by eye tracking technique.

The eye-tracking technique has been used to study the comprehensibility of the sign. A bivariate correlation study has been conducted to study the existence of a correlation between subjective data (subject-based technique) and eye-tracking data (eye-tracking technique) (Appendix G). The signs have been categorized under two levels- comprehensible and incomprehensible. Friedman's ANOVA test, followed by post hoc analysis, has been performed to see variations in comprehensibility levels among signs. A separate Wilcoxon signed rank test followed by effect size calculation has been done for the subjective data collected to show that signs with text are more comprehensible than signs without text (detailed in Chapter 4). Another ANOVA test (Kruskal-Wallis analysis followed by effect size) has been conducted to see the variation of the objective data across categories (two levels of comprehension- comprehensible and incomprehensible signs). Correlation (Pearson) plots demonstrate that the objective measures like average pupil diameter (left and right), visual intake duration, and saccade latency average tend to decrease with increased comprehensibility of signs, visual intake count, blink count, blink frequency, blink duration, saccade count, saccade amplitude, saccade duration total, and scan path length tends to increase with increase comprehensibility of the signs. The details can be seen in Chapter 4. Thus 'objective 5' is fulfilled.

Objective 6: To study the level of effectiveness of the computational method in comparison to the manual evaluation (subject-based and expert-based techniques).

To prove that *Objective 6* is fulfilled, *sub-objectives 6.1* and *6.2* must be fulfilled.

Sub-objective 6.1. To study the effectiveness of the computational technique in comparison to subject-based evaluation.

A Pearson correlation (bivariate correlation) test has been conducted, and it has been found that for all the studied signs, the Pearson r score is found to be above 0.35 with a p value<0.01 and p value<0.05, respectively (model validation). This value helps to validate the techniques and generalize the result confirmed by subject-based and computational techniques. The result is significant and encouraging, considering the p-value less than 0.05. The details can be found in Chapter 5, Table 5. 2, and Table 5. 3.

Sub-objective 6.2. To study the effectiveness of the computational technique in comparison to expert-based evaluation.

A Pearson correlation test has been conducted while generating the scores in the computational model, and it has been found that for all the studied signs, the Pearson r and Spearman's ρ score was found to be above 0.80 for the sentence semantic similarity of the proposed system against human judgments (model validation against STS benchmark). This value helped validate the techniques and generalize the result confirmed by expert and computational techniques. The result is significant and encouraging, considering the p -value less than 0.05. The details can be found in Chapter 3. A separate bivariate correlation study between the manual expert evaluation technique (score from the experts) and computational technique elicited a good correlation evident from their p -value. The details can be found in Chapter 5, Table 5. 2, and Table 5. 3.

Hence, from the two sub-objectives (*Sub-objective 6.1., and Sub-objective 6.2.*), it is seen that *objective 6* is fulfilled.

Objective 7: To study the level of effectiveness of the eye-tracking method in comparison to manual (subject-based technique) evaluation.

The effectiveness of the eye-tracking method in comparison to manual evaluation (subject-based data) has been studied using a Pearson plot (refer to 4.3.2.1) followed by bivariate correlation analysis (Pearson's r , Spearman's ρ , and Kendall's tau– refer to 4.3.2.2) to examine the existence of a relationship (correlation) between the two evaluation techniques. The effectiveness of the eye-tracking technique can be demonstrated by its correlation strength with the manual expert evaluation technique. There is a good correlation between the eye-tracking technique and the manual evaluation. Regression line or Pearson plots demonstrate that objective measures like average pupil diameter (left and right), visual intake duration, and saccade latency average tend to decrease with increased comprehensibility of signs; visual intake count, blink count, blink frequency, blink duration, saccade count, saccade amplitude, saccade duration total, scan path length tends to increase with increase comprehensibility of the signs. These are evident from the r square value. The results are in line with the previous literature. Thus, '*objective 7*' is fulfilled.

Objective 8: To develop and validate the framework for the comprehensibility evaluation of OSH signs.

The evaluation framework has been designed (Chapters 2, 3, and 4) and validated using 14 signs in Chapter 5, section 5.3.4. The outcome of the framework is the two categories of signs, comprehensible and incomprehensible. The signs' level of comprehensibility also validates the techniques by categorizing the signs into common groups. The incomprehensible signs are reconsidered for a redesign and revaluation. Finally, a UCD-based sign design/redesign cycle has been proposed in Chapter 6, Section 6.2, which validates the final design. A proof-of-concept COSACOEOT has been generated from the evaluation framework, which is one of the least comprehensible signs, and all the techniques (the evaluation framework) confirm the same. The sign COSACOEOT is the outcome of the

framework (discussed in Chapters 2, 3, and 4), which has been redesigned in Chapter 6 and validated by a ranking method (refer to Chapter 6, Section 6.2 -for details). The framework has been further validated by the redesign of the COSACOEOT into two variants COSACOE and COSIT.

7.1.3. Hypothesis testing

This section shows the proofs of hypothesis tests conducted and proved throughout the different chapters in the thesis. Three main hypotheses have been tested based on the research questions, with one having three working hypothesis proofs.

Hypothesis 1

Null Hypothesis 1 (H_0): Comprehensibility scores are significantly different and not correlated, while evaluations are compared among the subject-based technique, expert-based technique, and computational technique.

Alternative Hypothesis 1 (H_1): Comprehensibility scores are significantly different but correlated, while evaluations are compared among the subject-based technique, expert-based technique, and computational technique.

Proof: To prove this main hypothesis, three working hypotheses have been proved.

Working Hypothesis 1

Null Working Hypothesis (H_0): Comprehensibility score is significantly different and not correlated, while expert evaluation is compared with subject-based evaluation.

Working Hypothesis a (H_a): Comprehensibility score is significantly different but correlated, while expert evaluation is compared with subject-based evaluation.

For testing the working hypothesis 'a', direct comprehensibility scores have been collected from the subjects/ participants (refer to 2.2.3 for the subject-based technique data collection and Appendix B for scores). The comprehensibility scores from the experts were collected who rated or scored the textual responses of the participants. The subjects or participants wrote textual descriptive responses. Two experts evaluated the responses, and finally, scores were collected (refer to 2.2.3, an expert-based technique for data collection, and Appendix C for scores). Both correlation plots (Figure 5. 1- for COSACEOT only, all the plots for all the signs could not be included in the thesis due to space constraints) and bivariate correlations (refer to Table 5. 2 and Table 5. 3) have been conducted for all 14 signs (7 pairs- with and without text). The analysis shows we can reject the null hypothesis (p -value < 0.01 and p -value < 0.05). Hence, these findings favor the alternate working hypothesis that the *Comprehensibility score is significantly different but correlated, while expert evaluation is compared with subject-based evaluation, hence proved.*

Working Hypothesis 2

Null Working Hypothesis (H_0): Comprehensibility score is significantly different and not correlated, while expert evaluation is compared with computational evaluation.

Working Hypothesis b (H_b): Comprehensibility score is significantly different but correlated, while expert evaluation is compared with computational evaluation.

The comprehensibility scores from the experts are collected indirectly to test the working hypothesis. The subjects or participants wrote textual descriptive responses. The two experts rated the responses, and their scores were collected (refer to 2.2.3 for expert-based technique and Appendix C for scores). Similarly, the participants wrote textual descriptive responses, which are evaluated automatically by the computational technique (refer to 3.2.3 for the computational technique and Appendix D for scores). Both correlation plots (Figure 5. 2 for COSACEOT only, all the plots for all the signs could not be included in the thesis due to space constraints) and bivariate correlations (refer to Table 5. 2 and Table 5. 3) have been conducted for all 14 signs (7 pairs- with and without text). The analysis shows we can reject the null hypothesis (p -value < 0.01 and p -value < 0.05). Hence, these findings favor the alternate working hypothesis that the *Comprehensibility score is significantly different but correlated, while expert evaluation is compared with computational evaluation, hence proved.* In reality, the correlation between these techniques is very high.

Working Hypothesis 3

Null Working Hypothesis (H_0): Comprehensibility score is significantly different and not correlated, while subject-based evaluation is compared with computational evaluation.

Working Hypothesis c (H_c): Comprehensibility score is significantly different but correlated, while subject-based evaluation is compared with computational evaluation.

For testing the working hypothesis 'c', direct comprehensibility scores have been collected from the subjects or participants (refer to 2.2.3 for subject-based technique and Appendix B for scores). Similarly, the participants wrote textual descriptive responses, which are evaluated automatically by the computational technique (refer to 3.2.3 for computational technique and Appendix D for scores). Both correlation plots (Figure 5. 3 for COSACEOT only, all the plots for all the signs could not be included in the thesis due to space constraints) and bivariate correlations (refer to Table 5.2 and 5.3) have been conducted for all 14 signs (7 pairs- with and without text). The analysis shows we can reject the null hypothesis (p -value < 0.01 and p -value < 0.05). Hence, these findings favor the alternate working hypothesis that the *comprehensibility score is significantly different but correlated, while subject-based evaluation is compared with computational evaluation., hence proved.*

Hypothesis 2:

Null hypothesis (H₀): There is no correlation between the subjective response of the comprehensibility and the objective measures (eye tracking variables: average pupil diameter, visual intake count, visual intake duration, blink count, blink duration, blink frequency, saccade count, saccade amplitude, saccade latency, saccade duration, scan path length).

Alternative Hypothesis 2 (H₂): There is a significant correlation between the subjective response of the comprehensibility and the objective measures (eye tracking variables: average pupil diameter, visual intake count, visual intake duration, blink count, blink duration, blink frequency, saccade count, saccade amplitude, saccade latency, saccade duration, scan path length).

To test the hypothesis, direct comprehensibility scores have been collected from the subjects (refer to 4.2.3 for subject-based scores in the eye-tracking technique and Appendix E for scores). Similarly, the participants participated in the eye-tracking experiment, and eye-movement records were collected from the eye recorder using Begaze software (refer to 4.2.3 and Appendix F for eye-tracking data). Both correlation plots (refer to 4.3.2.1, Figure 4. 12 to Figure 4. 22 for Pearson plots) and bivariate correlations (refer to 4.3.2.2; Appendix G for bivariate correlation results) have been conducted for all 14 signs (7 pairs- with and without text). The analysis shows we can reject the null hypothesis (p -value < 0.01 and p -value < 0.05). Hence, these findings favor the alternate hypothesis that *there is a significant correlation between the subjective response of the comprehensibility and the objective measures (eye tracking variables: average pupil diameter, visual intake count, visual intake duration, blink count, blink duration, blink frequency, saccade count, saccade amplitude, saccade latency, saccade duration, scan path length), hence proved.*

Hypothesis 3:

Null hypothesis (H₀): The level of comprehensibility of the OSH sign is not significantly different when the sign is presented with or without textual information.

Alternative Hypothesis 3 (H₃): The level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information.

Testing this hypothesis involved collecting scores from each technique (Appendices B, C, D, and E).

- i. For the subject-based technique, a Wilcoxon signed rank test has been conducted. The effect size is significant, which reveals that for all the six sign pairs, it is found that the level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information evident from their mean difference, z value, p -value, and r value (Section 2.3.1.4, Table 2. 6 - Table 2. 9) with a minor variation for NAFUP-NAFUP_T.

- ii. For the expert-based technique, a Wilcoxon signed rank test has been conducted. The effect size is significant, which reveals that for all the seven sign pairs, it is found that the level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information evident from their mean difference, z value, p-value, and r value (Section 2.3.2.5, Table 2.20 – Table 2.23).
- iii. For the computational technique, a Wilcoxon signed rank test has been conducted. The effect size is significant, which reveals that for all the seven sign pairs, it is found that the level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information evident from their mean difference, z value, p-value, and r value (Section 3.3.5, Table 3.6 – Table 3.9).
- iv. For the eye-tracking technique, subjective data is collected separately, a Wilcoxon signed rank test is conducted, and the effect size is significant, which reveals that for all the seven sign pairs, it is found that the level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information evident from their mean difference, z value, p-value, and r value (Section 4.3.4, Table 4. 10 - Table 4. 13).

The analysis shows we can reject the null hypothesis ($p\text{-value} < 0.05$). Hence, these findings favor the alternate hypothesis that *the level of comprehensibility of the OSH sign is significantly different when the sign is presented with or without textual information, hence proved.*

7.1.4. Implications of the findings of the present research

Implication(s) from the perspective of primary data

The data collected for all the techniques are significant, considering that they are primary and first-hand data and can be used for many experimental purposes. The techniques implemented helped identify the comprehensible and incomprehensible signs from the OSH perspective and validate the techniques, data, and signs. These data can be further processed to create more categories of comprehensibility levels and will assist in capturing more minds of the target users. Such a discussion is out of the scope of this research.

Implication(s) from the perspective of designing a new tool/ decision-making system

The designed computational model acts as an assistive tool for the EHS/ OSH professionals/ sign designers in evaluating the comprehensibility of a sign. The system can decide whether to accept the sign as comprehensible or reject it and categorize it as incomprehensible. The designed technique is equipped with a feature (thresholding function) that can be tuned to any percentage (1 to 100) depending on the severity of the context, organization, danger, and situation for evaluating the sign. The tool works equally well when the responses of the participants are based on keywords or synonyms of the standard interpretation rather than any

complete sentence or the exact word used in the standard interpretation of the sign considered for the study.

Implication(s) from the perspective of the design process

- The overall research here is a glimpse of “research through design (RtD)” and “research for design.” “Research through Design (RtD) is an approach for conducting scholarly research which utilizes the methods, practices, and processes of design practice with the intent to generate new knowledge (J. Zimmerman & Forlizzi, 2014)”.
- Research for design is an approach where data is analyzed, and the outcome of the experiment/ research project is a type of knowledge/ new theory/ modified theory/ new information that is produced (Herriott, 2019; J. Zimmerman & Forlizzi, 2014). Research for design “includes almost all design research that may propose new method/ tools and techniques/ approaches or any work that uses exemplars, design implications, or problem framings which discuss improving the practice of design” (reworded from Zimmermann) (J. Zimmerman & Forlizzi, 2014).
- The sign developed underwent an incremental development process before getting selected. The research employed the methods/ techniques, practices, and process of design practice to generate new knowledge/ recommendations (J. Zimmerman & Forlizzi, 2014).
- Generally, a systematic design methodology or the basic structure of the design process follows analysis, synthesis, and evaluation (Cross, 2021; French et al., 1985). Archer (Archer, 1984) identified six types of activity commonly applied in designing a solution, which has been applied here. The model includes (a) Programming – establishing major and critical issues and proposing a course of action. (b) Data collection – collection, classification, and storing of data. (c) Analysis – identifying sub-problems; preparing performance (or design) specifications; reappraisal of the proposed program and estimation of the same. (d) Synthesis – preparing an outline of the design proposals. (e) Development – developing prototype design(s); preparing and executing validation studies. (f) Communication – preparing manufacturing/ design documentation.

There are classical models for designing software/ tools, such as the waterfall model (hardly used nowadays due to its sequential steps and no feedback loop to previous stages), which is the basis for other models. This waterfall model follows the same design process narrated by ‘Cross’ and falls under prescriptive process models.

Considering the automated model developed, the model undergoes the phases of the design process, i.e., analysis, synthesis, and evaluation. The model is finally validated using statistical analysis (bivariate correlation: model score with respect to the human expert score). Seven pairs of signs also undergo the phases of the design process, i.e., analysis, synthesis, and evaluation. These signs are evaluated and validated by the evaluation techniques implemented here. The whole process generated a knowledge base that is tacit, and this knowledge has been documented as recommendations and discussions.

In this study, the sign design process, the software design process, and the other evaluation techniques designed went hand in hand. The design process became a form of research (W. Gaver, 2012; Tekinbas & Zimmerman, 2003; J. Zimmerman et al., 2010).

- The thesis focuses on the analysis, design, and evaluation stage of the design process from the perspective of comprehensibility evaluation techniques. The study initializes with identifying the research gap, followed by analyzing the problem (s) from the literature and the necessity to study the evaluation techniques. Then research design has been planned to study the comprehensibility levels of safety signs. Then data is collected for studying the comprehensibility levels using four different techniques/ methods. Each technique is studied one by one. The first two techniques used subjective data for the study, and the study is based on manual comprehensibility evaluation of signs. In the third study, the data is subjective; however, the manual evaluation process has been transformed and designed into an automated system. The fourth study is based on an eye-tracking device that helped collect objective data and find significant measures and their correlation with two comprehensibility levels.
- The automated tool design process encountered empathizing, which necessitates understanding the existing expert's manual procedure of evaluating the comprehensibility of OSH signs. A new system is designed where the traditional method of evaluating comprehensibility has been maneuvered into an automated one. The change process involved understanding how the human experts rate/ score the surveys' responses, how comprehensibility is evaluated, and how the experts categorize signs as comprehensible or not comprehensible. A computational model is developed to automate the comprehensibility evaluation procedure based on the expert's feedback. The experts finally validate the model, evident from the bivariate correlation result and the model's performance.
- Similarly, the sign design phase encountered empathizing with the existing issues with the sign and evaluating the sign's comprehensibility. They are ideated, conceptualized, designed, and redesigned iteratively till the final design is validated by experts or any evaluation technique.
- Considering the sign design and the design of automated decision-making systems, both went through rigorous research, or the research is supported by design (OSH sign, techniques, principles, practices, concepts). The two terms research and design sound inseparable (Tekinbas & Zimmerman, 2003). Both activities generated new knowledge. The overall phases are shown in Figure 7. 1.

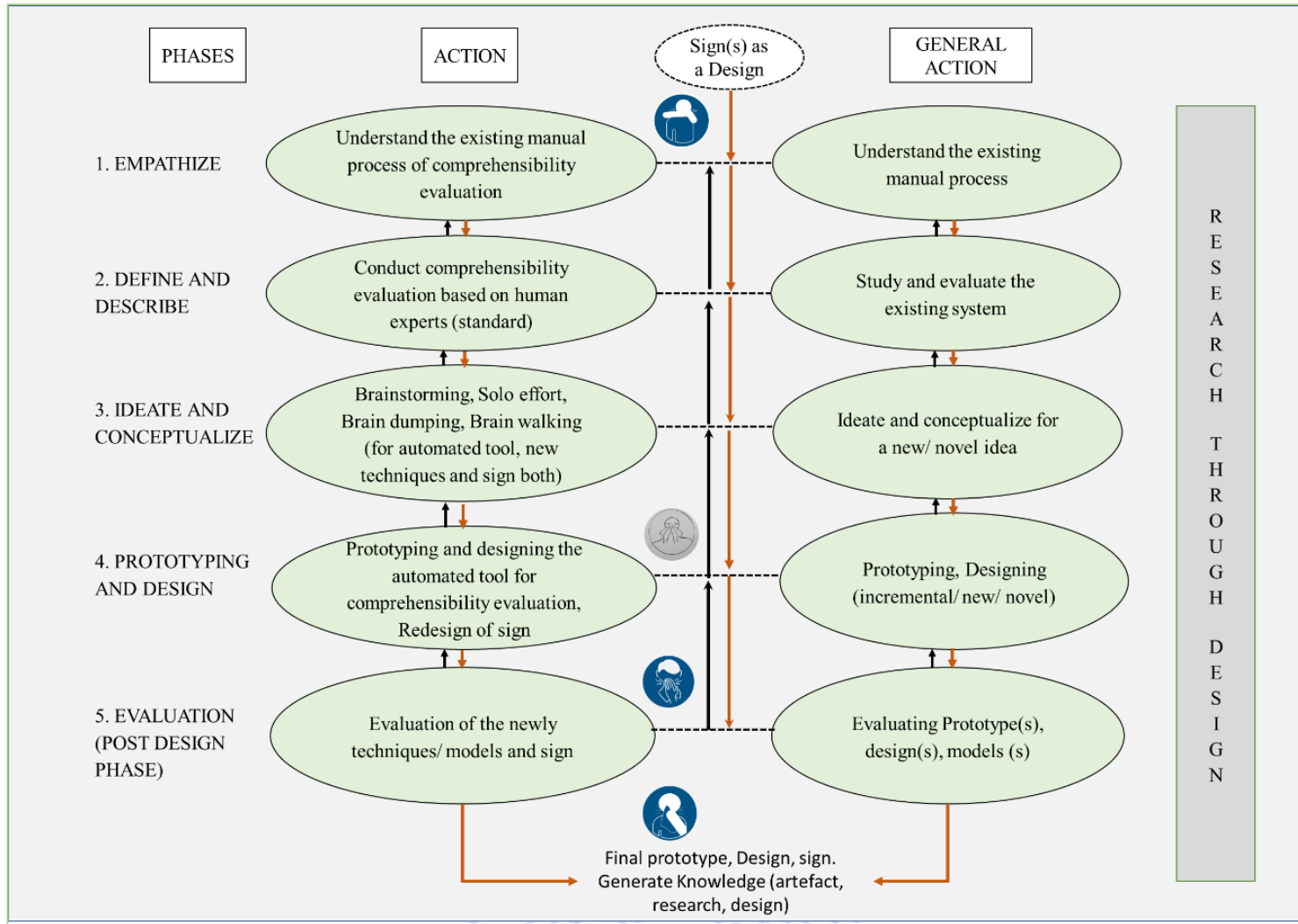


Figure 7. 1. Implication from the perspective of the Design process and Human-centered approach

Implication(s) from the perspective of a human-centered approach

Throughout the research, the focus is on humans to improve their lives, both in terms of users of the signs and OSH experts/ designers of the organizations who are humans too. The users are the people of the organizations for whom the signs are being designed. The proper comprehensibility evaluation techniques will improve the selection of the appropriate signs for the target users and eliminate incomprehensible signs, thus reducing the OSH hazards and dangers in the workplace. Appropriate sign selection will help humankind directly by preventing OSH hazards. All the techniques/ methods implemented here are based on a human-centered approach.

Implication(s) from the perspective of generalization of results and validation of techniques

Four techniques have been applied for the comprehensibility evaluation. Application of all the techniques/ triangulation of techniques/ methods helped better identify the comprehensible and incomprehensible signs. This triangulation property helped generalize the techniques/ methods and the results drawn from the study. For example, the comprehensible signs that are commonly found after evaluating with all the evaluation techniques are DNWSH_T, DNT_T, WAFS, and WAFS_T. COSACOEOT and WRP have been commonly found to be incomprehensible in all the techniques studied here. These common signs for the comprehensible and incomprehensible groups for all the techniques, accompanied by the bivariate correlation results between all the pairs of techniques, helped validate the techniques studied.

7.1.5. Key Contributions of the present research

The outcomes of the research have a significant contribution to the knowledge base; design methodology, design and development of automated human-centered comprehension evaluation strategy of a sign, UCD-based sign design/redesign cycle for comprehensibility evaluation, and finding significant features of eye movement to analyze sign comprehension by human (human behavior). Various statistical analyses and numbers have proved the robustness of the framework.

7.1.5.1. Contribution to knowledgebase

- The designed questionnaire can be used by the researchers or factory EHS expert(s) as a reference to study the comprehensibility of OSH signs and symbols in factories, educational institutions, and any organization. The questionnaire is validated and is a contribution to the knowledge base in the domain of OSH sign design and evaluation.
- There is a scarcity of literature discussing the techniques/ methodologies followed in the comprehensibility evaluation of signs. This research contributes by discussing four detailed techniques, their limits, and delimits. The research also discusses various techniques prevailing to date in the literature. This will assist future researchers and readers in applying the techniques according to the requirements of the organization and their feasibility.

- Binary classification at the end is considered where the comprehensibility level categories are comprehensible/ incomprehensible even though eleven-point rating scale responses have been collected. This is intentional, and the reason behind the eleven-point rating scale is to capture the broader ideas/ minds of the end users. The same database can be used again by scaling them down and easily converting to any other scale and studied further. So this knowledge base becomes a foundation for further research.
- An effort has been made to put forward the tacit knowledge of the OSH designers/ experts. “Designers are not used to accounting for what they know or do” (Pedgley, 2007). This is not an issue when practicing design. However, it becomes problematic as soon as the said knowledge needs to be explicated when teaching, publishing, or documenting research results and analysis. An attempt has been made to implement practice-led research in design or research through design (Pedgley, 2007).

7.1.5.2. *Contribution toward methodological perspective (design research methods)*

- Studying various techniques of comprehensibility evaluation simultaneously and finding their advantages and drawbacks is a contribution in itself. Hence, this study can be claimed as a contribution to the methodology.
- The literature survey for all the techniques elicited many new gaps and areas that can be the foundations for future research. The disadvantages or limitations discussed in each technique can trigger new or advanced research for readers of the thesis.
- The novel computational technique for OSH comprehensibility evaluation developed in a conglomeration of new and existing knowledge opened a new idea/ genre that evaluation can be automated without human intervention. This technique will assist in evaluating and fast processing of large amounts of subjective data which is otherwise time-consuming and demands the involvement of experts. It also helps to reduce biases by the experts in the comprehensibility evaluation process. Moreover, the computational technique can also be applied across various cultures and for large datasets.
- The study of the objective data extracted from the eye-tracking technique is also a novel contribution that helped identify the significant features/ measures correlated with levels of comprehensibility of OSH signs.
- The framework assists in identifying parameters that may help in selecting the technique(s) of evaluation for any particular organization.
- A significant contribution from the methods perspective is the proposed framework/ paradigm of evaluation, a step-by-step method to solve problems. All the techniques implemented in the thesis follow this approach. The sign design/ redesign cycle has been implemented from the perspective of the target users (User-centred design approach) with basic steps such as clarification (feasibility study), ideation (conceptualize), development (design), and implementation (Use and verification). Hence, a step-by-step procedure is also implemented for the sign design phase. The designed life cycle is also a contribution to the methodology of sign evaluation and design.

7.1.5.3. Contribution towards design community and design management

- The tacit knowledge gained from this research (research through evaluation and design) is very important for future studies and research in the same domain. This is a novel contribution to the design community.
- The automated evaluation model designed here eliminates the burden of experts/designers. The automated model is designed to acclimatize to any severity and context. It can be tuned to evaluate based on the threshold assigned by the experts.
- The automated sign evaluation in conglomeration with the sign design cycle may fasten the sign evaluation and design process so much that organizations will experience a drastic speed in the sign evaluation and design paradigm (evident from the accuracy and speed of the developed automated model). This is a significant contribution to design management.
- The sign evaluation and design in this scenario is not a sequential process but a concurrent process where both phases occur simultaneously (evident from the evaluation and design life cycle paradigm). This is indeed a contribution to design management. The evaluation framework (discussed in Chapter 2 to Chapter 4) with the UCD-based sign design/ redesign cycle is a novel contribution to the design community and humankind.

