

CRANFIELD UNIVERSITY

Teegan Laura Johnson

Integration of Physical and Psychological Human Data into Digital Human
Modelling for Manufacturing System Design

School of Aerospace, Transport & Manufacturing

PhD

Academic Year: 2014 - 2021

Supervisor: Dr. Sarah R. Fletcher
Associate Supervisor: Dr. Phil Webb
April 2021

CRANFIELD UNIVERSITY

School of Aerospace, Transport & Manufacturing

PhD

Academic Year 2014 - 2021

Teegan Johnson

Integration of Physical and Psychological Human Data into Digital
Human Modelling for Manufacturing System Design

Supervisor: Dr. Sarah R. Fletcher
Associate Supervisor: Dr. Phil Webb
April 2021

© Cranfield University 2021. All rights reserved. No part of this
publication may be reproduced without the written permission of the
copyright owner.

ABSTRACT

Industry 4, the current industrial revolution, is bringing an increase in digitisation to manufacturing. It involves, in part, a move towards using the digital human modelling (DHM) capabilities in computer aided design (CAD) programs to design and monitor products, workstations, and the new technology being introduced into manufacturing working environments. DHM in CAD involves modelling the impact manufacturing products and environments have on those who work in and with them. At present, the focus of CAD programs with integrated DHM tools is the physical impact of manufacturing environments on workers or operators; psychological impacts, however, are missing. This thesis aimed to identify and establish psychological parameters that could improve the capability of DHM for manufacturing system design.

Two studies identified psychological equations for integration into DHM CAD, the first established an equation between physical strain and mental workload. This study was preceded by an investigation into the effectiveness of an inertial motion capture suit to identify an effective means of measuring physical strain. The second study investigated the effect of human robot collaboration on perceived trust, workload, and mood, when manipulating robot speed. A predictive equation could only be established for trust; and this equation was integrated into a CAD program, and the reliability was assessed using an experimental mock-up of an Airbus installation process. The equation when integrated into the CAD program was able to successfully predict the trust response of participants in the reliability experiment and, therefore, achieved the aim of this thesis. The implications and limitations of the studies