7.1.5.4. Contribution to society and organizations

- This research is a boon for a society that might prevent incidents/ accidents by employing fast sign comprehensibility evaluation techniques.
- This research is a direct intervention in safety culture. Such kind of study is needed for the enrichment of knowledge. Safety culture prevents accidents and OSH emergencies in any organization, thus saving people's lives. Hence, it is a major contribution to organizations and society.
- The redesigned signs might be used for minimizing and controlling the spreading of communicable diseases like COVID, adenovirus, viral flu, etc., which spreads through coughing and sneezing openly.

7.1.6. Recommendations from the present thesis

- Comprehensibility evaluation of a sign should be mandated before a sign can be used for any organization.
- Proper evaluation techniques (one or more than one) must be followed before a sign can be standardized for use. Selection of the evaluation techniques again depends on the context of use, target users, availability of the experts, feasible technology, advantages and disadvantages of a technique, its limits, delimits, etc.

- A preliminary investigation must be done based on different parameters discussed in the thesis (Table 5. 4) so that a proper selection of comprehensibility evaluation technique (s) is possible.
- Each sign must be designed following ISO standards to make the sign comprehensible.
- For a design to be comprehensible, an incremental design of a sign should be practiced as it costs less time and money to design and consumes less time to train the target users with less cognitive load. Humans spend more time learning a new concept rather than learning a concept that is somewhat known to them.
- Besides, following the evaluation paradigm for an appropriate and suitable sign to be designed, a sign must undergo a sign design/redesign cycle (a kind of beta testing) before it is released for use.
- Each illustration should communicate only a single, distinct message.
- Signs must be appended with OSH-related textual information to be more comprehensible. This makes the sign more precise in terms of a particular message communication.
- Care must be taken to avoid overloading the sign with unusual excessive graphics or text, making the sign look clumsy and creating ambiguity. The textual information must match and conform with the graphical image or pictogram to create a more meaningful and specific message rather than making it ambiguous and complex (Further assessment in the future may validate the same).
- The incomprehensible sign(s) in the post-evaluation and redesign phase must undergo a proper sign design/ redesign life cycle for the sign to be comprehensible later. A slight modification might change the sign from the least incomprehensible to the comprehensible one. COSACOEOT, one of the most incomprehensible signs, with a slight modification, became comprehensible and can be used for OSH purposes.

7.1.6.1. Criteria/ conditions for the selection of a particular technique

Table 5. 4 is recommended to refer to understand the technique(s) that can be selected in a particular context/ situation.

- A subject-based technique is better if there is time/ facility to train the subjects and will cost less than other techniques.
- If there is no facility/ less time /no time for training the subjects for the experiment, then data as written text is easy to handle. The expert-based technique can help in this context.
- If in a survey, numerous signs need to be evaluated and designed/ redesigned simultaneously within a limited time, error-free and consistent, and experts are busy with

other important businesses, then the computational technique of evaluation will be helpful.

- An eye-tracking technique can be used if the data source needs to be free of user biases, unobtrusive, and real-time. However, the eye-tracking experiment must be supported by the comprehensibility score from the participants to study the variation of the eye measures across comprehensibility categories (comprehensible/ incomprehensible in this case). This technique is costly considering the eye-tracking experimental setup and time-consuming, which requires proper training of participants as well as the experts who are experimenting.

These are recommendations based on the present study, and the final selection of technique(s) is dependent on the experimenter who is designing the experiment.

Hence from the study, it can be concluded that selecting particular or combined techniques requires proper background research of the organization planning to implement the evaluation paradigm. The research elicits that all the techniques are robust enough with advantages and few disadvantages. The triangulation of the method/ techniques, if applied, will be very useful for organizations depending upon their requirement/ context/ feasibility of survey design/ technical feasibility/ time estimated for evaluation and design of sign(s).

The comparison of the techniques (in Chapter 5) based on parameters will provide new researchers and new OSH specialists/ experts with a preliminary idea about the feasibility of conducting such a study or implementing the technique (s) in their organization. They may first come up with a report of their organization's feasibility study (economic, technical, and behavioral) before selecting any particular type of evaluation technique or a combination of techniques.

7.1.7. Limitations and future scope of the present research

- Data is limited to India and a particular age group. Future studies may include data from different states and countries.
- Only white-collar employees and students could be included in the study. Blue-collar employees could be involved in future studies.
- None of the studies encountered a small age group (children) which can be a significant study as they, too, need to understand and be aware of the OSH signs and symbols. Future studies may include more female participants and small age groups (children).
- The study is about OSH-related signs only. Future work can be studying other signs also.
- The dataset size for the above experimental study is sufficient; however, it is too small for training any deep neural network (DNN) model. Hence this compelled the usage of pre-trained models with similar datasets from other sources like Wikipedia, web news, web question-answer pages, and discussion forums, SNLI corpus besides STS benchmark (English datasets), which are not directly related to OSH but can be used for Natural Language Processing tasks which is important for the computational technique. The

threshold considered for the final comprehensibility is the gold standard selected by the experts, and they are tunable manually. The experts judge the severity of the context and tune the threshold. Identification and recognition of the severity of the context are not automated/ computerized here. The model could not be finetuned at this stage, which can be future work; however, it gave good results when compared with other related models (Table 5.5) and compared with the expert judgment and STS benchmark (Pearson $r > 0.80$).

- The scale used for subjective comprehensibility rating is an 11-point ratio scale. The intention is to capture the whole idea of the user (how much they have comprehended the message from the sign, their ambiguity in understanding, etc.). The research considered binary categories, i.e., comprehensible and incomprehensible at the end. The question comes as to why the researcher collected 11-point scaled data. The reason behind this is that the researcher planned to keep the data as a knowledge base and use it later accordingly for further analysis in the future (by transforming the scale).
- The sample size is limited to forty-one (41) in the eye-tracking experiment, and getting those numbers is indeed a tough job. Getting participants without visual impairment is also a challenge. Sometimes, convincing participants about the experiment being harmless is difficult, though proper ethical procedures and guidelines are maintained. It took some time to acclimate them with the eye-tracking device, and before every experiment, the device was calibrated to capture the experiment correctly. The experimental protocol with different stimuli was shown to every participant before the commencement of each experiment. Getting the same participants again for subjective data is a real challenge for the researcher. The same data could have been taken from participants during the eye-tracking experiment. While watching the signs, they could have spoken what they comprehended from the message. This simultaneous data collection (both eye-tracking data and spoken data) was avoided as speaking during the experiment might distract the user and affect and interfere with capturing eye data. These are a few of the limitations faced during the eye-tracking experiment. Future studies can be designed to minimize the challenges.
- The present research is restricted to studying (four) techniques of comprehensibility evaluation. Many other evaluation techniques can be explored.
- Further studies can be conducted on how the length of text, or the size of pictograms affects comprehensibility. There can be rigorous studies on various factors affecting the comprehensibility of signs. The design elements of the signs can be studied explicitly in terms of how each element affects the comprehensibility of the target users. The detailed discussion is beyond the scope of this thesis.

7.2. Conclusion

This thesis is a novel attempt to study the comprehensibility evaluation techniques of Occupational safety and health (OSH) signs from the users' perspective in each phase of the evaluation and design process. User-centered design (UCD) has been practiced throughout the study keeping in mind that the design and evaluation employed here are directly related to the health and safety of the users. This study is an approach of research through design, and it

is seen that the design is directly linked with evaluation, and nowhere can it be separated in this research. Both research and design are a part of each other, where the outcome is the generated new knowledge that will improve design practices. The evaluation techniques are tools to evaluate the comprehensibility of a design which is a contribution. The redesigned sign became comprehensible, and this enhanced design outcome also contributed to this research. These are contributions to the design process too.

A total of four techniques have been studied here. All the techniques have their advantages and disadvantages. The first three techniques generated subjective data; the fourth generated objective data, and all the data were primary and new. Hence, it contributes to the database, which can be used later for further studies.

The four techniques studied here are excellent individually, evident from the results and validation of those techniques. The study's outcome reveals that triangulation of techniques will also assist better in evaluating the comprehensibility of a sign when there is a need for a stricter evaluation. The success of the techniques is evident from the match of the comprehensible and incomprehensible signs in all the techniques implemented here.

Each technique is studied, and based on its limitations, the following study has been designed, which is believed to eliminate the limitations of the previous techniques. The subject-based technique extracts scores directly from less experienced novices in the scoring procedure. Their data can be inconsistent. Expert-based is the standard procedure for the experiments; however, the experts may also suffer from biases due to previous experiences, notions, overwork, etc., which may affect the survey scores. Hence, to overcome this situation, a computational technique is developed, which assists as a tool to the experts by automatically generating scores from the written responses and deciding whether a sign is comprehensible. This tool or model is validated by the human experts (Pearson correlation >0.80) and their expert judgment. Developing this model is a novel idea, and this may help experts eliminate their involvement in this prolonged and monotonous evaluation procedure. This model has been compared with the related models and found to perform in line with them. This model is in its basic phase and can be upgraded later. This model has only one limitation. It uses the subjective views of the participants. It is impossible to remove subjective biases when the research is about users and their views are essential to the research. The views of the participants remain intact and unbiased when the participants' objective data can be captured. Hence, an eye-tracking model has been developed, which helps capture the unobtrusive data of the participants. These data are raw in nature, and participants' subjective views are necessary to compare or find a correlation between them. So, it is impossible to eliminate entire subjective views in this research. However, the eye-tracking technique helped to find significant variables correlated with the comprehensibility scores or categories. The data collected in all the techniques implemented here have been checked for descriptive statistics and data distribution. Based on that, statistical tests have been conducted to find variances between the signs (Friedman ANOVA followed by post hoc analysis) and whether OSH sign with OSH-related textual information affects the understanding or comprehensibility of the messages intended to convey (Wilcoxon signed rank test followed by calculation of effect size). These tests are based on the comprehensibility scores calculated by the experts (subject-based, expert-based, and eye-

tracking techniques) and computer (computational technique). The bivariate correlation test validates the comprehensibility scores and all the techniques.

In the case of the eye-tracking technique, one extra test has been performed to analyze the variation of objective data across the comprehensible groups. Kruskal Wallis test followed by post hoc analysis has been performed for the same. The correlation plot shows that the eye-tracking measures like blink count, blink frequency, and blink duration; saccade count, saccade amplitude, saccade amplitude average, and saccade duration; scan path length; revisits increase with the increased sign comprehensibility. Average pupil diameter, fixation duration, and saccade latency average decrease with increased comprehensibility. Hence, sign comprehensibility has significant effects on real-time nonintrusive eye movements. These findings demonstrate the potential to integrate physiological measures from eye movements with existing evaluation methods to assess OSH signs' comprehensibility. All three techniques found that DNWSH-T, DNT-T, WAFS, and WAFS-T are common in the comprehensible group. The incomprehensible signs commonly found are COSACOEOT and WRP. The techniques are compared and validated by the correlation study between all the techniques studied in this research.

This research can benefit the EHS/ OSH departments of any organization responsible for OSH in the workplace. The research also contributes to the design process, the design community, and human society. The data or knowledge base gathered can be food for thought for future researchers, and they can be used for further study.



8

References

8. References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., & others. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *ArXiv Preprint ArXiv:1603.04467*.
- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., & others. (2016). Tensorflow: A system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–283.
- Adachi, Y., Ozaki, M., & Iwahori, Y. (2014). Improvement of the Measurement Accuracy and Speed of Pupil Dilation as an Indicator of Comprehension. *Procedia Computer Science*, 35, 1202–1209.
- Ahmadi, M., Mortezaipoor, A., Kalteh, H. O., Emadi, A., Charati, J. Y., & Etemadinezhad, S. (2021). Comprehensibility of pharmaceutical pictograms: Effect of prospective-user factors and cognitive sign design features. *Research in Social and Administrative Pharmacy*, 17(2), 356–361.
- Ahmed, T., Hoque, A. S. M., Karmaker, C. L., & Ahmed, S. (2023). Integrated approach for occupational health and safety (OHS) risk Assessment: An Empirical (Case) study in Small enterprises. *Safety Science*, 164, 106143.
- Ajayi, A., Oyedele, L., Owolabi, H., Akinade, O., Bilal, M., Davila Delgado, J. M., & Akanbi, L. (2020). Deep learning models for health and safety risk prediction in power infrastructure projects. *Risk Analysis*, 40(10), 2019–2039.
- Alami, N., Meknassi, M., & En-nahnahi, N. (2019). Enhancing unsupervised neural networks based text summarization with word embedding and ensemble learning. *Expert Systems with Applications*, 123, 195–211.
- Aljabry, M. A., Alkadi, K., & Awawdeh, M. (2023). The impact of technology on patient knowledge acquisition and retention of oral hygiene instructions and associated risks. *Informatics in Medicine Unlocked*, 36, 101149.
- Anderson, D., Park, J., & others. (2012). An Examination of Test-Retest, Alternate Form Reliability, and Generalizability Theory Study of the easyCBM Reading Assessments: Grade 1. Technical Report# 1216. *Behavioral Research and Teaching*.
- Archer, L. B. (1984). Systematic method for designers. In Cross, N. *Developments in Design Methodology*, 57–82.
- Arcia, A., Grossman, L. V, George, M., Turchioe, M. R., Mangal, S., & Creber, R. M. M. (2019). Modifications to the ISO 9186 Method for Testing Comprehension of Visualizations: Successes and Lessons Learned. *2019 IEEE Workshop on Visual Analytics in Healthcare (VAHC)*, 41–47.
- Avramović, A., Sluga, D., Tabernik, D., Skočaj, D., Stojnić, V., & Ilc, N. (2020). Neural-Network-Based Traffic Sign Detection and Recognition in High-Definition Images Using Region Focusing and Parallelization. *IEEE Access*, 8, 189855–189868. <https://doi.org/10.1109/ACCESS.2020.3031191>

- Babić, D., Dijanić, H., Jakob, L., Babić, D., & Garcia-Garzon, E. (2020). Driver eye movements in relation to unfamiliar traffic signs: an eye tracking study. *Applied Ergonomics*, *89*, 103191.
- Bagagiolo, G., Vigoroso, L., Caffaro, F., Micheletti Cremasco, M., & Cavallo, E. (2019). Conveying safety messages on agricultural machinery: The comprehension of safety pictorials in a group of migrant farmworkers in Italy. *International Journal of Environmental Research and Public Health*, *16*(21), 4180.
- Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *ArXiv Preprint ArXiv:1409.0473*.
- Balali, V., Jahangiri, A., & Machiani, S. G. (2017). Multi-class US traffic signs 3D recognition and localization via image-based point cloud model using color candidate extraction and texture-based recognition. *Advanced Engineering Informatics*, *32*, 263–274.
- Bañares, J. R., Caballes, S. A., Serdan, M. J., Liggayu, A. T., & Bongo, M. F. (2018). A comprehension-based ergonomic redesign of Philippine road warning signs. *International Journal of Industrial Ergonomics*, *65*, 17–25.
- Belouafa, S., Habti, F., Benhar, S., Belafkih, B., Tayane, S., Hamdouch, S., Bennamara, A., & Abourriche, A. (2017). Statistical tools and approaches to validate analytical methods: methodology and practical examples. *International Journal of Metrology and Quality Engineering*, *8*, 9.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828.
- Bentivogli, L., Bernardi, R., Marelli, M., Menini, S., Baroni, M., & Zamparelli, R. (2016). SICK through the SemEval glasses. Lesson learned from the evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment. *Language Resources and Evaluation*, *50*(1), 95–124.
- Berger, P., & Luckmann, T. (2016). The social construction of reality. In *Social Theory Rewired* (pp. 110–122). Routledge.
- Berrio, S., Barrero, L. H., Zambrano, L., & Papadimitriou, E. (2022). Ergonomic factors affecting comprehension levels of traffic signs A Critical Review. *International Journal of Transportation Science and Technology*.
- Blignaut, P., Holmqvist, K., Nyström, M., & Dewhurst, R. (2014). Improving the accuracy of video-based eye tracking in real time through post-calibration regression. *Current Trends in Eye Tracking Research*, 77–100.
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in Public Health*, *6*, 149.
- Bowman, S. R., Angeli, G., Potts, C., & Manning, C. D. (2015). A large annotated corpus for learning natural language inference. *ArXiv Preprint ArXiv:1508.05326*.
- Brainard, R. W. ., Campbell, R. J., & Elkin, E. H. (1961). Design and interpretability of road

- signs. *Journal of Applied Psychology*, 45(2), 130–136.
- Brantley, K. A., & Wogalter, M. S. (1999). Oral and written symbol comprehension testing: The benefit of cognitive interview probing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 43(20), 1060–1064.
- Brewer, N. T., Gilkey, M. B., Lillie, S. E., Hesse, B. W., & Sheridan, S. L. (2012). Tables or bar graphs? Presenting test results in electronic medical records. *Medical Decision Making*, 32(4), 545–553.
- Bureau, N. W. C. S., & Hansen, C. M. (1914). *Universal Safety Standards: A Reference Book of Rules, Drawings, Tables, Formulae, Data Suggestions for Use of Architects, Engineers, Superintendents, Foremen, Inspectors, Mechanics and Students*. Universal Safety Standards Publishing Company.
<https://books.google.co.in/books?id=GE1GAAAAMAAJ>
- Burt, C. D. B., Henningsen, N., & Consedine, N. (1999). Prompting correct lifting posture using signs. *Applied Ergonomics*, 30(4), 353–359.
- Caffaro, F., & Cavallo, E. (2015). Comprehension of safety pictograms affixed to agricultural machinery: A survey of users. *Journal of Safety Research*, 55, 151–158.
- Caire, P., Genon, N., Heymans, P., & Moody, D. L. (2013). Visual notation design 2.0: Towards user comprehensible requirements engineering notations. *2013 21st IEEE International Requirements Engineering Conference (RE)*, 115–124.
- Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183–186.
- Caplette, L., Ince, R. A. A., Jerbi, K., & Gosselin, F. (2020). Disentangling presentation and processing times in the brain. *NeuroImage*, 218, 116994.
- Cer, D., Diab, M., Agirre, E., Lopez-Gazpio, I., & Specia, L. (2017). $\{S\}em\{E\}val-2017$ Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation. *Proceedings of the 11th International Workshop on Semantic Evaluation ($\{S\}em\{E\}val-2017)$* , 1–14. <https://doi.org/10.18653/v1/S17-2001>
- Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Céspedes, M., Yuan, S., Tar, C., & others. (2018). Universal sentence encoder. *ArXiv Preprint ArXiv:1803.11175*.
- Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Céspedes, M., Yuan, S., Tar, C., Strophe, B., & Kurzweil, R. (2018). Universal Sentence Encoder for $\{E\}nglish$. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 169–174.
<https://doi.org/10.18653/v1/D18-2029>
- Chan, A. H. S., & Ng, A. W. Y. (2010). Investigation of guessability of industrial safety signs: Effects of prospective-user factors and cognitive sign features. *International Journal of Industrial Ergonomics*, 40(6), 689–697.
<https://doi.org/10.1016/j.ergon.2010.05.002>
- Chan, K. L., & Chan, A. H. S. (2011). Understanding industrial safety signs: Implications for occupational safety management. *Industrial Management & Data Systems*.

- Chaudhuri, N. B., Dhar, D., & Yammiyavar, P. G. (2021). A human-centred deep learning approach facilitating design pedagogues to frame creative questions. *Neural Computing and Applications*, 1–28.
- Chaudhuri, S. B., Majhi, M., & Karmakar, S. (2021). A scoping review on role of communication media for effective OSH awareness and training. *International Journal of Reliability and Safety*, 15(1–2), 1–17. <https://doi.org/10.1504/IJRS.2021.119641>
- Chaudhuri, S. B., Majhi, M., & Karmakar, S. (2022). Evaluation of Comprehensibility of a Sign by Triangulation Method. In *Ergonomics for Design and Innovation: Humanizing Work and Work Environment: Proceedings of HWWE 2021* (pp. 1811–1822). Springer.
- Chen, D., & Dolan, W. B. (2011). Collecting highly parallel data for paraphrase evaluation. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 190–200.
- Chen, J., Wang, R. Q., Lin, Z., & Guo, X. (2018). Measuring the cognitive loads of construction safety sign designs during selective and sustained attention. *Safety Science*, 105, 9–21.
- Chok, N. S. (2010). *Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data*. University of Pittsburgh.
- Cholewa, N., Wołk, K., & Wołk, R. (2018). Precise eye-tracking technology in medical communicator prototype. *Procedia Computer Science*, 138, 264–271.
- Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons.
- Comelles, E., & Atserias, J. (2019). VERTa: a linguistic approach to automatic machine translation evaluation. *Language Resources and Evaluation*, 53(1), 57–86.
- Comello, M. L. G., Qian, X., Deal, A. M., Ribisl, K. M., Linnan, L. A., & Tate, D. F. (2016). Impact of game-inspired infographics on user engagement and information processing in an eHealth program. *Journal of Medical Internet Research*, 18(9), e237.
- Conneau, A., Kiela, D., Schwenk, H., Barrault, L., & Bordes, A. (2017). Supervised learning of universal sentence representations from natural language inference data. *ArXiv Preprint ArXiv:1705.02364*.
- Corbett, C. L., McLean, G. A., & Cospers, D. K. (2008). *Effective presentation media for passenger safety I: Comprehension of briefing card pictorials and pictograms*.
- Correll, D. J. (2007). The measurement of pain: objectifying the subjective. In *Pain management* (pp. 197–211). Elsevier.
- Cross, N. (2021). *Engineering design methods: strategies for product design*. John Wiley & Sons.
- Crow, D. (2017). *Visible Signs: Intro to Semiotics From David Crow's Visible Signs*. Retrieved November from <http://www.amazon.co.uk>.
- Darbyshire, P., & McDonald, H. (2004). Choosing response scale labels and length: Guidance for researchers and clients. *Australasian Journal of Market Research*, 12(2), 17–26.

- Davies, S., Haines, H., Norris, B., & Wilson, J. R. (1998). Safety pictograms: are they getting the message across? *Applied Ergonomics*, 29(1), 15–23.
- de la Escalera, A., Armingol, J. M., Pastor, J. M., & Rodriguez, F. J. (2004). Visual sign information extraction and identification by deformable models for intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 5(2), 57–68. <https://doi.org/10.1109/TITS.2004.828173>
- Deepthy, M. S., Harichandrakumar, K. T., Parameswaran, S., Kadhiravan, T., & Nair, N. S. (2022). Application of bivariate meta-analytic approach for pooling effect measures of correlated multiple outcomes in medical research. *Clinical Epidemiology and Global Health*, 15, 101029.
- Dewar, R. (1988). Criteria for the design and evaluation of traffic sign symbols. *Transportation Research Record*, 1160(1), 6.
- Ding, N., Ma, Y., Fan, Z., & Shi, J. (2022). Test the Effectiveness of Building Safety Guidance Signs in a T-junction Corridor based on Eye Movement Data. *Journal of Safety Science and Resilience*.
- Dowse, R. (2021). Pharmacists, are words enough? The case for pictograms as a valuable communication tool. *Research in Social and Administrative Pharmacy*, 17(8), 1518–1522.
- Duarte, M. E. C., & Rebelo, F. (2005). Comprehension of safety signs: internal and external variable influences and comprehension difficulties by disabled people. *Proceedings of CybErg, Johannesburg, Zuid-Afrika*.
- Elmogahzy, Y. E. (2019). *Engineering textiles: Integrating the design and manufacture of textile products*. Woodhead Publishing.
- Estevez, L., & Kehtarnavaz, N. (1996). A real-time histographic approach to road sign recognition. *Proceeding of Southwest Symposium on Image Analysis and Interpretation*, 95–100.
- Eygue, M., Richard-Forget, F., Cappelier, J.-M., Pinson-Gadais, L., & Membré, J.-M. (2020). Development of a risk-ranking framework to evaluate simultaneously biological and chemical hazards related to food safety: Application to emerging dietary practices in France. *Food Control*, 115, 107279.
- Fang, C.-Y., Chen, S.-W., & Fuh, C.-S. (2003). Road-sign detection and tracking. *IEEE Transactions on Vehicular Technology*, 52(5), 1329–1341.
- Fang, C.-Y., Fuh, C.-S., Chen, S.-W., & Yen, P.-S. (2003). A road sign recognition system based on dynamic visual model. *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, 1, I–I.
- French, M. J., Gravdahl, J. T., & French, M. J. (1985). *Conceptual design for engineers*. Springer.
- Galer, M. (1980). An ergonomics approach to the problem of high vehicles striking low bridges. *Applied Ergonomics*, 11(1), 43–46.
- Gaver, W. (2012). What should we expect from research through design? *Proceedings of the*

SIGCHI Conference on Human Factors in Computing Systems, 937–946.

- Gaver, W. W., Beaver, J., & Benford, S. (2003). Ambiguity as a resource for design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 233–240.
- Gilula, Z., McCulloch, R. E., Ritov, Y., & Urminsky, O. (2019). A study into mechanisms of attitudinal scale conversion: A randomized stochastic ordering approach. *Quantitative Marketing and Economics*, 17, 325–357.
- Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5–6), 602–610.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Handcock, H. E., Rogers, W. A., Schroeder, D., & Fisk, A. D. (2004). Safety symbol comprehension: Effects of symbol type, familiarity, and age. *Human Factors*, 46(2), 183–195.
- Herriott, R. (2019). What kind of research is research through design. *IASDR 2019*.
- Hinton, G. E., & others. (1986). *Proceedings of the eighth annual conference of the cognitive science society*.
- Hogg, R. V, Tanis, E. A., & Zimmerman, D. L. (1977). *Probability and statistical inference* (Vol. 993). Macmillan New York.
- Holmqvist, K., Örbom, S. L., Hooge, I. T. C., Niehorster, D. C., Alexander, R. G., Andersson, R., Benjamins, J. S., Blignaut, P., Brouwer, A.-M., Chuang, L. L., & others. (2023). Eye tracking: empirical foundations for a minimal reporting guideline. *Behavior Research Methods*, 55(1), 364–416.
- Hsu, S.-H., & Huang, C.-L. (2001). Road sign detection and recognition using matching pursuit method. *Image and Vision Computing*, 19(3), 119–129.
- Hsu, Y.-L., Li, W.-C., & Tang, C.-H. (2013). The use of eye tracking in the study of airline cabin safety communication. *International Conference on Engineering Psychology and Cognitive Ergonomics*, 134–143.
- ISO-7010. (2003). Standard Graphical Symbols: Safety Colours and Safety Signs--Registered Safety Signs (ISO 7010: 2003). 2003. *International Standards Organisation (ISO): Geneva, Switzerland*.
- ISO. (2008). ISO 9186-2:2008: Graphical symbols—Test methods—Part 2: Method for testing perceptual quality. *Geneva: International Organization of Standardization*. [Google Scholar], 2008, 13.
- ISO. (2014a). 9186--1: 2014 Graphical Symbols--Test Methods--Part 1: Method for Testing Comprehensibility. *International Organization for Standardization, Geneva*.
- ISO. (2014b). ISO 9186-3, Graphical symbols--Test methods--Part 3: Method for testing symbol referent association. *Geneva: International Organization of Standardization*. [Google Scholar], 14.

- ISO 3864-1, 2011. (2011). *Graphical symbols--Safety colours and safety signs--Part 1: Design principles for safety signs and safety markings.*
- Iyyer, M., Manjunatha, V., Boyd-Graber, J., & Daumé III, H. (2015). Deep unordered composition rivals syntactic methods for text classification. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1681–1691.
- Jahedi, S., & Méndez, F. (2014). On the advantages and disadvantages of subjective measures. *Journal of Economic Behavior and Organization*, 98. <https://doi.org/10.1016/j.jebo.2013.12.016>
- Jakobson, R. (2010). *Word and language* (Vol. 2). Walter de Gruyter.
- Janda, H. F., Volk, W. N., & Lloyd, M. G. (1935). Effectiveness of various highway signs. *Highway Research Board Proceedings*, 14.
- Jiang, W., Zhou, J., Su, H., & Wu, Z. (2022). The design of experimental courses in safety culture. *Heliyon*, 8(12), e11915.
- Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Citeseer.
- Kaufmann, D., & Ramirez-Andreotta, M. D. (2020). Communicating the environmental health risk assessment process: Formative evaluation and increasing comprehension through visual design. *Journal of Risk Research*, 23(9), 1177–1194.
- Kehtarnavaz, N., Griswold, N. C., & Kang, D. S. (1993). Stop-sign recognition based on color/shape processing. *Machine Vision and Applications*, 6(4), 206–208.
- Kibble, R. (2013). Introduction to natural language processing. *University of London*.
- Kim. (2013). Statistical notes for clinical researchers: Evaluation of measurement error 1: using intraclass correlation coefficients. *Restorative Dentistry & Endodontics*, 38(2), 98–102.
- Kim, H., Tamayo, D., Muhkin, M., Kim, J., Lam, J., Ohno-Machado, L., & Aronoff-Spencer, E. (2012). Pictogram evaluation and authoring collaboration environment. *NI 2012: 11th International Congress on Nursing Informatics, June 23-27, 2012, Montreal, Canada., 2012*.
- Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, {EMNLP} 2014, October 25-29, 2014, Doha, Qatar, {A} Meeting of SIGDAT, a Special Interest Group of the {ACL}*, 1746–1751. <http://aclweb.org/anthology/D/D14/D14-1181.pdf>
- Kirmizioglu, E., & Tuydes-Yaman, H. (2012). Comprehensibility of traffic signs among urban drivers in Turkey. *Accident Analysis & Prevention*, 45, 131–141.
- Krieber, M., Bartl-Pokorny, K. D., Pokorny, F. B., Einspieler, C., Langmann, A., Körner, C., Falck-Ytter, T., & Marschik, P. B. (2016). The relation between reading skills and eye movement patterns in adolescent readers: evidence from a regular orthography. *PLoS One*, 11(1), e0145934.

- Kummetha, V. C., Kondyli, A., & Devos, H. (2021). Evaluating driver comprehension of the roadway environment to retain accountability of safety during driving automation. *Transportation Research Part F: Traffic Psychology and Behaviour*, *81*, 457–471.
- Kwong, W. J., & Pathak, D. S. (2007). Validation of the eleven-point pain scale in the measurement of migraine headache pain. *Cephalalgia*, *27*(4), 336–342.
- la Escalera, A., Armingol, J. M., & Mata, M. (2003). Traffic sign recognition and analysis for intelligent vehicles. *Image and Vision Computing*, *21*(3), 247–258.
- La Fata, C. M., Giallanza, A., Micale, R., & La Scalia, G. (2021). Ranking of occupational health and safety risks by a multi-criteria perspective: inclusion of human factors and application of VIKOR. *Safety Science*, *138*, 105234.
- Lagopoulos, A. P., & Boklund-Lagopoulou, K. (2021). Chapter 2 A brief history of semiotics. In *The Tradition of Ferdinand de Saussure* (pp. 8–22). De Gruyter Mouton. <https://doi.org/doi:10.1515/9783110618808-002>
- Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, *10*, 90–115.
- Lazaridou, A., Elbaridi, N., Edwards, R. R., & Berde, C. B. (2018). Pain assessment. In *Essentials of pain medicine* (pp. 39–46). Elsevier.
- Lehto, M. R. (1992). Designing warning signs and warning labels: part II—scientific basis for initial guidelines. *International Journal of Industrial Ergonomics*, *10*(1–2), 115–138.
- Leiner, M., Peinado, J., Baylon, A., Lopez, I., & Pathak, I. (2018). Divide and conquer: improving parental understanding of health-related instructions using sequential pictorial instructions. *Health Education Research*, *33*(2), 104–113.
- Lerner, N. D., & Collins, B. L. (1980). *The assessment of safety symbol understandability by different testing methods*. US National Bureau of Standards.
- Li, G., Yang, Y., Li, S., Qu, X., Lyu, N., & Li, S. E. (2021). Decision making of autonomous vehicles in lane change scenarios: Deep reinforcement learning approaches with risk awareness. *Transportation Research Part C: Emerging Technologies*, 103452.
- Lin, S.-Y., Thompson, H. J., Hart, L. A., Fu, M. C. C., & Demiris, G. (2021). Evaluation of pharmaceutical pictograms by older “turkers”: a cross-sectional crowdsourced study. *Research in Social and Administrative Pharmacy*, *17*(6), 1079–1090.
- Liu, C., Kemper, S., & McDowd, J. (2009). The use of illustration to improve older adults’ comprehension of health-related information: Is it helpful? *Patient Education and Counseling*, *76*(2), 283–288.
- Liu, L., Hoelscher, U., & Gruchmann, T. (2005). Symbol comprehension in different countries: experience gained from medical device area. *Mensch & Computer 2005. Workshop-Proceedings Der 5. Fachübergreifenden Konferenz*.
- Liu, S., Yang, N., Li, M., & Zhou, M. (2014). *A recursive recurrent neural network for statistical machine translation*.

- Luo, H., Yang, Y., Tong, B., Wu, F., & Fan, B. (2018). Traffic Sign Recognition Using a Multi-Task Convolutional Neural Network. *IEEE Transactions on Intelligent Transportation Systems*, 19(4), 1100–1111. <https://doi.org/10.1109/TITS.2017.2714691>
- Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *ArXiv Preprint ArXiv:1508.04025*.
- Mainkar, V. V., Katkar, J. A., Upade, A. B., & Pednekar, P. R. (2020). Handwritten Character Recognition to Obtain Editable Text. *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 599–602. <https://doi.org/10.1109/ICESC48915.2020.9155786>
- Malhotra, R., Tan, Y. W., Suppiah, S. D., Tay, S. S. C., Tan, N. C., Liu, J., Koh, G. C.-H., Chan, A., Vaillancourt, R., Group, P. S., & others. (2022). Pharmaceutical pictograms: User-centred redesign, selection and validation. *PEC Innovation*, 100116.
- Mansoor, L. E., & Dowse, R. (2004). Design and evaluation of a new pharmaceutical pictogram sequence to convey medicine usage. *Ergonomics SA*, 16(2), 29–41.
- Martin, B., Hanington, B., & Hanington, B. M. (2012). *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport Pub.
- Maulina, D., Siregar, E. S., Rachma, T. A., Nashria, S. A., & Irwanda, D. Y. (2022). How effective is training for improving traffic sign comprehension? Examining the interaction between training and sign type among motorcyclists. *IATSS Research*, 46(4), 614–622.
- McDougald, B. R., & Wogalter, M. S. (2014). Facilitating pictorial comprehension with color highlighting. *Applied Ergonomics*, 45(5), 1285–1290.
- McDowell, M., Gigerenzer, G., Wegwarth, O., & Rebitschek, F. G. (2019). Effect of tabular and icon fact box formats on comprehension of benefits and harms of prostate cancer screening: a randomized trial. *Medical Decision Making*, 39(1), 41–56.
- Meek, S. R., & Tietz, M. A. (2022). Entrepreneurship and subjective vs objective institutional performance: A decade of US hospital data. *Research Policy*, 51(9), 104652.
- Megalingam, R. K., Thanigundala, K., Musani, S. R., Nidamanuru, H., & Gadde, L. (2022). Indian traffic sign detection and recognition using deep learning. *International Journal of Transportation Science and Technology*.
- Montagne, M. (2013). Pharmaceutical pictograms: a model for development and testing for comprehension and utility. *Research in Social and Administrative Pharmacy*, 9(5), 609–620.
- Morina, V. M., & Ahma, G. M. (2022). Building and Comparing AI-Powered Algorithms in Road Sign Detection. *IFAC-PapersOnLine*, 55(39), 404–407. <https://doi.org/https://doi.org/10.1016/j.ifacol.2022.12.066>
- Mugge, R., & Dahl, D. W. (2013). Seeking the ideal level of design newness: Consumer response to radical and incremental product design. *Journal of Product Innovation Management*, 30, 34–47.

- Ng, A. W. Y., & Chan, A. H. S. (2008). The effects of driver factors and sign design features on the comprehensibility of traffic signs. *Journal of Safety Research*, 39(3), 321–328.
- Ng, A. W. Y., & Chan, A. H. S. (2015). Effects of user factors and sign referent characteristics in participatory construction safety sign redesign. *Safety Science*, 74, 44–54.
- Ng, A. W. Y., Lo, H. W., & Chan, A. H. (2011). Measuring the Usability of Safety Signs: A use of system usability scale (SUS). *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2, 1296–1301.
- Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T. C., Kasneci, E., & Nyström, M. (2020). The impact of slippage on the data quality of head-worn eye trackers. *Behavior Research Methods*, 52, 1140–1160.
- Panchal, B. Y., Chauhan, G., Panchal, S. R., & Chaudhari, U. M. (2022). An investigation on feature and text extraction from images using image recognition in Android. *Materials Today: Proceedings*, 51, 798–802.
<https://doi.org/https://doi.org/10.1016/j.matpr.2021.06.237>
- Peckham, G. (2017). Safety Signs as Risk Communication. *Professional Safety*, 62(1), 26–27.
- Pedgley, O. (2007). Capturing and analysing own design activity. *Design Studies*, 28(5), 463–483.
- Pham, H. T. T. L., Rafieizonooz, M., Han, S., & Lee, D.-E. (2021). Current Status and Future Directions of Deep Learning Applications for Safety Management in Construction. *Sustainability*, 13(24), 13579.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879.
- Polančič, G., Brin, P., Heloisa Thom, L., Sosa, E., & Kocbek Bule, M. (2020). An empirical investigation of the intuitiveness of process landscape designs. *Enterprise, Business-Process and Information Systems Modeling: 21st International Conference, BPMDS 2020, 25th International Conference, EMMSAD 2020, Held at CAiSE 2020, Grenoble, France, June 8--9, 2020, Proceedings 21*, 209–223.
- Price, P. C., Jhangiani, R. S., Chiang, I. A., Leighton, D. C., & Cuttler, C. (2017). *Research Methods in Psychology (3rd American Edition)*. Pullman, WA: Publisher.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*.
- Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). Squad: 100,000+ questions for machine comprehension of text. *ArXiv Preprint ArXiv:1606.05250*.
- Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: a randomized trial. *Pain*, 104(1–2), 187–194.
- Reis, L., Duarte, E., & Rebelo, F. (2015). Research on workplace safety sign compliance: validation of a virtual environment prototype. *Procedia Manufacturing*, 3, 6599–6606.

- Ruiz, J. G., Andrade, A. D., Garcia-Retamero, R., Anam, R., Rodriguez, R., & Sharit, J. (2013). Communicating global cardiovascular risk: Are icon arrays better than numerical estimates in improving understanding, recall and perception of risk? *Patient Education and Counseling*, 93(3), 394–402.
- Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, 71–78.
- Sandoval, H., Hattori, T., Kitagawa, S., & Chigusa, Y. (2000). Angle-dependent edge detection for traffic signs recognition. *Proceedings of the IEEE Intelligent Vehicles Symposium 2000 (Cat. No. 00TH8511)*, 308–313.
- Sattler, B., Lippy, B., & Jordan, T. (1997). *Hazard communication: A review of the science underpinning the art of communication for health and safety*.
- Scharinger, C., Schüler, A., & Gerjets, P. (2020). Using eye-tracking and EEG to study the mental processing demands during learning of text-picture combinations. *International Journal of Psychophysiology*, 158, 201–214.
- Schiekel, C. (1999). A fast traffic sign recognition algorithm for gray value images. *International Conference on Computer Analysis of Images and Patterns*, 588–595.
- Schillaci, M. A., & Schillaci, M. E. (2022). Estimating the population variance, standard deviation, and coefficient of variation: Sample size and accuracy. *Journal of Human Evolution*, 171, 103230.
- Shadeed, W. G., Abu-Al-Nadi, D. I., & Mismar, M. J. (2003). Road traffic sign detection in color images. *10th IEEE International Conference on Electronics, Circuits and Systems, 2003. ICECS 2003. Proceedings of the 2003*, 2, 890–893.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591–611.
- Shieh, K.-K., & Lee, D.-S. (2007). Preferred viewing distance and screen angle of electronic paper displays. *Applied Ergonomics*, 38(5), 601–608.
- Shinyama, Y., & Sekine, S. (2006). Preemptive information extraction using unrestricted relation discovery. *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, 304–311.
- Siswandari, Y., Kim, W., & Xiong, S. (2014). Comprehension of newly introduced water-sport prohibitive signs in Korea by westerners. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 2300–2304.
- Siswandari, Y., & Xiong, S. (2015). Eye movements and brain oscillations to symbolic safety signs with different comprehensibility. *Journal of Physiological Anthropology*, 34(1), 1–12.
- Skansi, S. (2018). *Introduction to Deep Learning: from logical calculus to artificial intelligence*. Springer.
- Snyder, C. F., Smith, K. C., Bantug, E. T., Tolbert, E. E., Blackford, A. L., Brundage, M. D., & Board, P. R. O. D. P. S. A. (2017). What do these scores mean? Presenting patient-

- reported outcomes data to patients and clinicians to improve interpretability. *Cancer*, 123(10), 1848–1859.
- Soares, M. A. C., & Parreiras, F. S. (2020). A literature review on question answering techniques, paradigms and systems. *Journal of King Saud University-Computer and Information Sciences*, 32(6), 635–646.
- Socher, R., Huang, E. H., Pennington, J., Ng, A. Y., & Manning, C. D. (2011). Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. *NIPS*, 24, 801–809.
- Sung, S. H., Li, C., Chen, G., Huang, X., Xie, C., Massicotte, J., & Shen, J. (2021). How does augmented observation facilitate multimodal representational thinking? Applying deep learning to decode complex student construct. *Journal of Science Education and Technology*, 30(2), 210–226.
- Tam, C. M., Fung, I. W. H., Yeung, T. C. L., & Tung, K. C. F. (2003). Relationship between construction safety signs and symbols recognition and characteristics of construction personnel. *Construction Management and Economics*, 21(7), 745–753.
- Tao, D., Yuan, J., & Qu, X. (2018). Presenting self-monitoring test results for consumers: the effects of graphical formats and age. *Journal of the American Medical Informatics Association*, 25(8), 1036–1046.
- Tekinbas, K. S., & Zimmerman, E. (2003). *Rules of play: Game design fundamentals*. MIT press.
- Tripathy, J. K., Sethuraman, S. C., Cruz, M. V., Namburu, A., Mangalraj, P., Vijayakumar, V., & others. (2021). Comprehensive analysis of embeddings and pre-training in NLP. *Computer Science Review*, 42, 100433.
- van Beusekom, M. M., Land-Zandstra, A. M., Bos, M. J. W., van den Broek, J. M., & Guchelaar, H.-J. (2017). Pharmaceutical pictograms for low-literate patients: understanding, risk of false confidence, and evidence-based design strategies. *Patient Education and Counseling*, 100(5), 966–973.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 5998–6008.
- Vigoroso, L., Caffaro, F., & Cavallo, E. (2020). Occupational safety and visual communication: User-centred design of safety training material for migrant farmworkers in Italy. *Safety Science*, 121, 562–572.
- Vitabile, S., Gentile, A., & Sorbello, F. (2002). A neural network based automatic road signs recognizer. *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, 3, 2315–2320.
- Vukelich, M., & Whitaker, L. A. (1993). The effects of context on the comprehension of graphic symbols. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 37(8), 511–515.
- Wang, Y., & Gafurov, D. (2003). The cognitive process of comprehension. *The Second IEEE International Conference on Cognitive Informatics, 2003. Proceedings.*, 93–97.

- Waugh, P. (2006). *Literary theory and criticism: An Oxford guide*. Oxford University Press on Demand.
- Wogalter, M. S. (2018a). Communication-human information processing (C-HIP) model. In *Forensic Human Factors and Ergonomics* (pp. 33–49). CRC Press.
- Wogalter, M. S. (2018b). Hazard Analysis and Hazard-Control Hierarchy. In *Forensic Human Factors and Ergonomics* (pp. 17–32). CRC Press.
- Wogalter, M. S., & Laughery, K. R. (1996). Warning! Sign and label effectiveness. *Current Directions in Psychological Science*, 5(2), 33–37.
- Wogalter, M. S., & Sojourner, R. J. (1999). Research on pharmaceutical labeling: an information processing approach. In *Processing of medical information in aging patients* (pp. 329–350). Psychology Press.
- Xiang, L., Cai, M., Ren, C., & Ng, E. (2021). Modeling pedestrian emotion in high-density cities using visual exposure and machine learning: Tracking real-time physiology and psychology in Hong Kong. *Building and Environment*, 205, 108273.
- Xu, W., Callison-Burch, C., & Dolan, W. B. (2015). Semeval-2015 task 1: Paraphrase and semantic similarity in twitter (pit). *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 1–11.
- Xu, X., Jin, J., Zhang, S., Zhang, L., Pu, S., & Chen, Z. (2019). Smart data driven traffic sign detection method based on adaptive color threshold and shape symmetry. *Future Generation Computer Systems*, 94, 381–391.
<https://doi.org/https://doi.org/10.1016/j.future.2018.11.027>
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, 32.
- Yeow, J. A., Ng, P. K., Tan, K. S., Chin, T. S., & Lim, W. Y. (2014). Effects of stress, repetition, fatigue and work environment on human error in manufacturing industries. *Journal of Applied Sciences*, 14(24), 3464–3471.
- Yin, H. S., Parker, R. M., Sanders, L. M., Mendelsohn, A., Dreyer, B. P., Bailey, S. C., Patel, D. A., Jimenez, J. J., Kim, K.-Y. A., Jacobson, K., & others. (2017). Pictograms, units and dosing tools, and parent medication errors: a randomized study. *Pediatrics*, 140(1).
- Zaini, W. M. F., Lai, D. T. C., & Lim, R. C. (2022). Identifying patent classification codes associated with specific search keywords using machine learning. *World Patent Information*, 71, 102153.
- Zashchirinskaia, O. V. (2020). Patterns of interrelation between perception and understanding of images and texts with different degree of blur. *Attention, Perception, & Psychophysics*, 82(3), 1368–1377.
- Zeman, J. (1977). Peirce's theory of signs. *A Perfusion of Signs*, 22–39.
- Zhang, D., Chen, X., Li, F., Sangaiah, A. K., & Ding, X. (2020). Seam-carved image tampering detection based on the cooccurrence of adjacent lbps. *Security and Communication Networks*, 2020.

- Zhou, M., Duan, N., Liu, S., & Shum, H.-Y. (2020). Progress in neural NLP: modeling, learning, and reasoning. *Engineering*, 6(3), 275–290.
- Zhu, Z., Hu, Z., Dai, W., Chen, H., & Lv, Z. (2022). Deep learning for autonomous vehicle and pedestrian interaction safety. *Safety Science*, 145, 105479.
- Zimmerman, J., & Forlizzi, J. (2014). Research Through Design in HCI. In J. S. Olson & W. A. Kellogg (Eds.), *Ways of Knowing in HCI* (pp. 167–189). Springer New York. https://doi.org/10.1007/978-1-4939-0378-8_8
- Zimmerman, J., Stolterman, E., & Forlizzi, J. (2010). An analysis and critique of Research through Design: towards a formalization of a research approach. *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, 310–319.
- Zimmerman, M., & others. (1989). *Developing health and family planning print materials for low-literate audiences: A guide.*





Appendices

Appendix A: Eye tracking Set up (SMI ETG2 Wireless Analysis Pro)

Model: SMI Eye Tracking Glasses (Figure A. 1 and Figure A. 2)

1. Sample rate: 50/ 60 Hz (50/ 60 times per second does the eye tracker measure the eye position).
2. Accuracy: $<0.5^\circ$ (offset from the true gaze points)
3. Gaze tracking range: 80° horizontal, 60° vertical
4. Eye tracking technique: dark pupil
5. Binocular eye tracking: yes
6. Wavelength: 910 nm.

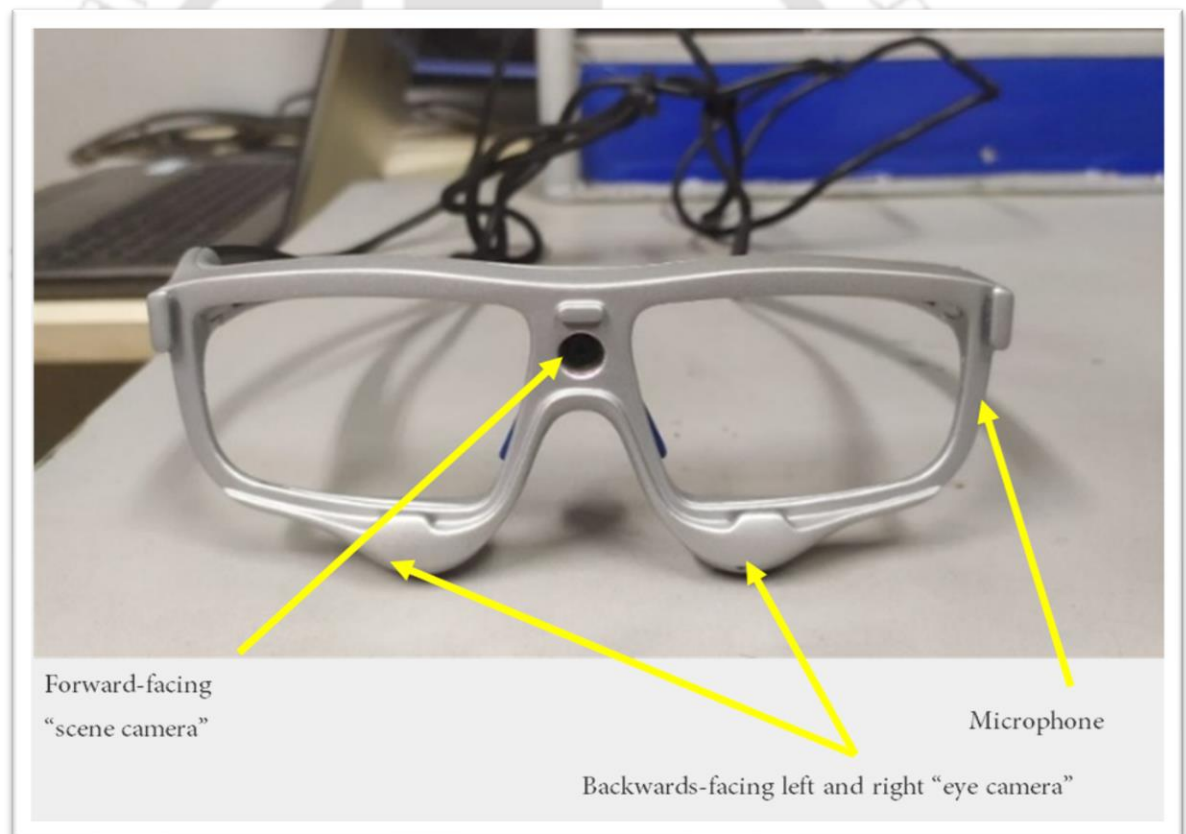


Figure A. 1. SMI Eye Tracking Glass

SMI ETG 2w Smart Recorder

1. Dimensions: 135x 69 x 23 mm (length x width x height)
2. Weight: 246g

3. Options Wireless control via Wi-Fi-connected Windows device.
4. Recording time: 3hrs+ without battery exchange
5. Storage capacity: 10 hrs recording

iView ETG

1. Allows to perform a task on the laptop and create and modify experiments- experiment settings and participant's properties or annotations.
2. It also allows working experiments on the smart recorder or with BeGaze.
3. Scene Camera mode: 1280*960, 24 FPS (standard frame rate- visual cues moving at a certain rate)
4. Calibration: 1 point

Model: Laptop

1. Laptop display: Resolution 1366*768
2. Storage: 913 GB
3. Windows 10: Home Single Language.
4. Processor: Intel (R) Core (TM) i5-6200 U CPU @ 2.30 GHz.
5. Memory (RAM): 4 GB.
6. Operating System: 64 bits.
7. Processor: X64-based processor.

Model: Plasma display

1. Plasma display (42.3" measured diagonally)
2. Make: Samsung
3. Model: PS-42 A 410C1



Figure A. 2. Eye experimental setup- big screen and laptop

Appendix B: Subject-based data

Subject-based data is presented in Table B. 1.

Table B. 1. Subject-based data for 14 signs and 137 participants

Subject-based data														
.No	NAFUP	UBC	DNT	DNWSH	COSACOEOT	WRP	WAFS	DNWSH-T	NAFUP-T	WAFS-T	COSACOEOT-T	UBC-T	WRP-T	DNT-T
1.	0.7	0.4	0.1	0.7	0.1	0.5	0.8	0.9	1	0.8	1	1	0.6	1
2.	1	0.4	0.5	0.5	0	1	1	0.4	0.9	1	0.4	1	1	1
3.	0.8	0.4	0.9	1	0.4	0.8	0.9	1	0.8	1	1	0.9	0.8	0.9
4.	0.8	0.4	0.8	0.3	0	0.6	0.8	1	0.7	0.7	1	0.7	0.5	0.8
5.	1	1	0.7	0.4	0.5	0.3	1	1	1	1	0.2	0	0.5	1
6.	0.8	0.2	0.7	0.8	0.3	0.2	0.8	0.8	0.7	0.9	0.5	0.8	0.5	0.9
7.	0.8	0.4	0.6	0	0.4	0.5	0.5	0.2	0	0.6	0.5	0.6	0.7	0.8
8.	0.3	0	1	0	0	0.2	0.5	0.5	0.5	1	0.3	0	0.9	0.7
9.	0.8	0.3	0.6	0.5	0.5	0.2	1	0.8	0.5	1	0.2	0.5	1	0.8
10.	0.9	1	0.8	1	0	0.8	0	1	0.9	1	1	1	1	1
11.	0.1	0.5	0.2	0	0	0.5	0.7	0	0.7	0.8	0	0.4	0.4	0.5
12.	0.3	0.4	0.6	1	0.3	0	0.5	0.8	1	1	1	1	1	1

13.	0.4	0.3	0.5	0.9	0.5	0.5	0.9	1	0.9	1	1	0.5	0.1	0.9
14.	0.4	0.3	0.6	0.7	0.2	0.4	0.5	0.5	0	0.7	0.2	0.5	0.4	0.7
15.	0.9	0.3	0.9	0.2	0.3	0.1	0.5	1	1	1	0	0.2	0	0.8
16.	0.2	0.4	0.8	0.8	0	0.7	0.9	1	0.9	1	1	1	1	1
17.	0.7	0.4	0.7	0	0	0	0.9	1	1	1	0.5	0.8	0	1
18.	1	0.4	1	0.9	0.8	0.3	1	1	1	1	0.2	1	1	1
19.	1	0.4	0.1	1	0.1	0.1	0.1	1	0.1	1	0.1	0.4	1	1
20.	0.7	0.4	0.5	0.5	0.1	0	1	1	0.3	1	0.1	0.2	0.1	1
21.	1	0.4	0	1	0	0	1	1	0.1	1	0.5	0.4	0	0
22.	0.9	0.4	0.5	0.4	0	0.2	0.9	0.5	0.8	0.9	0.9	0.7	0.8	0.8
23.	0.1	0.5	0.5	0.2	0.2	0.1	0.2	0.2	0	0.2	0.2	0.1	0.1	0.2
24.	0.8	0.9	0.9	1	0.9	1	1	0.8	0.8	1	0.8	0.9	1	1
25.	0.8	0.5	1	0.8	1	0.5	1	0.8	0.4	1	0	1	0.4	0.3
26.	1	1	1	1	1	1	1	1	1	1	0.1	0.1	0.1	0.1
27.	0.3	0.4	0	0.5	0	0.5	0.7	0.8	0.7	0.8	0.7	0.7	0.7	0.7
28.	0.8	0.5	0.8	0.9	0.9	0	0.6	0.5	1	0.4	0.3	0	0	0.1
29.	0.8	0.6	0.9	1	0	0.4	0.9	0.9	0.8	0.9	0.9	0.4	0.9	0.9
30.	0.6	0.4	0.2	0.3	0.4	0.5	0.1	0.3	0	0.6	0.8	0	0.2	0.1
31.	0.5	0.2	0.6	0.6	0	0.4	0.5	0.8	1	1	0.6	0.7	0.6	0.8
32.	1	1	1	1	1	1	1	1	1	1	1	1	1	1
33.	0.8	0.9	1	1	1	0	1	1	0.1	1	0.1	0.1	0.1	1

34.	0.7	0.3	0.9	0.5	0.9	0	0.8	0.8	0.9	0.5	0.4	0.3	0.1	0.5
35.	1	0.3	1	0.1	0.5	0.5	1	1	1	1	1	1	1	1
36.	0.6	0.3	0	1	0	0	0.5	1	0	0.8	0.2	0	0.6	0.4
37.	0.9	0.9	0.9	0.9	0	0	0	0.9	0.9	0.9	0.9	0.9	0.9	0.9
38.	1	1	0.6	1	0.8	0.6	1	1	0.1	1	1	0.4	0.8	1
39.	1	1	1	1	0.5	0.1	1	1	1	1	0.6	1	0.8	0.8
40.	0.8	0.4	0.5	0.8	0.3	0.1	0.7	0.4	0.6	0.8	0.6	0.3	0.3	0.8
41.	1	1	1	1	0	0	1	0.2	0.1	0.1	0.4	0	0	1
42.	0	1	0	1	1	1	1	1	0.1	1	0	0	1	0.1
43.	1	0.7	0.5	0.1	0.3	0.2	0.6	1	0.5	1	0.1	1	1	1
44.	1	0.9	0.8	0.3	0.5	0.4	0.7	0.6	0.6	0.8	0.6	0.7	0.3	0.9
45.	1	0.7	0.8	0.4	0	0.2	0.5	0.2	0	0.5	0.3	0.3	0	0.4
46.	0.6	1	0.5	0.2	0.8	0.4	1	1	0.6	1	0.5	0.4	0.2	0.6
47.	1	1	0	1	0	1	1	1	0.1	1	0.1	0.3	1	0.1
48.	0.8	0.5	0.7	0.8	0	0.6	0.5	0.9	0.3	0.9	0.1	0.5	0.7	0.7
49.	0.8	0.4	0.7	0.8	0	0.5	0.5	0.9	0.8	0.9	0.1	0.5	0.7	0.7
50.	1	0	0.5	0.8	0.2	0.7	0.6	1	1	0.7	0.1	0	0.5	1
51.	1	0.2	0.9	0.9	0.8	0.2	0.9	0.9	1	0.9	0.8	0.9	0.9	0.9
52.	0.9	0.3	0.5	0.8	0.1	0.9	1	1	0.4	1	0	0.5	0.9	1
53.	0.6	0.2	0.4	1	0.1	0.1	0.6	1	0.9	1	0.8	0.7	1	1
54.	0.8	0.3	0.7	0.9	0.4	0.4	0.8	0.9	0.8	0.8	0.7	0.2	0.7	0.9

55.	0.3	1	1	1	0	1	1	1	1	1	0.3	1	0.1	1
56.	1	1	1	0.1	0.5	0.2	1	1	0.2	1	0.6	0.6	0.4	1
57.	0.8	0	0.8	0.8	0.8	0.3	0.8	0.4	0.5	0.2	0.2	0.8	0.1	0.8
58.	0.6	0.5	0	0.4	0	0	0.5	0.8	0.3	0.6	0	0.5	0	0.6
59.	0.4	1	0.5	0.5	0	0.8	0.5	0.5	0.5	0.8	0.3	0.5	0.8	0.5
60.	0.5	0	0	0.8	0.7	0.8	0.8	0	0	0.8	0	0	0.5	0
61.	0.9	0.8	0	0.7	0	0	0.5	0.5	1	1	1	1	0.9	0
62.	0.8	0.5	0.8	0.8	0.8	0	0.8	0.8	0.8	0.9	0.8	0.8	0	0.9
63.	0.4	0.9	0	1	0.2	0.1	1	1	1	1	0.1	1	1	
64.	0.1	1	0	1	0.3	0.1	0.6	1	0.4	0.7	0.1	0.4	0.3	0.2
65.	0.7	1	0.9	0.7	0	1	1	0.2	0.8	1	0.6	0.8	1	1
66.	0.9	0.4	0.3	0.2	0.1	0.6	0.6	1	0.7	1	0	0.7	0.9	0.2
67.	0.9	1	0.8	0.9	0	0.3	1	0.2	0.8	1	0	0.9	0.1	0.9
68.	0.9	1	1	1	0	0.1	1	1	1	0.5	0.2	0.6	1	1
69.	0.1	0.9	0.9	0.9	0.8	0.8	1	1	1	1	1	1	0.5	1
70.	0.8	0	0.8	1	0	0	0.5	1	1	0.8	0	0	0	1
71.	0.5	1	0.5	0.5	0.1	0.1	1	1	1	1	1	1	0.4	0.5
72.	0.4	0.6	0.3	0.7	0	0.4	0.6	1	0.8	1	0	0.6	0.9	0.6
73.	0.7	0.7	0.4	0.8	0.1	0	0.6	0.8	0.7	0.8	0.7	0.8	0.7	0.9
74.	0.5	1	0.8	0	0	0.5	0.8	0.7	0.6	1	0.4	1	0.8	0.5
75.	0.5	0.8	0.6	0.9	0.2	0.3	1	1	0.8	1	0.2	0.7	0.5	0.8

76.	1	0.4	0	0.2	0.1	0.2	0.4	0.8	0.6	0.7	1	1	0.4	0.5
77.	0.4	0.8	0.6	1	0	1	1	1	1	1	0	1	1	0.9
78.	0.6	0.7	0.5	0.8	0	1	1	1	1	1	0	0.4	1	0.9
79.	0.5	0.6	0.5	0.7	0.5	0.8	0.9	1	0.5	1	0.7	0.6	1	1
80.	0.2	1	0.8	1	0.5	0.5	0.5	0.9	0.1	1	1	0.9	0.9	0.9
81.	0.5	1	0.8	1	0	0.5	1	1	0.5	1	0	0.5	0.5	1
82.	0.3	0.8	1	0.7	0.7	0.2	0.8	0.2	0.4	0.8	0.4	0.3	0.8	0.8
83.	0	0	0	1	0	1	1	1	0	1	0	0	0.1	0
84.	0.7	0.2	0.5	0.2	0.6	0.3	0.7	0.8	0.5	0.7	0.8	0.8	0.4	0.5
85.	1	1	0	1	0	0.4	1	0	0.5	1	0.5	0.5	0.5	1
86.	0	1	1	0	0	0	0.8	0	0.5	1	0.5	1	0	1
87.	0.8	1	0.5	0.5	0.4	1	1	0.5	0.5	1	0.5	0.5	1	0.3
88.	1	1	0.5	1	0	1	1	1	0.1	1	0	0.7	0.2	0.9
89.	0.3	0.7	0.1	0.7	0	0	0.5	1	0.9	0.9	0.8	0.6	0.5	0.5
90.	0.8	0.5	0.9	0.7	0.6	0.1	0.9	0.9	0.9	0.9	0.9	0.9	0.2	0.2
91.	0.3	0.7	0.6	1	0	0.1	1	1	0.7	1	0.2	0.7	1	0.3
92.	0.3	0.6	0.6	0.9	0	0.4	1	1	0.5	1	0.1	0.8	0.5	0.8
93.	1	1	0.5	0.8	0.6	0	0.4	0	0.4	0.5	0.4	0.8	0.9	1
94.	1	1	1	1	0	1	1	0.6	1	1	0	0.3	0.1	1
95.	0.8	0.3	1	0.9	0.1	0.3	0.8	1	0.9	0.9	0.8	0.8	0.2	0.8
96.	0.2	0.8	0.2	0.8	0.3	0.3	0.5	1	0.6	0.9	0.1	0.5	0.4	0.1

97.	0.5	0.4	0.4	0.4	0	0	0.5	0.5	1	0.5	0.5	0.5	1	0.4
98.	0.1	0.3	1	0.1	0	0	0.7	0.2	0.3	0.4	0.1	0	0	0.6
99.	0.8	1	1	0.1	0.5	1	1	0.4	0.5	1	0.2	1	0.9	1
100.	0.5	1	0.5	0.1	0	0	0	1	1	1	0	1	1	1
101.	0.8	0.7	0.7	0.8	0	0.4	0.6	1	0.6	0.5	0	0.2	0.6	0.7
102.	0.7	0.9	0.9	0.9	0.9	1	1	1	0.9	1	1	0.9	1	0.9
103.	0.6	0.1	0.2	1	0	0	0.5	1	0.1	1	0	0.5	0.8	0.5
104.	0.2	0.2	1	0.2	0.2	0	0.6	1	1	0.3	0.2	1	0.1	0.2
105.	1	0.5	1	0.6	1	0	1	1	1	1	1	1	1	1
106.	0.3	0.4	0	0.7	0	0.4	0	0.6	0.3	0.6	0.8	0.7	0.3	0.6
107.	0.8	0.9	0.8	0.9	0.3	0.4	0.3	1	0.9	0.9	0.3	0.9	0.1	1
108.	1	1	1	1	0	0.5	0.8	1	0.1	1	0	1	0.8	0.8
109.	0	0	0.1	1	0	0	0.9	0.8	0	1	0.8	0	0	0
110.	0.9	0.8	0.9	0.9	0.3	0.7	0.9	0.9	0.9	0.9	0.8	0.9	0.9	0.9
111.	0.2	0.1	0	0.4	0	0.7	1	0.8	1	0.8	0	1	0.5	0.3
112.	0.8	0.3	0	0.6	0	0	0.8	1	0.8	0.9	0.3	0.5	0.5	0.2
113.	0.7	1	0.8	0.4	0	0	0.5	1	0.5	0.8	0	1	1	1
114.	0.3	0.5	0.5	0.4	0.2	0.1	1	1	0.2	1	1	1	1	1
115.	1	1	0.5	1	0.6	0.2	0.5	0.5	0.5	1	0.5	0.5	0.5	0.7
116.	0	1	0.5	0.9	0	0.3	0.8	0.5	0	1	0.4	0.2	0.5	0
117.	0.8	0.7	0.4	0.8	0.1	0.9	0.9	0.7	0.8	0.9	0.6	0.8	0.8	0.4

118.	0.8	0	0	1	0.7	1	1	1	1	1	0.7	0	1	0.6
119.	0.5	0	0	0.6	0	0	0	0	0	0	0	0	0	0
120.	1	1	1	1	0.4	0.2	0.2	0.9	1	0.5	0.5	0.4	0.3	0.2
121.	0.5	0.4	0.4	0.5	0	0.4	0.6	0.6	0.6	0.8	0.3	0.6	0.6	0.4
122.	0.2	0.4	0.1	0.6	0.5	0.2	0.8	0.8	0.3	1	0.8	0.5	0.6	0.2
123.	1	1	0.5	1	0.5	0.6	0.8	1	1	1	0.9	1	0.5	0.9
124.	0.9	0.9	0.5	0.8	0	0	0.9	0.9	0.5	0.9	0	0.7	0.6	0.9
125.	0.5	1	0	0	0	0	1	1	0.7	1	0	0	0.5	0.5
126.	0.5	0.6	0.3	0.8	0	0.6	0.5	0.7	0.7	0.8	0.1	0.6	0.8	0.7
127.	0.3	0.4	0.5	0.7	0.2	0.8	0.6	1	1	1	1	1	1	1
128.	0.5	0.8	0.8	0	0.5	0.3	0.4	0.8	0.7	0.8	0.5	0.8	0.8	0.8
129.	0	1	1	0	0	1	1	0	0.5	1	1	1	1	1
130.	0	0.4	1	1	0	0	0	1	1	1	1	1	1	1
131.	1	0.3	1	0.1	0.1	1	1	0.1	0.1	1	0.1	1	1	1
132.	1	0.4	0	0.9	0	0.3	1	0.9	1	1	0	1	0	1
133.	0.9	0	0.5	1	0.5	1	1	1	0.9	0.9	0.5	0.6	0.6	0.8
134.	0.6	0.4	0.2	0.4	0	1	0.8	0.8	0.2	0.7	0.2	0.3	0	0.8
135.	0.8	0.4	0.6	0.8	0.1	0.2	0.9	0.5	0.7	1	0.1	1	0.9	1
136.	0.1	0.1	0.2	0.6	0	0.1	0.7	0.1	0.1	1	0.6	0	0	0
137.	1	0.7	0.9	0.9	0.7	0.8	0.8	1	1	1	0.9	0.2	0.9	1

Appendix C: Expert-based data

Subject-based data is presented in Table C. 2.

Table C. 2. Expert-based data for 14 signs and 137 participants

Expert-based data														
.No	NAFUP	UBC	DNT	DNWSH	COSACOEOT	WRP	WAFS	DNWSH-T	NAFUP-T	WAFS-T	COSACOEOT-T	UBC-T	WRP-T	DNT-T
1.	0.3	0.3	0.6	0.7	0	0.6	1	0.9	0.9	1	0.8	1	0.6	1
2.	0.3	0.3	1	0.6	0	0.6	1	0.2	0.8	1	0.7	0.4	0.6	1
3.	0.3	0.1	1	0.6	0	0.8	1	1	1	1	1	1	0.8	1
4.	0.4	0.3	1	0.4	0	0.6	1	1	1	1	1	1	1	1
5.	0	0.5	0.7	0.4	0	0.3	1	1	0.6	1	1	0	0.7	1
6.	0.2	0.1	1	0.6	0	0.3	0.9	1	0.1	1	0.8	1	0.6	1
7.	0.6	0.3	1	0	0.3	0.3	0.6	0	0	1	0.4	0.8	0.9	1
8.	0.3	0.3	1	0	0	0.6	1	0.5	0.6	1	0.4	0	0.9	1
9.	0.2	0.3	1	0.6	0	0.2	1	0.9	0.1	1	0	0.6	1	1
10.	0.2	0.3	0.5	0.5	0	0.3	0	0.9	0.5	1	0.5	0.5	0.5	1
11.	0.2	0.3	0.3	0	0	0.3	1	0	0.5	0.8	0	0.4	0.2	0.8
12.	0.2	0.3	1	0.8	0	0	1	0.9	0.8	1	0.5	1	1	1
13.	0.2	0.3	0.6	0.8	0	0.2	0.8	0.9	0.8	1	0.8	0.6	0.3	0.7
14.	0.2	0.4	1	0.6	0	0.3	1	0.7	0.1	1	0	0.6	0.3	1

15.	0.2	0.3	0.4	0.3	0	0	0.6	0.7	1	1	0	0.3	0	1
16.	0.2	0.3	1	0.8	0	0.5	0.6	0.7	0.1	0.8	1	1	1	1
17.	0	0.3	1	0	0	0	1	0.7	0.6	0.8	0.5	1	0.3	1
18.	0	0.3	1	0.6	0	0.3	0.6	1	1	1	0	0.8	0.9	1
19.	0.2	0.3	1	0.5	0	0.3	1	1	0.1	0.8	0	0.4	0.8	1
20.	0.2	0.3	1	0.5	0	0	1	1	0.1	1	0	0.3	0	1
21.	0.2	0.3	0	0.8	0	0	0.8	0.7	0.1	0.7	0.6	0.3	0	1
22.	0	0.3	0.3	0.4	0	0	1	0.7	0.8	1	0.8	1	1	1
23.	0	0.3	1	0.4	0	0	0	0	0	0	0	0	0	0
24.	0.2	0.3	1	1	0.6	0.6	1	1	1	1	0.9	1	1	1
25.	0.2	0.3	0.4	0.9	0	0.5	0.8	0.7	0.1	0.8	0	0.4	0.3	0.3
26.	0	0	0	1	0	0	0.8	1	1	1	1.6	0.1	0.1	0
27.	0.2	0.3	0	0.8	0	0.3	0.8	0.9	0.9	1	0.7	1	0.8	1
28.	0	0	0	1	0	0	0	1	1	0	1	0	0	0.5
29.	0.2	0.3	1	0.3	0	0.6	1	0.7	1	1	0.8	0.4	0.6	1
30.	0.2	0.3	0	0.3	0	0.5	0	0.3	0.1	1	1	0	0	0
31.	0	0.3	0.7	0.6	0	0.3	1	1	1	1	1	1	1	1
32.	1	1	1	1	1	1	1	1	1	1	1	1	1	1
33.	0.2	0.3	1	1	1	0	0.8	1	0	1	0	0.3	0	1
34.	0.3	0.3	0	0.6	0	0	0.8	0.6	0.8	0.8	0.7	0.4	0	0
35.	0.1	0.3	0.3	0.1	0	0.5	0.6	0.7	0.8	1	0.6	1	1	1

36.	0.2	0	0	0.4	0	0	0.8	1	0.1	1	0	0	0.6	0.6
37.	0	0.3	1	0.6	0	0	1	0.7	0.6	1	0.7	0.7	0.8	1
38.	0.2	0.3	1	0.5	0.7	0.6	1	1	0.1	1	0.8	0.4	0.6	1
39.	0.9	0.3	1	0.9	0	0.3	0.8	0.8	0.6	0.8	0.8	1	0.8	1
40.	0.2	0.3	0.7	0.5	0	0.3	0.6	0.5	0.4	1	0.8	0.3	0	1
41.	0.2	0.3	1	0.1	0	0	0.5	0	0	0.3	0	0	0	1
42.	0	0.3	0	0.5	0	0.6	0.5	0.7	0	1	0	0	0.6	0
43.	0	0.3	0.4	0.4	0	0	0.8	0.9	0.6	1	0.2	0.7	1	1
44.	0.2	0.3	1	0.4	0.7	0	0.5	0.5	0.1	0.6	0.8	0.3	0	1
45.	0.2	0.3	0	0.4	0.5	0	0.5	0.2	0	0.6	0.6	0.3	0	0.3
46.	0.2	0.3	0.3	0.3	0.4	0	1	0.5	0.4	1	0.1	0.3	0	0.6
47.	0.2	0.4	0	0.3	0	0.5	0.8	1	0.1	1	0	0.3	0.9	0
48.	0.2	0.4	1	0.6	0	0.6	1	0.7	0.1	1	0	0.3	0.6	1
49.	0.2	0.4	1	0.6	0	0.6	1	0.5	0.1	1	0	0.3	0.6	1
50.	0.2	0	1	0.5	0	0.3	0.5	0.5	0.6	0.7	0	0	0.3	1
51.	1	0.3	1	1	0.7	0.3	0.8	0.7	1	1	0.8	1	1	1
52.	0.2	0.3	1	0.7	0	0.6	1	0.5	0	1	0	0.3	1	1
53.	0.2	0.3	0	0.5	0	0	1	0.9	1	1	0.8	1	1	1
54.	0.2	0.3	0.5	0.6	0	0	0.6	0.8	1	1	0.8	0	1	1
55.	0.4	0.3	0.8	0.5	0	0.3	0.6	0.2	0.2	0.7	0	0.5	0.3	1
56.	0.2	0.3	1	0.4	0	0	0.8	0.8	0.1	1	1	0.5	0.5	1

57.	0	0	0	0.6	0	0	0	0.4	0	0	0	1	0	0.9
58.	0.2	0.3	0	0.4	0	0	0.5	0.6	0.1	0.8	0	0.5	0	1
59.	0.4	0.4	0.8	0.6	0	0.3	0.5	0.5	0.4	0.8	0.6	0.5	0.3	0.8
60.	0.2	0	0	0.6	0	0.5	0.7	0	0	1	0	0.5	0.5	0
61.	0.2	0.6	0	0.6	0	0	0.7	1	0.5	0.8	0.8	1	0.9	1
62.	0.1	0.3	1	0.6	0.4	0	0.8	0.6	0.1	1	1	0.6	0	1
63.	0.4	0.4	0.4	0.2	0	0	0.7	0.7	1	1	0	1	1	1
64.	0.1	0.3	0.2	0.5	0	0	0.9	0.6	0.1	0.8	0	0.6	0	0
65.	0.1	0.6	1	0.5	0	0.6	1	0.4	1	1	0.6	0.9	0.6	1
66.	0.2	0.3	0.7	0.7	0.1	0.6	0.8	0.8	0.6	1	0	1	0.6	1
67.	0.6	0.6	1	0.7	0	0.3	0.7	0	0.6	0.8	0	0.5	0.3	1
68.	0.2	0.3	1	0.7	0	0.3	0.5	0.8	0.6	0.8	0.5	0.9	0.9	1
69.	0.1	0.6	1	0.5	0.6	0.9	0.7	0.8	0.4	1	0.6	0.3	0.9	1
70.	0.1	0	1	0.4	0	0	1	0.8	0.1	1	0	0	0	1
71.	0.1	0.3	0	0.5	0	0	0.8	0.8	0.1	0.8	0.8	0.6	0	1
72.	0.5	0.3	0.6	0.9	0	0.4	1	0.9	0.6	1	0	0.6	0.8	0.7
73.	0.1	0.3	1	0.4	0	0	1	0.9	0.3	1	0.4	1	0.8	1
74.	0.1	0.6	0.2	0	0	0.6	0.8	1	0.8	1	0.8	1	1	1
75.	0	0.3	1	0.9	0	0.4	0.8	1	0.3	1	0.8	1	1	1
76.	0.1	0.3	0	0.5	0	0.3	0.5	1	0.8	1	0.8	1	0.3	1
77.	0.1	0.8	0.4	0.3	0	0.6	0.7	1	1	1	0	1	1	1

78.	0	0.3	0.4	0.8	0	0.6	0.7	0.8	0	1	0	0.3	0.6	1
79.	0	0.3	1	0.6	0.3	0.5	0.7	1	0.6	0.8	0.7	0.8	0.8	1
80.	0.1	0.3	1	0.4	0.6	0.5	0.7	0.6	0.4	0.8	0.6	0.5	0.5	1
81.	0.1	0.3	1	0.9	0	0.6	1	1	0.5	1	0	0.5	0.5	1
82.	0.2	0.3	1	0.2	0	0	0.7	0.1	0.4	0.8	0.3	0.3	0.9	1
83.	0	0.3	0	0.5	0	0	1	0.3	0.4	0.8	0.7	0	0.3	0.9
84.	0.3	0.3	0	0	0	0	0.6	1	0.6	0.5	0.4	0.3	0	1
85.	0	0.3	0.3	1	0	0.3	1	1	0.6	1	0.7	0.9	0.8	1
86.	0	0.3	1	0	0	0	0.9	0	0.1	0.9	1	0.6	0	1
87.	0.1	0.3	0.1	0.2	0	0.7	0.8	0.6	0.9	0.8	0.7	0.6	0.8	1
88.	0.1	0.1	1	0.5	0	0.5	0.6	1	0.1	0.8	0.6	0.6	0.3	1
89.	0.1	0.3	1	0.5	0	0	0.6	1	1	1	0.8	1	1	1
90.	0.1	0	0.2	0.1	0	0	0.8	0.8	0.9	1	0.5	1	0	0
91.	0.4	0.3	0.6	0.4	0	0.1	0.8	0.8	0.4	0.8	0.7	0.9	0.3	0.5
92.	0	0.3	1	0.9	0	0.2	0.8	0.8	0.3	0.8	0	0.9	0.3	1
93.	0	0.3	0.2	0.3	0	0	0	0	0	1	0	0.3	1	0
94.	0	0.3	1	0.2	0	0.5	0.7	0.6	1	1	0	0.3	0	1
95.	0.1	0.3	1	0.2	0	0	0.8	1	0.3	1	0.7	0.5	0	1
96.	0.1	0.3	0.6	0.5	0	0.5	0.6	0.5	0.6	1	0	0.9	0.3	0
97.	0.2	0.3	1	0.3	0	0	0.6	0.8	1	1	0.6	0.9	1	1
98.	0.2	0.3	1	0.2	0	0	0.8	1	0.1	0.8	0	0.3	0	1

99.	0.1	0.5	1	0.2	0	0.4	0.8	0.1	0.1	0.8	0	0.9	0.6	1
100.	0.1	0.2	1	0.3	0	0	0	1	1	1	0	0.5	1	1
101.	0.1	0.3	1	0.9	0	0.4	0.8	1	0.6	1	0	0.5	0.6	1
102.	0.1	0.6	1	0.4	0	0.6	0.8	0.7	0.8	1	0.5	0.5	0.6	1
103.	0	0.3	0.5	0.3	0	0	0.5	1	0	1	0.8	0.9	0.6	1
104.	0	0.3	1	0.3	0	0	0.5	1	1	1	1	1	0	1
105.	0.9	0.3	1	0.8	0.7	0	0.6	1	0.9	1	0.7	1	1	1
106.	0.1	0.3	0	0.4	0	0.4	0	0.5	0.1	0.8	0.8	1	0.3	1
107.	0.2	0.6	1	0.6	0	0.4	0	1	0.5	0.8	0	1	0.3	1
108.	0.2	0.3	0.4	0.3	0	0.6	0.9	0.2	0.1	1	0	1	0.6	0.6
109.	0	0	0	0.3	0	0	0.9	0.8	0	1	0.8	0	0	0
110.	1	0.6	1	0.5	0	0.7	0.8	0.8	1	0.8	0.9	1	1	1
111.	0.1	0.3	0	0.4	0	0.9	1	1	1	1	1	1	1	1
112.	0.1	0.3	0	0.9	0	0.3	0.8	1	1	1	0	0.5	0.1	0
113.	0.1	0.3	1	0.3	0	0	0.8	1	1	1	1	1	1	1
114.	0.1	0.3	1	0.3	0	0	1	1	1	1	1	1	1	1
115.	0.2	0.6	1	0.3	0	0	0.8	0.8	0.6	1	0.7	0.9	0.6	1
116.	0	0.3	0.5	0.9	0	0.3	0.6	0.3	0	1	0.8	0.3	1	1
117.	0.1	1	1	0.9	0	0.9	0.8	0.8	1	0.8	0.8	1	0.8	1
118.	0.6	0	0	0.9	0.7	0.7	1	0.8	0.1	1	0.8	0	0.6	0
119.	0.1	0	0	0.2	0	0	0	0	0	0	0	0	0	0

120.	0.1	0.3	0.4	0.3	0.2	0.3	0	0.6	0.6	1	0.8	0.3	0	0.4
121.	0.1	0.3	0.3	0.9	0	0.3	1	0.6	0.6	1	0.7	0.9	0.6	1
122.	0.3	0.3	0	0.3	0.5	0	0.8	1	0.8	0.9	0.8	1	0.6	1
123.	0	0.2	1	0.3	0.6	0.6	0.4	0	0.6	0.5	0.7	0.3	0.6	0.8
124.	0.2	0.3	1	0.9	0	0	0.5	0.5	0.1	1	0	0.3	1	1
125.	0	0.2	0	0	0	0	1	0.7	0.9	1	0	0	1	1
126.	0	0.3	0.6	0.7	0	0.6	1	0.9	1	1	0	0.4	0.6	1
127.	0	0.3	0.6	0.4	0	0.6	0.5	1	0.9	0.8	0.8	1	0.9	1
128.	0	0.3	0.8	0	0	0	0	0.6	0.8	0.8	0.4	0.6	0.9	1
129.	0.1	0.6	1	0	0	0.6	1	0	0.7	1	0.8	0.6	0.6	1
130.	0	0.3	1	0.9	0	0	0	1	0.6	1	0.8	0.5	1	1
131.	0	0.2	1	0	0	0.6	0.8	0	0	0.8	0	0.3	0.5	1
132.	0.1	0.3	0	0.9	0	0	0.6	0.8	0.1	0.8	0	0.3	0	0.7
133.	0.1	0	0	0.4	0	0.6	0.8	0.7	1	1	0.8	0.7	1	1
134.	0	0.3	0	0.4	0	0	0.8	1	0	1	0	0.8	1	1
135.	0	0.2	0.5	0.4	0	0.3	0.8	0.4	0	1	0	0.5	0.6	1
136.	0	0.3	0	0.5	0	0.3	0.8	0.9	0.6	1	0.7	0.5	1	0.5
137.	0	0.8	1	0.9	0.4	0.9	0.7	0.6	0.4	1	0.4	0.8	0.9	1

Appendix D: Computational data

Computational data is presented in Table D. 3.

Table D. 3. Computational data for 14 signs and 137 participants

Expert-based data														
.No	NAFUP	UBC	DNT	DNWSH	COSACOEOT	WRP	WAFS	DNWSH-T	NAFUP-T	WAFS-T	COSACOEOT-T	UBC-T	WRP-T	DNT-T
1.	0.3	0.3	0.5	0.6	0	0.5	0.9	0.6	0.8	1	0.7	1	0.6	1
2.	0.3	0.2	0.8	0.5	0	0.6	0.9	0.2	0.7	1	0.6	0.4	0.6	0.6
3.	0.1	0.1	1	0.5	0.2	0.7	0.9	1	1	1	1	0.7	0.7	1
4.	0.4	0.2	1	0.2	0.1	0.5	0.6	1	1	1	1	1	1	1
5.	0.1	0.3	0.7	0.4	0	0.3	0.9	1	0.6	0.9	1	0	0.7	1
6.	0.1	0.1	0.9	0.4	0	0.3	0.8	0.9	0.1	0.9	0.8	1	0.5	0.9
7.	0.5	0.2	1	0	0.3	0.4	0.5	0	0.1	0.9	0.3	0.7	0.8	0.9
8.	0.3	0.3	0.9	0.1	0	0.5	0.8	0.4	0.6	0.8	0.5	0	0.8	1
9.	0.1	0.2	0.9	0.6	0.2	0.2	0.8	0.6	0.1	0.8	0.1	0.5	1	1
10.	0.1	0.2	0.5	0.4	0	0.4	0.1	0.8	0.5	0.8	0.6	0.4	0.5	1
11.	0.1	0.2	0.3	0.3	0	0.3	0.8	0.1	0.5	0.8	0	0.2	0.3	0.6
12.	0.2	0.3	1	0.5	0.2	0	0.9	0.6	0.8	0.9	0.6	1	1	1
13.	0.1	0.4	0.2	0.6	0	0.2	0.4	0.6	0.7	0.9	0.8	0.5	0.4	0.6

14.	0.2	0.4	1	0.5	0	0.4	0.9	0.8	0.4	0.9	0	0.5	0.4	1
15.	0.1	0.2	0.3	0.2	0.1	0.1	0.5	0.7	0.9	1	0.1	0.2	0.1	1
16.	0.1	0.2	1	0.7	0	0.5	0.6	0.7	0.2	0.7	1	1	1	1
17.	0.2	0.4	0.7	0.1	0	0.1	0.8	0.6	0.4	0.8	0.7	1	0.4	1
18.	0.1	0.3	0.9	0.5	0	0.4	0.6	1	0.8	0.9	0	0.7	0.8	1
19.	0	0.2	1	0.4	0	0.4	0.8	0.7	0	0.8	0	0.4	0.8	1
20.	0.3	0.3	1	0.4	0	0.1	0.9	0.8	0.2	0.9	0	0.3	0.1	1
21.	0.2	0.2	0.2	0.6	0	0	0.7	0.6	0.2	0.7	0.5	0.3	0	0.2
22.	0.1	0.4	0.3	0.2	0	0.2	0.7	0.6	0.7	1	0.5	1	1	0.9
23.	0.2	0.1	0.7	0.2	0	0	0.1	0	0.1	0.1	0	0	0	0.2
24.	0.1	0.3	1	0.9	0.6	0.6	1	1	0.9	1	0.9	1	1	1
25.	0.1	0.2	0.6	0.6	0	0.5	0.8	0.7	0.1	0.8	0	0.4	0.4	0.4
26.	0	0	0.1	0.9	0	0.1	1	0.9	1	1	0.1	0.1	0.2	0.1
27.	0.2	0.2	0.2	0.6	0	0.4	0.4	0.8	0.8	0.9	0.7	1	0.8	1
28.	0	0	0.2	0.8	0	0.1	0.1	0.8	1	0.1	1	0	0.1	0.1
29.	0.2	0.3	0.9	0.2	0	0.6	0.9	0.7	1	1	0.7	0.3	0.6	0.9
30.	0.1	0.2	0.1	0.3	0.1	0.5	0.1	0.2	0.1	0.9	1	0.1	0	0.2
31.	0.2	0.4	0.1	0.3	0	0.4	0.9	1	1	0.9	1	1	1	1
32.	0.9	1	1	1	1	1	0.9	1	0.9	1	1	1	1	1
33.	0.3	0.3	0.6	0.8	1	0.2	0.8	0.8	0.3	1	0.1	0.2	0.2	1
34.	0.3	0.2	0.2	0.4	0	0	0.8	0.6	0.8	0.8	0.6	0.4	0.1	0.3

35.	0.2	0.2	0.1	0.1	0	0.5	0.6	0.6	0.7	1	0.7	1	1	1
36.	0.2	0.1	0.3	0.4	0	0.2	0.8	1	0	0.9	0	0	0.6	0.5
37.	0.1	0.5	0.9	0.6	0	0.1	0.9	0.7	0.6	0.9	0.6	0.7	0.8	1
38.	0.1	0.4	0.6	0.2	0.7	0.6	0.9	0.9	0.1	0.9	0.7	0.4	0.6	1
39.	0.5	0.3	0.9	0.6	0	0.4	0.8	0.6	0.5	0.8	0.8	0.9	0.9	0.9
40.	0.1	0.3	0.6	0.4	0	0.4	0.6	0.4	0.3	1	0.7	0.1	0	1
41.	0.4	0.3	0.6	0.2	0	0	0.5	0.1	0	0.3	0	0	0	0.7
42.	0	0.2	0.2	0.3	0	0.6	0.5	0.7	0	0.9	0	0	0.6	0.2
43.	0.1	0.3	0.6	0.2	0	0.1	0.2	0.9	0.5	0.9	0.2	0.7	1	0.9
44.	0.1	0.3	0.3	0.3	0.7	0.2	0.5	0.4	0.4	0.5	0.8	0.3	0.1	0.4
45.	0.1	0.3	0.2	0.4	0.4	0.2	0.6	0.2	0	0.5	0.4	0.3	0	0.2
46.	0.1	0.3	0.3	0.4	0.3	0.1	1	0.4	0.4	0.9	0.2	0.3	0.1	0.7
47.	0.1	0.4	0.2	0.3	0	0.5	0.7	0.9	0.1	0.9	0	0.3	0.8	0.3
48.	0.1	0.3	0.7	0.4	0	0.6	1	0.4	0.1	1	0	0.4	0.6	0.7
49.	0.1	0.3	0.7	0.4	0	0.6	1	0.4	0.1	1	0	0.4	0.6	0.7
50.	0.1	0.1	1	0.5	0	0.4	0.4	0.5	0.6	0.6	0	0	0.4	1
51.	0.9	0.3	0.9	0.9	0.7	0.4	0.8	0.7	0.9	0.9	0.7	1	1	1
52.	0.2	0.2	0.9	0.7	0	0.6	0.9	0.7	0	1	0	0.2	0.6	1
53.	0.1	0.3	0.3	0.5	0	0	0.9	0.7	0.9	0.9	0.9	1	1	1
54.	0.1	0.2	0.6	0.6	0	0	0.7	0.6	1	1	0.9	0.3	1	0.5
55.	0.4	0.2	0.3	0.2	0	0.4	0.6	0.1	0.3	0.6	0	0.5	0.4	0.6

56.	0.2	0.3	1	0.1	0	0	0.8	0.6	0.2	1	0.9	0.5	0.7	1
57.	0	0	0.3	0.4	0	0.1	0.1	0.4	0	0	0	1	0	0.7
58.	0.1	0.2	0.2	0.4	0.1	0	0.5	0.6	0.1	0.8	0.1	0.4	0	0.9
59.	0.4	0.4	0.5	0.4	0	0.4	0.5	0.5	0.4	0.7	0.7	0.4	0.4	0.5
60.	0.1	0	0.4	0.7	0	0.5	0.7	0.2	0	0.9	0	0	0.6	0.4
61.	0.2	0.5	0.3	0.4	0	0.1	0.6	1	0.4	0.7	0.7	1	0.8	1
62.	0.1	0.3	1	0.5	0.3	0.1	0.8	0.5	0.1	0.7	0.9	0.5	0.1	1
63.	0.3	0.4	0.4	0.2	0	0.1	0.7	0.6	0.8	0.9	0	0.8	1	1
64.	0.1	0.2	0.3	0.3	0	0	0.7	0.5	0.1	0.8	0	0.5	0.1	0.1
65.	0.1	0.5	0.6	0.3	0	0.6	0.9	0.4	0.8	0.9	0.7	0.7	0.6	0.8
66.	0.2	0.3	0.7	0.6	0.1	0.6	0.8	0.6	0.6	1	0	1	0.6	1
67.	0.5	0.5	0.7	0.1	0	0.4	0.8	0.1	0.5	0.8	0	0.5	0.4	0.7
68.	0.1	0.2	0.9	0.6	0	0.4	0.4	0.6	0.6	0.8	0.5	0.8	0.8	0.9
69.	0.1	0.5	0.8	0.4	0.4	0.8	0.7	0.6	0.3	1	0.4	0.2	0.8	1
70.	0.2	0	0.9	0.4	0	0.1	0.9	0.4	0.2	0.9	0	0.1	0.1	0.9
71.	0.1	0.3	0.3	0.5	0	0	0.8	0.6	0	0.8	0.6	0.6	0	1
72.	0.3	0.2	0.5	0.5	0	0.4	0.9	0.6	0.3	0.9	0.1	0.6	0.7	0.5
73.	0.1	0.2	1	0.3	0	0	0.9	0.7	0.3	0.9	0.5	1	0.8	1
74.	0.1	0.5	0.2	0.1	0	0.5	0.8	0.7	0.6	1	0.7	1	0.9	1
75.	0.1	0.3	1	0.6	0	0.4	0.8	1	0.3	1	0.8	0.9	1	0.8
76.	0.1	0.2	0.2	0.5	0	0.4	0.5	1	0.8	1	0.8	0.7	0.4	0.7

77.	0	0.7	0.3	0.3	0	0.5	0.6	1	0.9	1	0	1	1	1
78.	0.1	0.2	0.6	0.6	0	0.6	0.7	0.6	0.1	0.7	0	0.2	0.6	0.6
79.	0.2	0.5	0.6	0.6	0.4	0.5	0.8	0.9	0.6	0.8	0.5	0.7	0.6	1
80.	0.1	0.2	0.7	0.4	0.6	0.5	0.6	0.4	0.4	0.7	0.6	0.4	0.5	1
81.	0.1	0.3	1	0.8	0.1	0.7	0.9	1	0.4	0.9	0	0.5	0.5	1
82.	0.2	0.3	1	0.4	0	0	0.8	0.3	0.4	0.8	0.3	0.2	0.9	1
83.	0.1	0.3	0.2	0.5	0	0.2	0.9	0.3	0.4	0.9	0.6	0	0.4	0.8
84.	0.3	0.3	0.4	0.2	0.4	0.1	0.6	1	0.2	0.4	0.4	0.3	0.1	0.9
85.	0.1	0.3	0.5	0.9	0	0.4	0.8	0.9	0.5	0.9	0.7	0.7	0.7	0.9
86.	0.1	0.3	0.9	0.1	0	0	0.8	0.1	0.1	0.8	0.9	0.5	0	0.9
87.	0.1	0.1	0.2	0.3	0.5	0.6	0.8	0.7	0.7	0.8	0.5	0.4	0.6	1
88.	0.2	0.1	0.9	0.3	0	0.5	0.6	0.6	0.1	0.8	0.5	0.5	0.4	0.9
89.	0.1	0.3	0.9	0.5	0.2	0	0.7	0.9	1	1	0.9	1	1	0.9
90.	0.1	0.1	0.4	0.2	0	0.1	0.8	0.6	0.6	0.6	0.7	1	0.1	0
91.	0.2	0.3	0.1	0.3	0	0.2	0.8	0.6	0.3	0.8	0.7	0.7	0.3	0.4
92.	0.1	0.3	0.7	0.6	0	0.2	0.8	0.6	0.1	0.8	0	0.7	0.4	1
93.	0	0.2	0.3	0.1	0	0.1	0.1	0.1	0	0.7	0.1	0.2	1	0.1
94.	0.1	0.2	1	0.2	0	0.5	0.7	0.5	0.1	1	0	0.2	0.2	1
95.	0	0.3	1	0.3	0	0.1	0.8	1	0.4	1	0.8	0.5	0.1	1
96.	0.1	0.3	0.5	0.5	0.1	0.6	0.5	0.5	0.6	0.9	0.1	0.7	0.4	0.4
97.	0.1	0.3	0.9	0.3	0	0	0.6	0.7	0.9	0.9	0.6	0.7	1	0.9

98.	0.1	0.2	0.9	0.3	0	0	0.8	0.9	0.1	0.8	0	0.2	0.1	0.9
99.	0.1	0.4	0.7	0.2	0	0.4	0.8	0.2	0.1	0.8	0	0.5	0.6	0.4
100.	0.1	0.2	0.8	0.4	0.1	0	0.1	0.9	0.9	1	0	0.5	1	0.9
101.	0.1	0.3	1	0.6	0	0.2	0.8	0.9	0.6	0.9	0	0.4	0.6	1
102.	0.3	0.5	0.6	0.4	0.1	0.6	0.8	0.5	0.7	0.9	0.6	0.5	0.6	0.9
103.	0.1	0.2	0.3	0.2	0	0.2	0.5	0.8	0.1	1	0.8	0.7	0.6	1
104.	0	0.1	0.3	0.2	0.2	0.1	0.6	0.8	1	1	1	1	0	1
105.	0.6	0.2	1	0.6	0.6	0	0.7	1	0.7	1	0.6	0.9	0.9	1
106.	0.1	0.3	0.2	0.4	0	0.4	0	0.4	0.1	0.4	0.8	1	0.4	0.9
107.	0.2	0.5	0.9	0.5	0	0.4	0.1	0.9	0.4	0.9	0	1	0.4	0.9
108.	0.1	0.1	0.4	0.3	0	0.6	0.7	0.2	0	1	0	1	0.6	0.6
109.	0.1	0.1	0.5	0.3	0	0.1	0.7	0.6	0.1	1	0.7	0.1	0.1	0.5
110.	0.8	0.5	1	0.5	0	0.5	0.8	0.7	0.8	0.8	0.9	1	1	1
111.	0.1	0.3	0.1	0.4	0	0.7	1	1	1	1	1	1	1	1
112.	0.4	0.2	0.4	0.7	0	0.3	0.7	1	1	1	0	0.5	0.2	0.3
113.	0.1	0.2	1	0.2	0	0	0.8	1	0.8	1	1	1	1	1
114.	0.1	0.3	0.8	0.3	0	0	0.9	0.9	0.9	0.9	0.9	1	1	0.8
115.	0.1	0.5	0.9	0.4	0	0.1	0.8	0.6	0.6	0.9	0.5	0.7	0.6	0.9
116.	0.1	0.2	0.4	0.5	0	0.1	0.5	0.2	0.2	1	0.6	0.2	1	1
117.	0.1	1	0.9	0.5	0	0.6	0.7	0.6	0.9	0.7	0.8	1	0.7	1
118.	0.6	0.1	0.2	0.6	0.8	0.6	0.9	0.6	0.1	0.9	0.8	0	0.6	0.2

119.	0.1	0	0.1	0.3	0	0	0.1	0.1	0.1	0.1	0	0	0	0.1
120.	0.1	0.3	0.4	0.4	0.2	0.2	0.2	0.6	0.6	1	0.8	0.2	0.1	0.4
121.	0.2	0.3	0.5	0.6	0	0.4	0.9	0.7	0.6	1	0.5	0.7	0.6	0.9
122.	0.3	0.2	0.3	0.4	0.5	0.1	0.8	1	0.8	0.8	0.7	1	0.6	1
123.	0.1	0.2	0.7	0.4	0.6	0.5	0.4	0.3	0.4	0.4	0.6	0.2	0.4	0.7
124.	0.1	0.3	0.9	0.5	0	0.1	0.5	0.4	0.1	0.9	0	0.2	0.8	0.9
125.	0	0.2	0.2	0.1	0	0	0.9	0.5	0.7	0.8	0	0	1	1
126.	0	0.4	0.5	0.5	0	0.6	0.9	0.3	0.5	0.8	0	0.4	0.6	0.7
127.	0.1	0.2	0.6	0.4	0.2	0.5	0.5	0.9	0.9	0.8	0.8	1	0.8	1
128.	0	0.2	0.6	0.1	0.1	0.1	0.1	0.5	0.6	0.8	0.3	0.5	0.8	0.9
129.	0.4	0.5	1	0.2	0	0.6	0.8	0.2	0.6	0.9	0.7	0.5	0.6	1
130.	0.1	0.2	0.9	0.4	0	0	0	0.8	0.5	0.9	0.7	0.4	1	0.9
131.	0.2	0.2	0.6	0.2	0	0.5	0.8	0.2	0.2	0.8	0	0.2	0.5	0.6
132.	0.1	0.2	0.3	0.7	0	0	0.6	0.7	0.1	0.6	0	0.2	0	0.7
133.	0.1	0.1	0.5	0.2	0	0.6	0.7	0.6	1	0.9	0.7	0.7	1	0.9
134.	0.1	0.2	0.2	0.3	0	0	0.8	0.9	0	0.9	0	0.5	1	0.9
135.	0	0.3	0.4	0.4	0	0.4	0.9	0.4	0.1	0.9	0	0.4	0.6	0.9
136.	0.2	0.3	0.4	0.5	0	0.3	0.7	0.7	0.6	0.8	0.6	0.4	1	0.4
137.	0.3	0.6	0.7	0.7	0.5	0.7	0.7	0.7	0.4	0.9	0.2	0.7	0.8	1

Appendix E: Subject-based data for the eye-tracking experiment

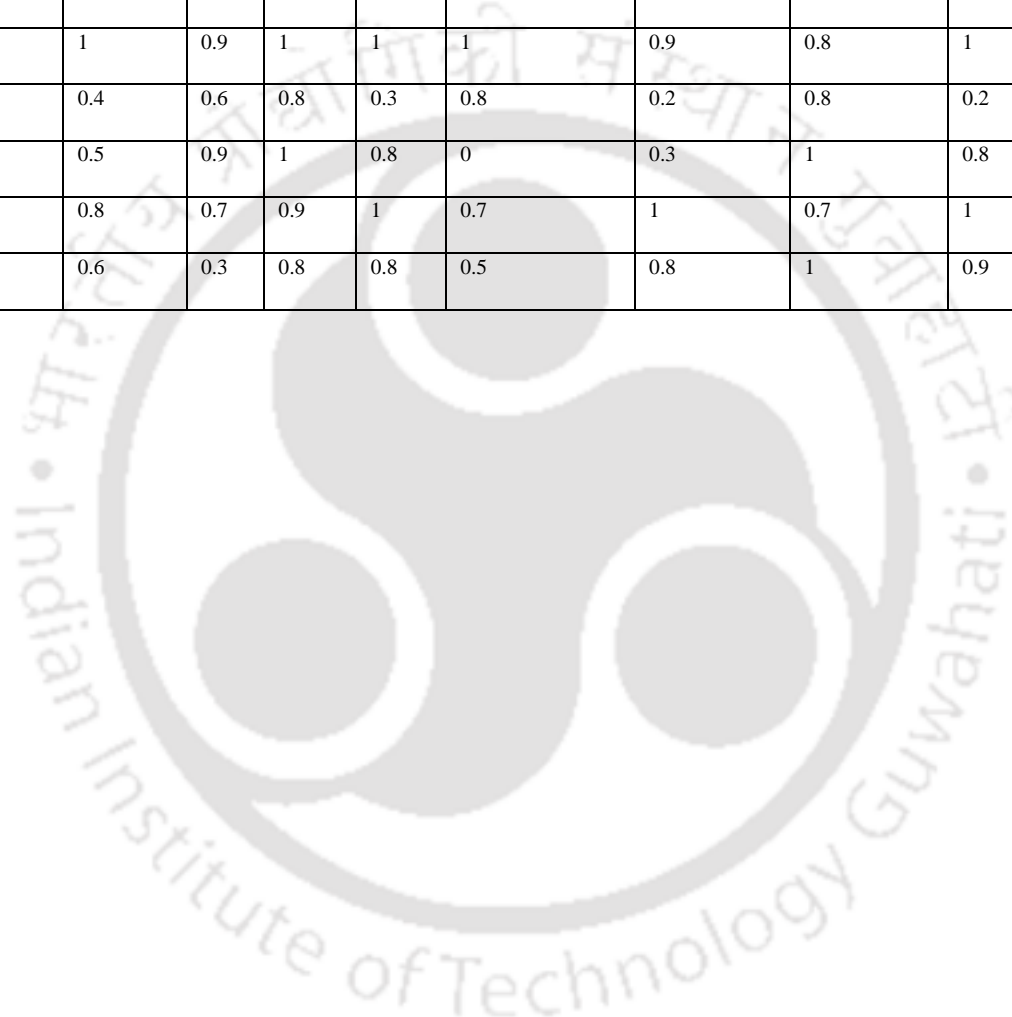
Subject-based data for the eye-tracking experiment with 14 signs and 41 participants is presented in Table E. 4

Table E. 4. Subject-based data for the eye-tracking experiment with 14 signs and 41 participants

Subject-based data for the eye-tracking experiment														
.No	COSACOEOT	DNT	DNWSH	NAFUP	UBC	WAFS	WRP	COSACOEOT-T	DNT-T	DNWSH-T	NAFUP-T	UBC-T	WAFS-T	WRP-T
1	0.6	0.7	0.6	0.6	0.6	0.9	0.6	0.6	0.5	0.6	0.6	0.6	0.9	0.7
2	0.4	0.9	1	0.9	0.8	0.8	0	0.8	1	1	1	0.8	0.9	0.6
3	0	0.1	0.2	1	0.5	1	0.5	0	0.4	0.6	0.6	0.3	0.9	0.7
4	0.7	0.9	0.9	0.8	0.9	0.8	0.3	0.7	1	1	0.9	1	1	0.3
5	0.5	0.6	0.8	0.3	0.7	0.8	0.3	0.9	0.9	0.9	0.9	1	1	0.9
6	0.7	0.3	0.5	0.8	0.7	0.2	0.5	0.8	0.8	0.7	0.7	0.8	0.2	0.6
7	0.7	0.8	0.7	0.6	0.7	0.8	0.6	0.7	0.8	0.7	0.8	0.7	0.8	0.6
8	1	0.9	1	1	0.9	1	0.9	1	1	1	1	1	1	1
9	0.9	0.9	1	0.9	0.8	0.9	0.9	0.9	0.8	1	1	1	1	1
10	0.8	0.9	0.7	0.3	0.9	0.3	0.8	0.9	0.9	0.8	0.9	0.9	0.9	0.9
11	0.6	0	0.8	0	0.1	0.8	0.8	0.8	0.5	1	0.8	0.4	0.8	0.8
12	0.1	0.8	0.8	0.1	0.5	0.7	0.1	0.7	0.8	0.8	0.8	0.8	0.7	0.8
13	0.8	0.9	0.9	0.6	0.9	0.5	0.3	0.9	0.9	0.8	0.9	0.9	0.9	0.8

14	0.3	0.8	0.8	0.3	0.3	0.5	0.3	1	1	1	1	1	1	1
15	0.4	0.8	0.7	0.3	0.7	1	0.9	0.6	1	1	0.8	1	1	0.9
16	0	0.9	0.8	0.6	0.1	0.6	1	0.6	1	1	1	0.5	1	1
17	0.2	0.8	0.7	0.7	0.8	0.8	0.8	0.8	0.9	1	0.9	1	0.9	1
18	0.6	0.3	0.6	0.7	0.5	0.6	0.6	0.9	1	0.9	0.9	0.5	0.9	0.8
19	0	0.3	1	0.3	1	1	1	1	1	1	1	1	1	1
20	0	0.2	1	0.6	0.4	0.3	0.8	0.8	0.7	1	0.9	0.9	0.5	0.9
21	0.3	0.9	0.9	0.9	0.9	0.9	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9
22	0	0.3	1	0.4	0.4	0.9	0.5	0.9	1	1	1	0.9	1	0.9
23	0	0.7	1	0.8	0.8	0.8	0.2	1	1	1	1	0.9	0.9	0.8
24	0.3	0.7	0.8	0.7	0.7	1	0.8	0.9	0.9	0.8	0.9	0.8	0.9	0.9
25	0	0.5	1	0.3	0.3	1	1	1	1	1	1	1	1	1
26	0	0.7	0.8	0.6	0.4	0.7	0.9	0.4	0.8	1	0.8	0.8	0.7	1
27	0.3	0.3	0.7	0.4	0.5	0	0.6	0.9	0.8	1	0.8	1	0.7	0.9
28	0.7	0.9	1	0.9	0.5	0.9	0.3	0.6	0.9	0.9	0.8	0.6	0.9	0.8
29	0.1	0.4	0.3	0.6	0.7	0.8	0.7	0.8	0.8	0.8	0.8	0.8	0.8	0.9
30	0.5	0.4	0.5	0.5	0.5	0.9	0.9	1	0.9	0.9	1	0.9	0.9	0.9
31	0.4	0.5	0.7	0.9	1	1	0.7	0.5	0.8	1	1	0.9	1	1
32	0.5	0.7	0.9	0.6	0.4	0.6	0.3	0.4	0.7	0.8	0.5	0.4	0.6	0.4
33	0.3	0.3	1	0.8	0.7	1	0	0.8	0.5	1	0.9	0.9	1	0.3
34	0.5	0.8	1	0.7	0.1	0.8	0.3	0.7	1	1	0.9	0.9	0.8	0.7

35	0	1	1	0.7	0.5	0.9	0	0.7	1	0.9	0.7	0.2	1	0.8
36	0.5	0.4	0.9	0.9	0.9	1	0.2	1	1	0.9	0.9	1	1	1
37	0.6	0.8	0.7	1	0.9	1	1	1	0.9	0.8	1	0.9	1	1
38	0.7	0.4	0.9	0.4	0.6	0.8	0.3	0.8	0.2	0.8	0.2	0.7	0.8	0.2
39	0	0.4	1	0.5	0.9	1	0.8	0	0.3	1	0.8	0.8	1	0.9
40	0.4	0.9	0.8	0.8	0.7	0.9	1	0.7	1	0.7	1	0.9	1	0.9
41	0.2	0.3	1	0.6	0.3	0.8	0.8	0.5	0.8	1	0.9	0.5	0.7	1



Appendix F: Eye-tracking data for the eye-tracking experiment

Eye tracking data for the eye-tracking experiment with 14 signs and 41 participants is presented in Table F. 5.

Table F. 5. Eye-tracking data for the eye-tracking experiment with 14 signs and 41 participants

Sign	Avg Pupil Diameter Right [mm]	Avg Pupil Diameter Left [mm]	Visual Intake Count	Visual Intake Frequency [count/s]	Visual Intake Duration Total [ms]	Visual Intake Duration Average [ms]	Visual Intake Duration Maximum [ms]	Visual Intake Duration Minimum [ms]	Visual Intake Dispersion Minimum [px]	Saccade Count
COSACOEOT	2.257684	2.291357	16.09756	1.609756	8755.878	1159.202	2547.878	564.9122	7.895122	13.80488
DNT	2.29137	2.332108	14.2439	1.42439	9088.332	1005.544	3040.98	363.2171	8.02439	12.43902
WRP	2.169133	2.215844	15.31707	1.531707	8479.398	752.9341	2619.705	145.7244	5.439024	13.46341
NAFUP	2.219713	2.28931	16.36585	1.636585	9021.954	767.0951	2457.393	176.0732	5.926829	14.31707
UBC	2.294989	2.287233	14.31707	1.431707	8704.871	952.0415	2623.62	393.495	6.89	12.56098
WAFS	2.191843	2.226164	17.60976	1.760976	8422.034	615.9415	2195.21	119.2927	5.463415	15.46341
COSACOEOT-T	2.217403	2.178909	27.02439	2.702439	7960.68	315.8683	1630.693	87.22927	3.809756	24.90244
UBC-T	2.076639	2.101576	21.85366	2.185366	7951.28	465.7707	1734.185	99.3878	4.480488	19
WRP-T	2.237742	2.212861	22.14634	2.214634	8425.771	447.3171	2045.024	97.78293	4.802439	20.14634

DNT-T	2.185874	2.181893	18.58537	1.858537	8304.346	540.5122	2231.724	101.8439	4.960976	15.90244
DNWSH	2.236398	2.267188	14.73171	1.473171	8948.278	991.622	2919.263	335.6073	7.221951	12.60976
NAFUP-T	2.148664	2.100601	22.39024	2.239024	8222.463	419.978	1983.766	95.76585	4.443902	20.12195
WAFS-T	2.122327	2.142578	19.73171	1.973171	8015.256	587.2293	1969.961	136.3512	5.358537	17.53659
DNWSH-T	2.196645	2.166237	21.58537	2.158537	7891.771	435.5366	1708.31	99.42195	4.719512	18.95122
Sign	Saccade Frequency [count/s]	Saccade Duration Total [ms]	Saccade Duration Average [ms]	Saccade Duration Maximum [ms]	Saccade Duration Minimum [ms]	Saccade Amplitude Total [°]	Saccade Amplitude Average [°]	Saccade Amplitude Maximum [°]	Saccade Amplitude Minimum [°]	Saccade Velocity Total [°/s]
COSACOEOT	1.380488	783.0756	57.61	124.395	35.23	30.69756	2.1875	9.4025	0.5875	898.8512
DNT	1.243902	699.3439	59.165	124.8125	39.43	46.7878	3.8175	30.685	0.6	1077.995
WRP	1.346341	750.7341	56.22195	99.48293	35.23659	69.43659	5.209756	32.30976	0.534146	1148.715
NAFUP	1.431707	765.422	52.53659	92.58049	35.23659	25.2878	1.790244	5.814634	0.639024	676.0683
UBC	1.256098	696.0683	54.625	104.49	35.6175	25.89756	2.09	6.2075	0.595	961.4707
WAFS	1.546341	851.0902	55.93415	119.3439	34.03171	87.4	6.546341	34.36829	0.487805	1594.61
COSACOEOT-T	2.490244	1631.771	65.91707	169.661	35.64878	118.2756	4.807317	30.06341	0.526829	2697.371
UBC-T	1.9	1190.359	66.76829	156.2488	35.24634	58.76098	3.192683	12.55854	0.592683	1397.71
WRP-T	2.014634	1289.349	68.01463	189.122	35.63659	74.37317	3.843902	16.0439	0.626829	1268.81

DNT-T	1.590244	1023.956	66.72439	153.8537	36.05122	45.05854	2.802439	12.11463	0.497561	1239.634
DNWSH	1.260976	681.9171	55.3725	101.94	37.345	23.12683	1.835	5.555	0.6525	793.7073
NAFUP-T	2.012195	1291.42	65.5439	166.4122	37.26341	77.07561	4.509756	24.84146	0.609756	1671.498
WAFS-T	1.753659	1079.751	59.02683	149.7805	33.59512	51.62439	2.782927	13.20976	0.529268	1205.705
DNWSH-T	1.895122	1288.2	69.9561	173.3122	38.09756	136.5146	7.326829	43.9439	0.617073	2029.544
Sign	Saccade Velocity Total [°/s]	Saccade Velocity Average [°/s]		Saccade Velocity Maximum [°/s]		Saccade Velocity Minimum [°/s]	Saccade Latency Average [ms]	Blink Count	Blink Frequency [count/s]	
COSACOEOT	898.8512	65.62		313.7275		17.3775	722.2732	2.365854	0.236585	
DNT	1077.995	116.85		326.57		45.8575	876.4659	1.682927	0.168293	
WRP	1148.715	95.33659		369.861		17.08049	838.7683	1.853659	0.185366	
NAFUP	676.0683	42.40732		183.0854		17.35366	743.6317	2.04878	0.204878	
UBC	961.4707	79.5325		377.4025		16.8475	766.222	1.707317	0.170732	
WAFS	1594.61	121.2976		440.1732		17.11707	684.8512	2.097561	0.209756	
COSACOEOT-T	2697.371	108.778		833.0854		16.37317	328.2732	1.853659	0.185366	
UBC-T	1397.71	126.4244		417.0951		22.52439	491.8561	2.560976	0.256098	
WRP-T	1268.81	68.52927		337.6317		18.6439	462.7098	1.731707	0.173171	

DNT-T	1239.634	104.3463	477.7463	17.30488	573.9585	2.439024	0.243902
DNWSH	793.7073	86.4825	290.8275	18.795	822.1951	2.097561	0.209756
NAFUP-T	1671.498	99.08537	525.5537	17.58049	442.4854	2.146341	0.214634
WAFS-T	1205.705	75.5561	343.7659	17.37073	494.1073	1.926829	0.192683
DNWSH-T	2029.544	107.9732	605.0171	17.83902	490.3683	2.390244	0.239024
Sign	Blink Duration Total [ms]	Blink Duration Average [ms]	Blink Duration Maximum [ms]	Blink Duration Minimum [ms]		Scanpath Length [px]	
COSACOEOT	731.3293	282.9423	387.1346	216.2808		686.3902	
DNT	552.3561	291.3435	413.8478	226.4783		1144.683	
WRP	617.2366	199.5098	279.6098	153.7927		982.4146	
NAFUP	561.6805	256.2292	326.5792	199.6792		573.3171	
UBC	812.4805	496.0286	766.2095	326.4571		1006.659	
WAFS	669.2366	173.3829	231.7488	129.0366		1619.049	
COSACOEOT-T	580.7707	293.264	398.688	226.272		2163.61	
UBC-T	1104.678	311.0585	520.2732	202.5098		1640.146	
WRP-T	698.3805	235.3951	367.2756	126.2049		1422.488	

DNT-T	883.9073	308.0567	422.09	227.3733	1324.268	
DNWSH	781.3122	313.65	511.3769	215.1538	549.1951	
NAFUP-T	1100.988	398.9296	950.263	240.337	1426.585	
WAFS-T	756.4854	230.8976	346.9854	174.5415	1274.537	
DNWSH-T	929.3268	371.7	613.8111	208.9111	2094.146	
AOI 004- Outerbox						
Sign	Time to First Appearance [ms]	First Visual Intake Duration [ms]	Revisits	Visual Intake Count	Visual Intake Time [ms]	Average Visual Intake Duration [ms]
COSACOEOT	89135.67561	889.6292683	0.12195122	15.90243902	8657.25122	1158.27561
DNT	98499.35854	661.9170732	0.097560976	14.14634146	9068.44878	1005.719512
WRP	94819.12195	1038.590244	0.073170732	15.24390244	8469.646341	753.6439024
NAFUP	90153.99024	431.3878049	0.05	16.24390244	8646.55122	645.6292683
UBC	90463.63659	645.2634146	0.15	14.12195122	8671.2	954.0756098
WAFS	92455.08537	429.804878	0.097560976	17.41463415	8383.878049	615.7829268
COSACOEOT-T	263350.7268	244.3292683	0.219512195	26.41463415	7841.797561	318.195122
UBC-T	254446.5098	221.997561	0.268292683	21.48780488	7870.129268	466.3317073
WRP-T	254084.8805	257.505	0.15	21.375	8327.7325	454.9825

DNT-T	249410.7927	330.7658537	0.317073171	18.26829268	8236.560976	542.9292683
DNWSH	96543.13171	643.2609756	0.048780488	14.65853659	8925.543902	993.6585366
NAFUP-T	264794.3293	230.5170732	0.170731707	22	8155.070732	424.7317073
WAFS-T	262112.161	236.1853659	0.025	19.65853659	8008.35122	587.8463415
DNWSH-T	254104.0341	260.5878049	0.12195122	21.19512195	7725.763415	436.0804878
AOI 005- Image Circle						
Sign	Time to First Appearance [ms]	First Visual Intake Duration [ms]	Revisits	Visual Intake Count	Visual Intake Time [ms]	Average Visual Intake Duration [ms]
COSACOEOT	89135.67561	884.3487805	0.268292683	15.41463415	8488.421951	1164.043902
DNT	98499.35854	661.9195122	0.170731707	13.80487805	9008.387805	1006.085366
WRP	94819.12195	1038.995122	0.365853659	14.68292683	8161.197561	754.3219512
NAFUP	90153.99024	430.9756098	0.125	16.04878049	8479.343902	641.2926829
UBC	90463.63659	642.4219512	0.375	13.63414634	8520.221951	964.8073171
WAFS	92455.08537	450.0926829	0.365853659	16.97560976	8200.039024	624.3487805
COSACOEOT-T	263350.7268	323.0268293	1.4	8.195121951	3742.52439	533.5121951
UBC-T	254446.5098	254.0658537	1.85	13.43902439	5645.360976	532.4853659

WRP-T	254084.8805	400.5731707	1.487179487	11.95121951	5335.612195	603.8195122
DNT-T	249410.7927	449.6829268	1.487179487	12.65853659	6495.821951	666.4853659
DNWSH	96543.13171	643.2609756	0.170731707	14.12195122	8646.331707	1001.704878
NAFUP-T	264794.3293	239.0341463	1.9	12.34146341	5819.841463	554.3878049
WAFS-T	262112.161	272.302439	1.552631579	12.63414634	6027.817073	637.0926829
DNWSH-T	254104.0341	678.3073171	1.634146341	11.31707317	5168.926829	757.7878049
AOI 006- Text box						
Sign	Time to First Appearance [ms]	First Visual Intake Duration [ms]	Revisits	Visual Intake Count	Visual Intake Time [ms]	Average Visual Intake Duration [ms]
COSACOEOT	89135.68	31.66098	0.097561	0.243902	98.62927	30.33415
DNT	98499.36	6.904878	0.073171	0.170732	25.98049	7.712195
WRP	94819.12	27.19512	0.02439	0.170732	77.92195	31.17317
NAFUP	90153.99	0	0	0	0	0
UBC	90463.64	0	0	0	0	0
WAFS	92455.09	19.88293	0	0.02439	19.88293	19.88293
COSACOEOT-T	263350.7	216.7366	1.780488	14.90244	3337.473	221.7146

UBC-T	254446.5	206.5585	1.075	6.390244	1812.829	298.4244
WRP-T	254084.9	214.678	0.756098	6.926829	2351.478	248.3756
DNT-T	249410.8	185.5	1.025	4.170732	1249.249	239.161
DNWSH	96543.13	32.46341	0.02439	0.292683	73.44878	33.27561
NAFUP-T	264794.3	180.6122	1.365854	7.853659	1954.92	233.739
WAFS-T	262112.2	152.2268	1.085714	5.585366	1578.771	236.7024
DNWSH-T	254104	209.0171	1.146341	8.243902	1992.278	265.7659

Appendix G: Bivariate correlation results (Pearson's r, Kendall's tau_b, and Spearman's rho) for the subject-based data versus eye-tracking data for the significant signs or the significant eye measures in the eye-tracking experiment

The following tables (Table G. 6, Table G. 7) show bivariate correlation results for the subject-based data versus eye-tracking data for the significant signs or the significant eye measures in the eye-tracking experiment.

Table G. 6. Bivariate correlation results (Parametric- Pearson correlation) for the subject-based data versus eye-tracking data for the significant signs in the eye-tracking experiment

Bivariate correlation results (Parametric- Pearson correlation) for the subject-based data versus eye-tracking data for the significant signs in the eye-tracking experiment										
Parametric Correlations		Subjective scores of signs								
Eye-tracking measures	N=41	NAFUP	WAFS	COSACOEOT	WRP	DNT	NAFUP_T	WAFS_T	COSACOEOT_T	DNT_T
AvgpupilDR	Pearson Correlation	0.158	0.187	0.255	0.105	0.083	-0.034	0.102	-0.003	0.030
	Sig. (2-tailed)	0.324	0.241	0.108	0.515	0.608	0.833	0.526	0.987	0.852
AvgpupilDL	Pearson Correlation	0.182	0.195	0.205	0.028	0.076	-0.118	0.049	-0.101	0.012

	Sig. (2-tailed)	0.254	0.222	0.198	0.861	0.637	0.464	0.763	0.530	0.940
vis_intk_cnt	Pearson Correlation	-0.280	0.269	-.368*	-0.207	-0.141	-0.198	-0.007	-0.003	0.065
	Sig. (2-tailed)	0.076	0.090	0.018	0.195	0.380	0.214	0.967	0.987	0.685
vis_intk_freq	Pearson Correlation	-0.280	0.269	-.368*	-0.207	-0.141	-0.198	-0.007	-0.003	0.065
	Sig. (2-tailed)	0.076	0.090	0.018	0.195	0.380	0.214	0.967	0.987	0.685
Vis_intk_durtot	Pearson Correlation	0.299	.370*	0.180	0.219	-0.005	0.263	0.106	0.124	0.207
	Sig. (2-tailed)	0.057	0.017	0.260	0.170	0.976	0.097	0.511	0.441	0.194
Vis_intk_durAvg	Pearson Correlation	0.300	-0.126	0.296	0.289	0.062	0.233	0.080	0.009	0.051
	Sig. (2-tailed)	0.057	0.434	0.061	0.067	0.702	0.142	0.620	0.957	0.751
Vis_intk_durMax	Pearson Correlation	0.245	-0.028	.330*	.322*	0.023	0.262	0.109	0.174	0.025
	Sig. (2-tailed)	0.123	0.862	0.035	0.040	0.885	0.098	0.497	0.276	0.877
Vis_intk_durMin	Pearson Correlation	.320*	0.107	0.240	0.090	0.051	0.172	-0.001	0.015	-0.105
	Sig. (2-tailed)	0.041	0.506	0.130	0.574	0.749	0.282	0.994	0.925	0.514
Sac_cnt	Pearson Correlation	-0.194	.343*	-.395*	-0.238	-0.098	-0.052	0.029	0.101	0.185
	Sig. (2-tailed)	0.223	0.028	0.011	0.134	0.540	0.746	0.860	0.532	0.246
Sac_freq	Pearson Correlation	-0.194	.343*	-.395*	-0.238	-0.098	-0.052	0.029	0.101	0.185
	Sig. (2-tailed)	0.223	0.028	0.011	0.134	0.540	0.746	0.860	0.532	0.246
Sac_durtot	Pearson Correlation	-0.247	0.174	-.447**	-0.131	-0.195	0.077	0.025	0.069	0.232
	Sig. (2-tailed)	0.119	0.276	0.003	0.416	0.223	0.634	0.874	0.667	0.144
Sac_durAvg	Pearson Correlation	-0.249	-0.225	-0.139	0.178	-.391*	0.181	-0.111	-0.044	0.007
	Sig. (2-tailed)	0.116	0.157	0.387	0.265	0.011	0.258	0.488	0.783	0.963
Sac_durMax	Pearson Correlation	-0.201	-0.011	-0.220	0.237	-0.232	0.254	0.076	-0.064	0.074
	Sig. (2-tailed)	0.208	0.945	0.167	0.135	0.145	0.109	0.636	0.692	0.645

Sac_durMin	Pearson Correlation	0.182	0.016	-0.239	-0.006	-.333*	0.077	0.005	-0.059	-0.131
	Sig. (2-tailed)	0.255	0.920	0.132	0.971	0.033	0.631	0.974	0.715	0.414
Sac_ampltot	Pearson Correlation	-0.077	-0.254	-.345*	0.129	-0.242	0.067	0.123	0.100	0.071
	Sig. (2-tailed)	0.633	0.109	0.027	0.422	0.128	0.678	0.442	0.534	0.660
Sac_amplAvg	Pearson Correlation	0.162	-0.295	-0.265	0.130	-0.249	0.056	0.115	0.059	-0.137
	Sig. (2-tailed)	0.312	0.061	0.095	0.419	0.116	0.727	0.473	0.716	0.393
Sac_amplMax	Pearson Correlation	0.109	-0.267	-0.094	0.126	-0.243	0.057	0.230	0.045	-0.164
	Sig. (2-tailed)	0.498	0.092	0.561	0.433	0.125	0.723	0.149	0.778	0.305
Sac_amplMin	Pearson Correlation	0.152	0.260	-0.003	0.022	-0.192	-0.242	-0.022	-.389*	-0.038
	Sig. (2-tailed)	0.344	0.100	0.984	0.893	0.228	0.128	0.892	0.012	0.813
Sac_veltot	Pearson Correlation	-0.108	-0.271	-0.253	0.099	-0.283	0.075	-0.300	0.071	0.014
	Sig. (2-tailed)	0.501	0.086	0.110	0.540	0.073	0.642	0.057	0.659	0.932
Sac_velAvg	Pearson Correlation	-0.078	-0.306	-0.030	0.123	-.412**	0.074	-.348*	0.067	-0.079
	Sig. (2-tailed)	0.627	0.052	0.852	0.443	0.007	0.647	0.026	0.676	0.623
Sac_velMax	Pearson Correlation	-0.109	-0.211	-0.123	0.131	-.315*	0.120	-0.255	0.080	-0.045
	Sig. (2-tailed)	0.496	0.186	0.442	0.416	0.045	0.454	0.108	0.619	0.778
Sac_velMin	Pearson Correlation	0.180	-.317*	0.018	-0.155	-.359*	-0.059	-0.068	-0.207	0.042
	Sig. (2-tailed)	0.260	0.044	0.911	0.334	0.021	0.714	0.672	0.195	0.795
Sac_latAvg	Pearson Correlation	0.142	-0.261	.309*	0.012	0.233	0.100	0.054	-0.199	-0.066
	Sig. (2-tailed)	0.375	0.100	0.049	0.941	0.143	0.533	0.737	0.213	0.683
Blink_cnt	Pearson Correlation	-0.205	-0.171	-0.020	0.032	-0.114	-.345*	-0.131	-0.241	-.367*
	Sig. (2-tailed)	0.199	0.285	0.899	0.842	0.477	0.027	0.413	0.129	0.018
Blink_freq	Pearson Correlation	-0.205	-0.171	-0.020	0.032	-0.114	-.345*	-0.131	-0.241	-.367*

	Sig. (2-tailed)	0.199	0.285	0.899	0.842	0.477	0.027	0.413	0.129	0.018
Blink_dur	Pearson Correlation	-0.166	-0.300	0.115	-0.009	-0.061	-.316*	-0.176	-0.212	-0.288
	Sig. (2-tailed)	0.300	0.056	0.475	0.954	0.704	0.044	0.270	0.183	0.068
Blink_durAvg	Pearson Correlation	-0.105	-.475**	0.153	0.251	0.150	-0.261	-0.126	-0.121	-0.219
	Sig. (2-tailed)	0.512	0.002	0.339	0.113	0.348	0.099	0.434	0.451	0.169
Blink_durMax	Pearson Correlation	-0.092	-.399**	0.184	0.080	0.078	-0.280	-0.298	-0.133	-0.268
	Sig. (2-tailed)	0.567	0.010	0.248	0.620	0.629	0.076	0.058	0.407	0.091
Blink_durMin	Pearson Correlation	-0.124	-.419**	0.093	.412**	0.196	0.007	0.001	-0.046	-0.205
	Sig. (2-tailed)	0.439	0.006	0.562	0.007	0.218	0.963	0.993	0.775	0.198
Scanpath_len	Pearson Correlation	-0.201	-0.275	-.426**	0.080	-0.237	-0.130	-.330*	-0.045	-0.055
	Sig. (2-tailed)	0.208	0.082	0.006	0.617	0.135	0.419	0.035	0.781	0.731
	N	41	41	41	41	41	41	41	41	41

Table G. 7. Bivariate correlation results (Parametric- Pearson correlation) for the subject-based data versus eye-tracking data for the significant signs in the eye-tracking experiment (AOI-based)

Bivariate correlation results (Parametric- Pearson correlation) for the subject-based data versus eye-tracking data for the significant signs in the eye-tracking experiment (AOI-based)						
Eye-tracking measures (AOI based)	N=41	Subjective scores of signs				
Parametric Test		DNWSH	WAFS	COSACOEOT	WRP	WAFS_T
AOI_size004First_vis_intkdur	Pearson Correlation	0.069	-0.044	0.230	.371*	0.047
	Sig. (2-tailed)	0.670	0.784	0.149	0.017	0.769
AOI_size004_Revisits	Pearson Correlation	0.216	-0.259	-.352*	0.108	-.363*
	Sig. (2-tailed)	0.175	0.103	0.024	0.503	0.020
AOI_size004_Vis_intkCnt	Pearson Correlation	-0.117	.313*	-.349*	-0.213	0.005
	Sig. (2-tailed)	0.467	0.046	0.026	0.180	0.977
AOI_size004_Vis_intkTime	Pearson Correlation	-0.060	.379*	0.179	0.210	0.109
	Sig. (2-tailed)	0.711	0.015	0.261	0.187	0.497
AOI_size004_Avg_vis_intkdur	Pearson Correlation	0.049	-0.125	0.294	0.290	0.079
	Sig. (2-tailed)	0.759	0.437	0.062	0.066	0.623
AOI_size005First_vis_intkdur	Pearson Correlation	0.069	-0.019	0.229	.371*	-0.006
	Sig. (2-tailed)	0.670	0.908	0.151	0.017	0.970
AOI_size005_Revisits	Pearson Correlation	0.051	-0.105	-0.212	0.115	-0.007
	Sig. (2-tailed)	0.753	0.514	0.183	0.476	0.964
AOI_size005_Vis_intkCnt	Pearson Correlation	-0.062	0.280	-.326*	-0.245	0.113
	Sig. (2-tailed)	0.698	0.076	0.037	0.123	0.481
AOI_size005_Vis_intkTime	Pearson Correlation	0.012	0.308	0.199	0.092	0.029
	Sig. (2-tailed)	0.943	0.050	0.212	0.567	0.855
AOI_size005_Avg_vis_intkdur	Pearson Correlation	0.060	-0.109	0.295	0.278	0.041
	Sig. (2-tailed)	0.708	0.497	0.061	0.079	0.801
AOI_size006First_vis_intkdur	Pearson Correlation	-0.273	0.079	-0.044	0.154	-0.110
	Sig. (2-tailed)	0.084	0.622	0.783	0.338	0.494
AOI_size006_Revisits	Pearson Correlation	-0.256	. ^b	-0.177	0.108	-0.035
	Sig. (2-tailed)	0.107		0.269	0.503	0.830
AOI_size006_Vis_intkCnt	Pearson Correlation	-0.265	0.079	-0.139	0.097	-0.085
	Sig. (2-tailed)	0.094	0.622	0.386	0.546	0.596
AOI_size006_Vis_intkTime	Pearson Correlation	-.345*	0.079	-0.095	0.136	0.100
	Sig. (2-tailed)	0.027	0.622	0.554	0.395	0.534
AOI_size006_Avg_vis_intkdur	Pearson Correlation	-0.280	0.079	-0.042	0.160	0.103
	Sig. (2-tailed)	0.077	0.622	0.796	0.317	0.521
TimeToFirstAppear_allAOI	Pearson Correlation	-0.123	-0.042	0.188	-.356*	0.196
	Sig. (2-tailed)	0.443	0.796	0.239	0.022	0.220

** . Correlation is significant at the 0.01 level (2-tailed).
* . Correlation is significant at the 0.05 level (2-tailed).

Table G. 8. Bivariate correlation- Non-parametric (Kendall's tau_b) results for significant measures of eye-tracking or significant signs only

Bivariate correlation- Non-parametric (Kendall's tau_b) results for significant measures of eye-tracking or significant signs only												
Eye-tracking measures	N=41	NAFUP	WAFS	COSACOEOT	WRP	DNWSH-T	NAFUP_T	WAFS_T	COSACOEOT_T	UBC_T	WRP_T	DNT_T
vis_intk_cnt	Kendall's tau_b	-0.215	.237*	-.277*	-0.199	0.003	-0.187	0.106	-0.016	0.073	-0.157	0.040
	Sig. (2-tailed)	0.065	0.047	0.017	0.085	0.980	0.124	0.385	0.891	0.540	0.189	0.743
vis_intk_freq	Kendall's tau_b	-0.215	.237*	-.277*	-0.199	0.003	-0.187	0.106	-0.016	0.073	-0.157	0.040
	Sig. (2-tailed)	0.065	0.047	0.017	0.085	0.980	0.124	0.385	0.891	0.540	0.189	0.743
Vis_intk_durtot	Kendall's tau_b	0.206	0.129	0.094	.246*	0.093	0.198	0.063	0.084	0.062	0.212	0.145
	Sig. (2-tailed)	0.071	0.268	0.406	0.031	0.442	0.096	0.602	0.465	0.595	0.071	0.220
Vis_intk_durAvg	Kendall's tau_b	.246*	-0.059	0.204	0.199	0.019	.260*	0.014	0.084	-0.065	0.204	0.072
	Sig. (2-tailed)	0.031	0.612	0.072	0.080	0.874	0.029	0.906	0.465	0.579	0.082	0.544
Vis_intk_durMin	Kendall's tau_b	.329**	0.027	0.209	-0.011	0.053	0.088	-0.201	0.031	0.029	.304*	-0.126
	Sig. (2-tailed)	0.004	0.817	0.067	0.927	0.668	0.478	0.102	0.799	0.808	0.012	0.301
Sac_cnt	Kendall's tau_b	-0.123	.267*	-.329**	-0.204	0.063	-0.059	0.124	0.052	0.092	-0.079	0.175
	Sig. (2-tailed)	0.284	0.024	0.004	0.078	0.608	0.630	0.311	0.655	0.438	0.508	0.147
Sac_freq	Kendall's tau_b	-0.123	.267*	-.329**	-0.204	0.063	-0.059	0.124	0.052	0.092	-0.079	0.175
	Sig. (2-tailed)	0.284	0.024	0.004	0.078	0.608	0.630	0.311	0.655	0.438	0.508	0.147
Sac_durtot	Kendall's tau_b	-0.144	0.210	-.328**	-0.151	0.150	0.001	0.071	0.033	0.063	-0.071	0.170
	Sig. (2-tailed)	0.207	0.072	0.004	0.184	0.218	0.991	0.553	0.775	0.587	0.547	0.151
Sac_durAvg	Kendall's tau_b	-0.147	0.112	-0.111	0.113	0.171	.271*	-0.053	-0.166	0.038	-0.113	0.197
	Sig. (2-tailed)	0.199	0.338	0.328	0.323	0.160	0.024	0.660	0.150	0.746	0.336	0.097
Sac_durMax	Kendall's tau_b	-.237*	0.156	-0.172	0.165	.248*	.245*	0.063	-0.075	0.089	-0.098	0.161
	Sig. (2-tailed)	0.041	0.184	0.135	0.154	0.044	0.041	0.601	0.522	0.452	0.404	0.179

Sac_ampltot	Kendall's tau_b	-0.116	.238*	-.357**	-0.093	0.061	-0.147	0.137	-0.007	0.067	-0.193	0.145
	Sig. (2-tailed)	0.311	0.041	0.002	0.413	0.617	0.219	0.254	0.954	0.564	0.100	0.220
Sac_amplMax	Kendall's tau_b	-0.208	0.202	-0.213	0.000	0.094	-0.071	.245*	-0.020	0.160	-0.102	0.064
	Sig. (2-tailed)	0.070	0.085	0.062	1.000	0.442	0.550	0.042	0.864	0.173	0.385	0.591
Sac_amplMin	Kendall's tau_b	0.042	.245*	-0.014	0.083	0.110	-.288*	-0.193	-.290*	-.250*	-0.091	0.072
	Sig. (2-tailed)	0.722	0.046	0.908	0.489	0.383	0.020	0.124	0.016	0.040	0.455	0.563
Sac_veltot	Kendall's tau_b	-0.214	.253*	-0.184	-0.072	0.203	0.106	0.023	0.147	0.067	-0.090	0.151
	Sig. (2-tailed)	0.061	0.030	0.106	0.525	0.095	0.374	0.849	0.201	0.564	0.444	0.203
Sac_velMax	Kendall's tau_b	-0.197	0.226	-0.121	0.138	0.226	.249*	0.105	0.122	0.016	-0.060	0.084
	Sig. (2-tailed)	0.084	0.053	0.285	0.224	0.062	0.037	0.380	0.288	0.890	0.610	0.476
Blink_cnt	Kendall's tau_b	-0.145	-0.196	-0.019	0.073	-0.092	-0.218	0.023	-0.093	0.032	-0.053	-.345**
	Sig. (2-tailed)	0.240	0.119	0.878	0.560	0.476	0.092	0.862	0.460	0.795	0.683	0.006
Blink_freq	Kendall's tau_b	-0.145	-0.196	-0.019	0.073	-0.092	-0.218	0.023	-0.093	0.032	-0.053	-.345**
	Sig. (2-tailed)	0.240	0.119	0.878	0.560	0.476	0.092	0.862	0.460	0.795	0.683	0.006
Blink_dur	Kendall's tau_b	-0.143	-0.219	0.025	0.124	-0.047	-0.188	-0.020	-0.103	-0.020	-0.053	-.311*
	Sig. (2-tailed)	0.233	0.072	0.834	0.293	0.709	0.129	0.874	0.396	0.870	0.675	0.010
Blink_durAvg	Kendall's tau_b	-0.124	-.318**	0.117	.246*	0.009	-0.105	-0.014	-0.087	-0.105	0.011	-0.228
	Sig. (2-tailed)	0.304	0.009	0.327	0.036	0.940	0.396	0.912	0.472	0.380	0.931	0.061
Blink_durMax	Kendall's tau_b	-0.122	-.255*	0.067	0.197	-0.016	-0.157	-0.045	-0.081	-0.082	-0.008	-0.237
	Sig. (2-tailed)	0.310	0.036	0.575	0.096	0.901	0.205	0.723	0.502	0.497	0.951	0.051
Blink_durMin	Kendall's tau_b	-0.115	-.305*	0.133	.347**	0.025	0.028	0.082	-0.001	-0.117	0.123	-0.204
	Sig. (2-tailed)	0.344	0.013	0.262	0.003	0.842	0.820	0.516	0.991	0.331	0.330	0.094
Scanpath_len	Kendall's tau_b	-0.218	.236*	-.302**	-0.186	0.049	-0.156	-0.063	-0.039	0.051	-0.122	0.101

	Sig. (2-tailed)	0.056	0.042	0.008	0.102	0.687	0.190	0.602	0.732	0.661	0.297	0.394
**. Correlation is significant at the 0.01 level (2-tailed).												
*. Correlation is significant at the 0.05 level (2-tailed).												

Table G. 9. Bivariate correlation- Non-parametric (Kendall's tau_b) results for significant measures of eye-tracking or significant signs only (AOI-based)

Bivariate correlation- Non-parametric (Kendall's tau_b) results for significant measures of eye-tracking or significant signs only (AOI-based)									
		WAFS	COSACOEOT	WRP	DNWSH-T	NAFUP_T	COSACOEOT_T	WRP_T	DNT_T
AOI_size004First_vis_intkdur	Kendall's tau_b	0.061	0.135	0.201	.266*	.282*	0.019	0.096	-0.067
	Sig. (2-tailed)	0.604	0.237	0.078	0.030	0.019	0.873	0.417	0.575
AOI_size004_Revisits	Kendall's tau_b	-0.026	-.320*	0.076	0.169	0.268	0.091	-0.067	0.080
	Sig. (2-tailed)	0.853	0.019	0.579	0.244	0.059	0.504	0.635	0.567
AOI_size004_Vis_intkCnt	Kendall's tau_b	.265*	-.261*	-0.208	-0.015	-0.205	-0.004	-0.082	0.031
	Sig. (2-tailed)	0.026	0.024	0.072	0.903	0.091	0.973	0.493	0.797
AOI_size004_Vis_intkTime	Kendall's tau_b	0.121	0.102	.246*	0.084	0.179	0.082	.231*	0.123
	Sig. (2-tailed)	0.299	0.369	0.031	0.487	0.134	0.479	0.049	0.299
AOI_size004_Avg_vis_intkdur	Kendall's tau_b	-0.059	0.191	0.202	0.034	.265*	0.063	0.201	0.080
	Sig. (2-tailed)	0.612	0.092	0.076	0.779	0.026	0.583	0.086	0.498
AOI_size005First_vis_intkdur	Kendall's tau_b	0.055	0.130	0.202	-0.036	0.161	0.005	0.023	-0.051
	Sig. (2-tailed)	0.636	0.255	0.076	0.770	0.178	0.964	0.844	0.665
AOI_size005_Revisits	Kendall's tau_b	0.053	-0.122	0.004	0.070	-0.154	-0.099	0.102	0.041
	Sig. (2-tailed)	0.700	0.363	0.974	0.598	0.246	0.439	0.428	0.754
AOI_size005_Vis_intkCnt	Kendall's tau_b	.234*	-0.221	-0.227	-0.014	-0.109	-0.125	-0.074	0.024
	Sig. (2-tailed)	0.049	0.057	0.051	0.912	0.372	0.287	0.538	0.842

AOI_size005_Vis_intkTime	Kendall's tau_b	0.078	0.113	0.191	-0.007	0.134	-0.103	0.076	0.011
	Sig. (2-tailed)	0.504	0.322	0.093	0.951	0.261	0.373	0.516	0.926
AOI_size005_Avg_vis_intkdur	Kendall's tau_b	-0.056	0.202	0.194	0.040	0.161	-0.011	0.136	0.030
	Sig. (2-tailed)	0.628	0.076	0.088	0.742	0.178	0.927	0.246	0.797
AOI_size006First_vis_intkdur	Kendall's tau_b	0.067	-0.054	0.038	0.021	-0.143	0.056	0.018	0.170
	Sig. (2-tailed)	0.634	0.690	0.781	0.864	0.232	0.631	0.880	0.157
AOI_size006_Revisits	Kendall's tau_b		-0.092	0.076	0.199	-0.032	-0.069	0.054	0.044
	Sig. (2-tailed)		0.501	0.579	0.139	0.807	0.581	0.682	0.742
AOI_size006_Vis_intkCnt	Kendall's tau_b	0.067	-0.063	0.027	0.067	0.000	0.222	0.147	-0.003
	Sig. (2-tailed)	0.634	0.642	0.840	0.589	1.000	0.059	0.222	0.981
AOI_size006_Vis_intkTime	Kendall's tau_b	0.067	-0.054	0.031	0.025	0.045	.237*	0.113	0.123
	Sig. (2-tailed)	0.634	0.690	0.820	0.836	0.708	0.040	0.336	0.303
AOI_size006_Avg_vis_intkdur	Kendall's tau_b	0.067	-0.054	0.038	0.158	0.067	0.109	0.086	.255*
	Sig. (2-tailed)	0.634	0.690	0.781	0.192	0.574	0.343	0.465	0.032
TimeToFirstAppear_allAOI	Kendall's tau_b	0.048	0.153	-0.209	0.007	-0.126	-0.095	0.060	0.090
	Sig. (2-tailed)	0.678	0.180	0.066	0.951	0.292	0.411	0.610	0.448
<p>** . Correlation is significant at the 0.01 level (2-tailed).</p> <p>* . Correlation is significant at the 0.05 level (2-tailed).</p>									

Table G. 10. Bivariate correlation- Non-parametric (Spearman's rho) results for significant measures of eye-tracking or significant signs only

Bivariate correlation- Non-parametric (Spearman's rho) results for significant measures of eye-tracking or significant signs only										
		DNWSH	WAFS	COSACOEOT	WRP	NAFUP_T	WAFS_T	COSACOEOT_T	WRP_T	DNT_T
AvgpupilDR	Spearman's rho	-0.013	0.023	0.176	0.005	-0.038	-0.042	0.022	-0.129	0.028
	Sig. (2-tailed)	0.937	0.888	0.272	0.975	0.814	0.795	0.892	0.422	0.860
AvgpupilDL	Spearman's rho	-0.043	0.031	0.174	0.057	-0.154	-0.151	-0.058	-0.177	-0.009
	Sig. (2-tailed)	0.790	0.848	0.277	0.724	0.337	0.347	0.718	0.267	0.957
vis_intk_cnt	Spearman's rho	-0.103	.317*	-.364*	-0.284	-0.251	0.137	-0.020	-0.205	0.062
	Sig. (2-tailed)	0.520	0.044	0.019	0.072	0.113	0.394	0.900	0.199	0.702
vis_intk_freq	Spearman's rho	-0.103	.317*	-.364*	-0.284	-0.251	0.137	-0.020	-0.205	0.062
	Sig. (2-tailed)	0.520	0.044	0.019	0.072	0.113	0.394	0.900	0.199	0.702
Vis_intk_durtot	Spearman's rho	0.222	0.177	0.151	.354*	0.269	0.082	0.122	0.284	0.195
	Sig. (2-tailed)	0.163	0.267	0.345	0.023	0.090	0.609	0.448	0.072	0.222
Vis_intk_durAvg	Spearman's rho	0.019	-0.075	0.270	0.302	.350*	0.016	0.126	0.302	0.100
	Sig. (2-tailed)	0.905	0.640	0.088	0.055	0.025	0.920	0.432	0.055	0.533
Vis_intk_durMax	Spearman's rho	0.092	-0.034	0.308	0.300	0.277	0.019	0.152	0.233	0.029
	Sig. (2-tailed)	0.569	0.835	0.050	0.057	0.079	0.908	0.342	0.143	0.858
Vis_intk_durMin	Spearman's rho	-0.047	0.048	0.292	-0.007	0.111	-0.263	0.028	.402**	-0.160
	Sig. (2-tailed)	0.772	0.767	0.064	0.963	0.489	0.096	0.862	0.009	0.319
Sac_cnt	Spearman's rho	-0.052	.355*	-.427**	-0.285	-0.079	0.149	0.074	-0.112	0.228
	Sig. (2-tailed)	0.749	0.023	0.005	0.071	0.622	0.351	0.645	0.487	0.152
Sac_freq	Spearman's rho	-0.052	.355*	-.427**	-0.285	-0.079	0.149	0.074	-0.112	0.228
	Sig. (2-tailed)	0.749	0.023	0.005	0.071	0.622	0.351	0.645	0.487	0.152
Sac_durtot	Spearman's rho	0.005	0.271	-.447**	-0.225	-0.010	0.108	0.040	-0.098	0.232

	Sig. (2-tailed)	0.974	0.087	0.003	0.157	0.950	0.503	0.803	0.544	0.145
Sac_durAvg	Spearman's rho	0.111	0.147	-0.138	0.140	.340*	-0.071	-0.233	-0.152	0.264
	Sig. (2-tailed)	0.489	0.358	0.391	0.383	0.030	0.657	0.143	0.341	0.096
Sac_durMax	Spearman's rho	0.198	0.211	-0.239	0.209	.328*	0.076	-0.112	-0.140	0.214
	Sig. (2-tailed)	0.214	0.186	0.132	0.189	0.036	0.639	0.485	0.382	0.178
Sac_durMin	Spearman's rho	-0.167	-0.068	0.102	0.035	-0.047	0.011	0.173	-0.157	-0.136
	Sig. (2-tailed)	0.296	0.674	0.526	0.826	0.771	0.947	0.280	0.327	0.395
Sac_ampltot	Spearman's rho	-0.023	.320*	-.468**	-0.150	-0.200	0.212	-0.011	-0.272	0.199
	Sig. (2-tailed)	0.887	0.041	0.002	0.349	0.209	0.184	0.947	0.086	0.213
Sac_amplAvg	Spearman's rho	0.036	.324*	-0.229	-0.005	-0.106	0.107	-0.118	-0.165	0.047
	Sig. (2-tailed)	0.825	0.039	0.149	0.976	0.510	0.504	0.464	0.302	0.769
Sac_amplMax	Spearman's rho	0.064	0.272	-0.286	-0.023	-0.086	.322*	-0.027	-0.137	0.059
	Sig. (2-tailed)	0.690	0.086	0.070	0.887	0.591	0.040	0.867	0.394	0.714
Sac_amplMin	Spearman's rho	-0.279	.311*	-0.023	0.096	-.381*	-0.243	-.385*	-0.113	0.101
	Sig. (2-tailed)	0.077	0.048	0.888	0.549	0.014	0.126	0.013	0.480	0.530
Sac_veltot	Spearman's rho	0.030	.316*	-0.250	-0.115	0.155	0.029	0.207	-0.137	0.211
	Sig. (2-tailed)	0.853	0.044	0.114	0.474	0.333	0.859	0.193	0.393	0.184
Sac_velAvg	Spearman's rho	0.030	0.266	-0.110	0.118	0.304	0.061	0.221	-0.105	0.149
	Sig. (2-tailed)	0.850	0.093	0.494	0.461	0.053	0.704	0.166	0.514	0.353
Sac_velMax	Spearman's rho	0.113	0.308	-0.167	0.175	.332*	0.138	0.167	-0.083	0.127
	Sig. (2-tailed)	0.482	0.050	0.297	0.273	0.034	0.390	0.296	0.606	0.430
Sac_velMin	Spearman's rho	-.311*	-0.049	0.013	-0.164	-0.080	-0.233	-0.091	0.021	0.126
	Sig. (2-tailed)	0.048	0.761	0.937	0.307	0.618	0.143	0.572	0.898	0.431
Sac_latAvg	Spearman's rho	0.166	-0.261	0.260	0.072	0.106	-0.011	-0.129	0.197	-0.155
	Sig. (2-tailed)	0.301	0.099	0.100	0.657	0.509	0.946	0.421	0.217	0.332

Blink_cnt	Spearman's rho	-0.044	-0.235	-0.025	0.112	-0.261	0.012	-0.116	-0.066	-0.399**
	Sig. (2-tailed)	0.787	0.139	0.878	0.485	0.099	0.940	0.472	0.681	0.010
Blink_freq	Spearman's rho	-0.044	-0.235	-0.025	0.112	-0.261	0.012	-0.116	-0.066	-0.399**
	Sig. (2-tailed)	0.787	0.139	0.878	0.485	0.099	0.940	0.472	0.681	0.010
Blink_dur	Spearman's rho	-0.066	-0.270	0.019	0.180	-0.238	-0.035	-0.137	-0.067	-.393*
	Sig. (2-tailed)	0.680	0.088	0.908	0.261	0.134	0.826	0.392	0.677	0.011
Blink_durAvg	Spearman's rho	-0.137	-.419**	0.156	.316*	-0.138	-0.034	-0.120	0.000	-0.301
	Sig. (2-tailed)	0.393	0.006	0.331	0.044	0.390	0.833	0.457	0.999	0.056
Blink_durMax	Spearman's rho	-0.103	-.333*	0.082	0.266	-0.193	-0.068	-0.113	-0.020	-.314*
	Sig. (2-tailed)	0.521	0.033	0.612	0.093	0.227	0.675	0.482	0.903	0.046
Blink_durMin	Spearman's rho	-0.174	-.387*	0.167	.432**	0.046	0.090	0.002	0.158	-0.270
	Sig. (2-tailed)	0.277	0.012	0.296	0.005	0.774	0.575	0.992	0.323	0.088
Scanpath_len	Spearman's rho	-0.050	0.298	-.407**	-0.267	-0.215	-0.059	-0.063	-0.180	0.132
	Sig. (2-tailed)	0.758	0.058	0.008	0.092	0.177	0.712	0.698	0.261	0.412

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Table G. 11. Bivariate correlation- Non-parametric (Spearman's rho) results for significant measures of eye-tracking or significant signs only (AOI-based)

Bivariate correlation- Non-parametric (Spearman's rho) results for significant measures of eye-tracking or significant signs only (AOI-based)								
		WAFS	COSACOEOT	WRP	DNWSH_T	NAFUP_T	COSACOEOT_T	WRP_T
AOI_size004First_vis_intkdur	Spearman's rho	0.063	0.169	0.271	.341*	.357*	0.029	0.121
	Sig. (2-tailed)	0.697	0.289	0.087	0.029	0.022	0.857	0.451
AOI_size004_Revisits	Spearman's rho	-0.032	-.372*	0.088	0.188	0.303	0.108	-0.078
	Sig. (2-tailed)	0.841	0.017	0.585	0.240	0.054	0.501	0.628

AOI_size004_Vis_intkCnt	Spearman's rho	.354*	-.355*	-0.295	-0.028	-0.286	-0.004	-0.103
	Sig. (2-tailed)	0.023	0.023	0.061	0.864	0.070	0.979	0.520
AOI_size004_Vis_intkTime	Spearman's rho	0.167	0.151	.354*	0.099	0.237	0.110	.331*
	Sig. (2-tailed)	0.295	0.347	0.023	0.539	0.135	0.494	0.035
AOI_size004_Avg_vis_intkdur	Spearman's rho	-0.075	0.253	0.306	0.056	.359*	0.100	0.290
	Sig. (2-tailed)	0.640	0.110	0.051	0.730	0.021	0.532	0.066
AOI_size005First_vis_intkdur	Spearman's rho	0.067	0.161	0.272	-0.051	0.199	0.021	0.024
	Sig. (2-tailed)	0.678	0.313	0.085	0.751	0.212	0.898	0.880
AOI_size005_Revisits	Spearman's rho	0.058	-0.144	0.008	0.089	-0.178	-0.123	0.125
	Sig. (2-tailed)	0.719	0.368	0.958	0.580	0.266	0.442	0.437
AOI_size005_Vis_intkCnt	Spearman's rho	.316*	-0.303	-.322*	-0.027	-0.134	-0.176	-0.104
	Sig. (2-tailed)	0.044	0.054	0.040	0.867	0.404	0.272	0.516
AOI_size005_Vis_intkTime	Spearman's rho	0.116	0.170	0.293	-0.002	0.192	-0.165	0.111
	Sig. (2-tailed)	0.471	0.288	0.063	0.992	0.229	0.302	0.488
AOI_size005_Avg_vis_intkdur	Spearman's rho	-0.071	0.265	0.298	0.061	0.206	-0.006	0.204
	Sig. (2-tailed)	0.660	0.094	0.059	0.704	0.196	0.969	0.201
AOI_size006First_vis_intkdur	Spearman's rho	0.075	-0.066	0.043	0.032	-0.196	0.075	0.018
	Sig. (2-tailed)	0.640	0.684	0.789	0.842	0.218	0.640	0.909
AOI_size006_Revisits	Spearman's rho		-0.106	0.088	0.236	-0.041	-0.095	0.068
	Sig. (2-tailed)		0.508	0.585	0.138	0.800	0.554	0.672
AOI_size006_Vis_intkCnt	Spearman's rho	0.075	-0.074	0.035	0.097	0.007	.311*	0.206
	Sig. (2-tailed)	0.640	0.645	0.828	0.545	0.966	0.048	0.196
AOI_size006_Vis_intkTime	Spearman's rho	0.075	-0.066	0.041	0.056	0.061	.333*	0.162
	Sig. (2-tailed)	0.640	0.684	0.801	0.729	0.703	0.033	0.313

AOI_size006_Avg_vis_intkdur	Spearman's rho	0.075	-0.066	0.043	0.216	0.082	0.148	0.110
	Sig. (2-tailed)	0.640	0.684	0.789	0.174	0.611	0.357	0.494
TimeToFirstAppear_allAOI	Spearman's rho	0.075	0.208	-0.291	0.004	-0.156	-0.118	0.069
	Sig. (2-tailed)	0.640	0.193	0.065	0.979	0.331	0.461	0.667
<p>** . Correlation is significant at the 0.01 level (2-tailed).</p> <p>* . Correlation is significant at the 0.05 level (2-tailed).</p>								



Appendix H: Example of Questionnaire Sample for Chapters 2, 3 and 4

2/2/24, 3:29 PM Questionnaire

Questionnaire

Your voluntary participation in this study is highly appreciated. Participants are requested to fill up the survey form. Filling up the form indicates your consent to participate in the program voluntarily. It will take hardly 15 to 20 minutes for the whole survey. Your patience is highly appreciated. The questions are divided into demographic questions and study-related questions. Kindly answer all the questions.

This questionnaire is intended to help assess the understanding of the COVID-related signages which is essential to minimize the communal spreading of the disease. The intended subjects are the corporate people, workers engaged in small and big factories, etc. The intention is to know how much the signages are understandable or meaningful.

The anonymity of the participants will be maintained and none of the data will be revealed to the outside world or revealed for any commercial purpose. Any usage of data for publication will be strictly maintaining the rights of the Confidentiality of an individual. The researcher is abided by the ethical rules and regulations.

** Indicates required question*

Demographic Question

Please answer the questions given below. It might take a maximum of 2 minutes.

1. Name (optional)

2. Age *

<https://docs.google.com/forms/d/1qwCfeDzS9g-cfZYeosq9DFMZGVHmcqPM93K0cpYXtCs/edit> 1/55

3. 3. State and district of work *

4. 4. Educational Background *

Mark only one oval.

- 10th
- 12th
- Diploma
- Bachelors
- Masters/ Post graduate
- Doctors/ PhD
- Other: _____

5. 5. Gender *

Mark only one oval.

- Male
- Female
- Transgender
- Others

6. 6. Role : (Please Tick) *

Mark only one oval.

- Safety-Officer / Manager
- Officer/ Manager
- Engineer/ Sr. Engineer
- Factory Supervisor
- HR
- Worker
- Other: _____

7. 7. Number of months/ years of Work experience in the workplace: Please mention month or year (e.g. 11 months or 7 years) *

8. 8. Industry and Sector *

Mark only one oval.

- Food, drink and tobacco
- Chemical industries
- Energy, Oil and gas production; oil refining
- Health services
- Mining
- Shipping; ports; fisheries; inland waterways
- Textiles; clothing; leather; footwear
- Transport equipment manufacturing
- Utilities (water; gas; electricity)
- Education
- Other: _____

9. 9. Broad category of Product/ Product line *

Study-related Data

The participants are requested to view each signage only once (for 5 seconds approx) and answer the question accordingly. There are 14 signages and it must take a maximum of 13 minutes. Each question is compulsory and needs to be answered rightly. The participants are requested not to go back and correct anything in the previous question. They are also requested not to refer to any book or internet while answering the question. The participants should not try to copy and paste any text from the signages which consists of text as well as images. They must try to write whatever they have understood from the signage. The intention is to know how much the signages are exactly understandable by common people. I thank you for your patience and your participation in this study is highly appreciated.

Signage 1



10. 1.1. Have you ever seen this signage or related signage before? (Please Tick) *

Mark only one oval.

Yes

No

Not Sure

2/2/24, 3:29 PM

Questionnaire

11. 1.2. Assign a score for your understanding of signage¹ that it intends to convey. (Please Tick) Note: 0.0 is the minimum * value and 1.0 is the maximum value. (0.0 signifies no understanding, 0.1 signifies slightly better understanding than 0.0,....., 0.5 signifies medium understanding, 1.0 signifies complete understanding.)

Mark only one oval.

- 0.0
- 0.1
- 0.2
- 0.3
- 0.4
- 0.5
- 0.6
- 0.7
- 0.8
- 0.9
- 1.0

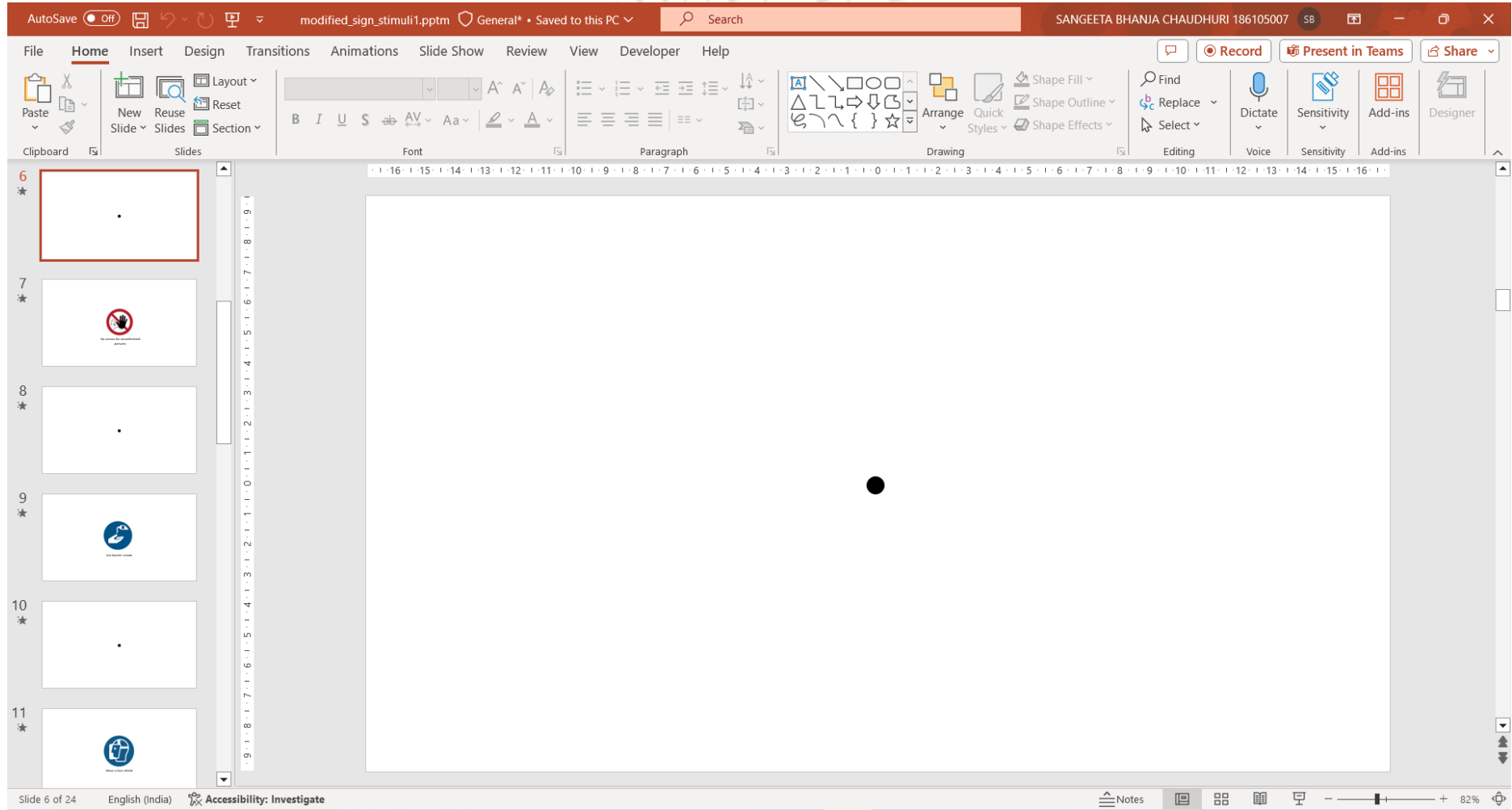
12. 1.3. What message does the signage convey? Explain. *

13. 1.4. Rate the difficulty level of understanding of signage 1. (Please Tick) *

Mark only one oval.

- Very Easy
- Easy
- Neutral (Nor easy nor difficult)
- Difficult
- Very difficult

Appendix I: Stimuli for eye-tracking experiment in Chapter 4



Appendix J: Sign variant-Ranking Questionnaire for Chapter 6

2. Please see the sign variants a, b, c, and d and assign rank to the variants as 1, ^{*} 2, 3, 4 (in each row) according to the comprehensibility level of sign (variants) you think. 1 is the highest comprehensible variant and 4 is the least comprehensible variant. Please do not assign similar rank to any variant.



a



b



c



d

Cough or sneeze in a tissue

1

2

3

4

a. Variant 1 (sele...

Select any option which is not selected before. *

	1	2	3	4
b. Variant 2 (sele...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Select any option which is not selected before *

	1	2	3	4
c. Variant 3 (sele...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Select any option which is not selected before *

	1	2	3	4
d. Variant 4 (sele...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Appendix K: Glossary

AOI	Areas of Interest (AOI) is a region of the stimuli that are selected to extract metrics from that region(s) during or after the eye-tracking experiment.
AOI Scope	“Scope of AOI - local or global (<i>quoted from SMI BeGaze Manual</i>).
AOI Size[px]	Size of AOI in pixel - the part overlaying the stimulus is considered, and parts outside the stimuli area are ignored. <i>For dynamic AOIs, the size is the sum of sizes at each sample timestamp (as defined above) where the AOI is visible averaged by the number of samples where the AOI is visible (quoted from SMI BeGaze Manual).</i>
Average fixation count	Average fixation count on each AOI (<i>quoted from SMI BeGaze Manual</i>).
Average fixation duration	The mean of fixation duration on each AOI. (i.e., Gaze duration mean) (<i>quoted from SMI BeGaze Manual</i>).
Blink count	Number of blinks in the trial (<i>quoted from SMI BeGaze Manual</i>).
Blink duration total [ms]	A sum of the duration of all blinks (<i>quoted from SMI BeGaze Manual</i>).
Blink frequency [count/s]	Number of blinks per second (<i>quoted from SMI BeGaze Manual</i>).
Comprehensibility	<i>Comprehensibility refers to the ability of the individual reading a label, warning, or material safety data sheet to understand the information sufficiently to take the desired action. Comprehensibility is a measure of how well the receiver of the information understood” (direct quote from Sattler et al., (1997)).</i>
Comprehension Test	<i>Procedure for quantifying the degree of understanding of a proposed graphical symbol (direct quote from ISO 9186-1: 2014 (E)).</i>
Computational model	The computational model uses computers to simulate and study complex systems in a fast, easy, reliable, and efficient way.
Education	It’s a longer-term, process that includes the goals of training. It explains why certain information must be known. Education imparts special importance to the scientific foundation of the material presented (<i>NIOSH, (1999)</i>).
Fire equipment sign	A Safety sign indicating the location or identification of fire equipment (<i>ISO-3864-1</i>).
First fixation duration/	Time spent on the first fixation/ visual intake (<i>quoted from SMI</i>

visual intake duration	<i>BeGaze Manual</i>).
Fixation count / visual-intake count	Number of fixations/ visual intake for selected subjects (<i>quoted from SMI BeGaze Manual</i>).
Fixation duration Minimum/ visual intake duration Minimum [ms]	Shortest fixation duration of all selected trials (<i>quoted from SMI BeGaze Manual</i>).
Fixation duration Average/ visual intake duration Average [ms]	A sum of the duration of all fixations of all selected trials divided by the number of selected trials (<i>quoted from SMI BeGaze Manual</i>).
Fixation duration Maximum/ visual intake duration Maximum [ms]	Longest fixation duration of all selected trials (<i>quoted from SMI BeGaze Manual</i>).
Fixation duration Total/ visual intake duration Total [ms]	A sum of the duration of all fixations of all selected trials (<i>quoted from the SMI BeGaze Manual</i>).
Fixation frequency / visual intake frequency [count/s]	Number of fixations of all selected trials per second divided by the number of selected trials (<i>quoted from SMI BeGaze Manual</i>).
Fixation/ visual intake [px]	The largest value for the sum of X and Y (<i>quoted from SMI BeGaze Manual</i>).
Gaze duration	Total fixation duration within a word or an AOI (<i>quoted from SMI BeGaze Manual</i>).
Gaze points	Shows the elements of a stimulus the eyes are looking at. For example, If an eye tracker collects data with a sampling rate of, for example, 60 Hz, one will have 60 individual gaze points per second.
Glances	Increments the counter each time a fixation hits the AOI if not hit before.
Ishihara Test	A color vision test that helps determine general and specific color vision deficits. This consists of a series of color plates in which the test participants are asked to identify numbers or geometric shapes composing specific colors that are hidden within a field of spots of very similar colors. People with normal color vision can see numbers or geometric figures while people with abnormal color vision cannot see them.
ISO 3864-1 and ISO	<i>ISO 3864-1</i> and <i>ISO 3864-2</i> standards together outline the best practices for the design of product safety signs and safety labels.

3864-2	These standards create a unified design for safety signs and labels to be used on products and in the workplace.
Literacy	It is defined as the ability to “use printed and written information to function in society, to achieve one’s goals, and to develop one’s knowledge and potential.” (<i>direct quoted from National Coalition for Literacy, 2009</i>).
Mandatory action sign	A Safety sign indicating that a specific course of action is to be taken is forbidden (<i>ISO 3864-1</i>)
Multiple signs	Sign that combines two or more safety signs and associated supplementary signs on the same rectangular carrier (<i>ISO 3864-1</i>)
Overfitting	Overfitting is an undesirable machine learning behavior that occurs when the machine learning model gives accurate predictions for training data (e.g. pre-trained with STS, SNLI, Wikipedia, question-answer, blogs, etc.) but not for new data (test data e.g. - the safety written message used for the study).
Pre-trained models	A model that uses saved weights and biases, trained by somebody else before to solve similar kinds of problems. It saves time as it avoids training the neural network from scratch, and requires less data. The pre-trained model can also be used to test data directly.
Prohibition sign	A safety sign indicating that specific behavior is forbidden (<i>ISO 3864-1</i>)
Revisits [count]	Average Revisits = (Number of glances divided by selected subjects with at least one visit) - 1 (<i>quoted from SMI BeGaze Manual</i>).
Saccade amplitude total [°]	The sum of all saccades amplitude of all selected trials (<i>quoted from SMI BeGaze Manual</i>).
Saccade amplitude average [°]	The sum of all saccades amplitude of all selected trials divided by the number of saccades in the trial (<i>quoted from SMI BeGaze Manual</i>).
Saccade amplitude	Max. saccade amplitude of all selected trials (<i>quoted from SMI</i>

maximum [°]	<i>BeGaze Manual</i>).
Saccade amplitude Minimum [°]	Min. saccade amplitude of all selected trials (<i>quoted from SMI BeGaze Manual</i>).
Saccade count	Total number of saccades counted within an AOI (<i>quoted from SMI BeGaze Manual</i>).
Saccade duration	A sum of saccadic time spent within an AOI (<i>quoted from SMI BeGaze Manual</i>).
Saccade duration Average [ms]	A sum of all saccade duration of all selected trials divided by the number of selected trials (<i>quoted from SMI BeGaze Manual</i>).
Saccade duration Maximum [ms]	Longest saccade duration of all selected trials (<i>quoted from SMI BeGaze Manual</i>).
Saccade duration Minimum [ms]	Shortest saccade duration of all selected trials (<i>quoted from SMI BeGaze Manual</i>).
Saccade duration Total ms]	A sum of all saccade duration of all selected trials (<i>quoted from SMI BeGaze Manual</i>).
Saccade frequency [count/s]	Number of saccades per second of all selected trials divided by the number of selected trials (<i>quoted from SMI BeGaze Manual</i>).
Saccade length	Distance between two consecutive fixations (<i>quoted from SMI BeGaze Manual</i>).
Saccade velocity Total [°/s]	A sum of all saccades velocities of all selected trials (<i>quoted from SMI BeGaze Manual</i>).
Saccade velocity Maximum [°/s]	Max. value of the saccade velocity of all selected trials (<i>quoted from SMI BeGaze Manual</i>).
Saccade velocity Minimum [°/s]	Min. value of the saccade velocity of all selected trials (<i>quoted from SMI BeGaze Manual</i>).
Saccade velocity Average [°/s]	A sum of all saccades velocities of all selected trials divided by the number of saccades in the trial (<i>quoted from SMI BeGaze Manual</i>).
Safety color	Color with special properties with which a safety meaning is attributed (<i>ISO 3864-1</i>).
Safety sign	A sign that gives a general safety message, obtained by a combination of a color and geometric shape and which, by the addition of a graphical symbol, gives a particular safety message (<i>ISO 3864-1</i> .)
Scanpath Length [px]	A sum of the lengths (distance from start to end) of all the saccades in

	the scan paths of all selected trials (<i>quoted from SMI BeGaze Manual</i>).
Semantic Transparency	The meaning of a symbol is clear (<i>Caire et al., 2013</i>). The idea is that if novices can understand/ comprehend a particular sign, experts/ professionals can easily do so.
Signs	A sign is comprised of its physical form and a mental concept associated with it. This concept is a comprehension of external reality. Sign relates to reality through concepts and people who use it (<i>Salvendy & Karwowski, (2016)</i>).
Subjective Tests	Subjects write their original answers (long or short descriptions) based on their understanding of the signage concerned here. The question framed here requires an answer of short answer type which is decipherable from the keyword.
Supplementary sign	Sign that is supportive of a safety sign. The main purpose of the supportive sign is to provide additional clarification (<i>ISO 3864-1</i>).
Time to first fixation/ Time to first visual intake	Time spent from stimuli onset to the first fixation arrival (<i>quoted from SMI BeGaze Manual</i>).
Total fixation count/ Total visual intake count	Total number of fixations counted in an AOI or a task (<i>quoted from SMI BeGaze Manual</i>).
Total fixation duration/ Total visual intake duration	Total time spent on fixations (<i>quoted from SMI BeGaze Manual</i>).
Training (employees)	Communication is directed at a defined population to develop skills, modify behavior, and increase competence (<i>NIOSH, (1999)</i>).
Training (machine/ model)	Learning good values for all the weights and biases in a neural network.
Warning sign	Safety signs indicate a specific source of potential harm (<i>ISO 3864-1</i>).

List of Publications

- **Journal published**

1. Chaudhuri, S.B., Majhi, M. and Karmakar, S. " A scoping review on role of communication media for effective OSH awareness and training "International Journal of Reliability and Safety (IJRS)", Vol. 15 (1-2), 1-17 (2021) <https://doi.org/10.1504/IJRS.2021.119641> [Scopus Indexed, Compendex journal]

- **Conference accepted**

1. Chaudhuri, S.B., Yadav, P., Majhi, M. and Karmakar, S. "Human-Centered Sign Design/ Redesign Life Cycle for Comprehension of Safety Sign". International Joint Conference- ACED 2023, HWWE 2023 & BRICSplus HFE 2023, Mumbai, SpringerNature.

- **Conference and Book Chapters published**

1. Chaudhuri, S.B., Majhi, M. and Karmakar, S. " Evaluation of Comprehensibility of a Sign by Triangulation Method", In Chakrabarti, D., Karmakar, S., Salve, U.R. (eds) Ergonomics for Design and Innovation. HWWE 2021. Lecture Notes in Networks and Systems, vol 391. Springer, Cham. https://doi.org/10.1007/978-3-030-94277-9_154 [Scopus Indexed]

- **Journal(s) submitted/ under review**

1. Chaudhuri, S.B., Majhi, M. and Karmakar, S. "An Analytical Comparative Study between Two Techniques of Data Collection for Comprehensibility Evaluation of Safety Signs. " [Communicated]
2. Chaudhuri, S.B., Majhi, M. and Karmakar, S. "An Analytical Man-Machine-Hybrid Model for Comprehensibility Evaluation of Safety Messages Using Deep Learning Approach." [Communicated]
3. Chaudhuri, S.B., Majhi, M. and Karmakar, S. "Exploration and analysis of significant eye-tracking measures, subjective measures, and their inter-relationship: A study of comprehensibility of OHS signs." [Communicated]

Curriculum vitae

Social Contact details:

Linked in: <https://www.linkedin.com/in/ms-sangeeta-bhanja-chaudhuri-6486128/>

Google Scholar: <https://scholar.google.co.in/citations?user=wQtFy0cAAAAJ&hl=en>

Research Gate: <https://www.researchgate.net/profile/Sangeeta-Bhanjachaudhuri>

Publons: <https://publons.com/researcher/4667762/sangeeta-bhanjachaudhuri/>

Educational Qualifications:

Doctor of Philosophy (Ph.D.): From the Department of Design, Indian Institute of Technology Guwahati, India. PhD Viva-Voce: 05/04/2024, Cumulative Performance Index (CPI): 8.56/10.

M.Tech: In Information Technology from Gauhati University Institute of Science & Technology, Guwahati, India. Year of Completion: June 2016, Cumulative Performance Index (CGPA): 9.80/10.

M.Phil: In Computer Science from Vinayaka Missions University, Tamil Nadu, India. Year of Completion: June 2009, Percentage: 73%.

M.C.A: In Computer Applications from Bangalore University, Bengaluru, India. Year of Completion: June 2007, Percentage: 83%.

B.Sc: In Science (General) from Calcutta University, Kolkata, India. Year of Completion: June 2001, Percentage: 57.48%.

Job Experiences:

1. Assistant Professor, Department of Information Science and Engineering, CMR Institute of Technology, AECS Layout, Bengaluru from October 2023.
2. Senior Content Strategist (Research and Development- Data Science) in GREAT LAKES E-LEARNING SERVICES PVT. LTD., Bengaluru from May 2022.
3. Teaching Assistant at IIT Guwahati (Department of Design), Assam, India., from July 2016 to May 2022.
4. Teaching Assistant at Gauhati University, Institute of Science & Technology (Department of IT), Assam, India., from July 2014 to May 2016.
5. Assistant Professor (Computer Science & Engineering) at the NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY, BPUT, Odisha, India., from March 2011 to July 2013.
6. Lecturer (Computer Science & Engineering) at the INSTITUTE OF ENGINEERING & INDUSTRIAL TECHNOLOGY, WBUT, India., from August 2010 to February 2011.
7. Lecturer (Computer Science) at KRISTU JAYANTI COLLEGE OF MANAGEMENT & TECHNOLOGY, BANGALORE UNIVERSITY from January 2010 to June 2010.

8. Lecturer (MCA) at THE OXFORD COLLEGE OF SCIENCE, BANGALORE UNIVERSITY from February 2008 to January 2010.
9. Lecturer (Computer Science) at DAYANANDA SAGAR COLLEGE OF MGMT & INFO. TECHNOLOGY, BANGALORE UNIVERSITY from September 2007 to January 2008.

Proficiency:

1. Programming/ scripting language: HTML, CSS, Javascript (Developed websites for the department of Design, IIT Guwahati, on various platforms), SQL, C, Basics of Python.
2. Tools: Brackets, Matlab programming, Rose2, Microsoft Office, SPSS, Dreamweaver, Pedagogy platform (CodeTantra), Vosviewer.
3. Courses: Usability Engineering, System Analysis and design, Software Engineering, Experiential Design, Human-Computer Interface and Typography, Research Methodology, statistics, Design Research Methodology, OSH safety communication, User Experience Design, UI/UX, Design-based Project, Data Science, Internet and Web Technologies, Application of Artificial Intelligence and machine learning besides core Computer science and Information science subjects.
4. Instruments: Expert in working with Eye-tracking devices (SensoMotoric Instruments) and analysis software.

Research experience in MTech:

Thesis Title: Design of a Modern Clinical Expert System Using Data Mining Algorithms and Application of Statistical Approach for Performance Evaluation.

The main goal of this work is to design a clinical Expert System that will simulate the work of a medical practitioner. The model developed seems to predict the presence or absence of diseases (cardiovascular disease/ heart disease, premenstrual syndrome (PMS)). Rose2 and MATLAB R2010 have been used for implementing Rough Set Theory and Neural Networks. A comparative study was done to record the models' performances. Both models seem to classify heart and PMS data with encouraging accuracy. The model using a neural network with the BFGS algorithm gave better results than the Rough Set Theory.

Other research experiences:

Design-based Project: An interactive toy was designed and developed for the vernacular medium school children to help newly admitted kids learn the alphabet or number systems. There is a need for the kids to enjoy playing with it while simultaneously, learning the basics of ABCD or rhymes.

Environment and Experience Design Project: IIT Guwahati intended to honor two Heroes and Martyrs groups by setting up a war memorial on its campus. The intention is to display the portraits of war warriors decorated with Param Veer Chakra “Wall of Heroes.” and design a “Monument for Martyrs” in memory of the martyrs who laid down their lives during the Assam Agitation against foreign nationals. A conceptual design is proposed for the above Project.

Animation Projects (Under the supervision of eminent animator, Mr. Vaibhav Kumaresh):

- * Design of a dancing doll, theme, “Ballerina alone in Love” using Puppet making and pixilation.
- * Simple animation experiments with coins using stop motion.
- * Created animation movies using displacement techniques. The theme is: “Chai”
- * Movie with sound using Pixilation. The theme is “Soul”

Graphics Design Project: Visiting card design.

Awards:

- * Ph.D.- MHRD Scholarship, M.Tech (NEC-Merit Scholarship)
- * Got selected for the Google Graduate Symposium 2021 and attended for 5 days.
- * Secured 2nd Position in M.Tech at Gauhati University (GUIST) in 2016.
- * Gold medal for observatory power and patience (2008).
- * Received honorarium- a gift voucher for securing the first position in MCA (First Semester) in “The Oxford College of Science”, Bangalore

Professional membership:

Lifetime member of the Linguistic Society of India

Lifetime member of IAENG, International Association of Engineers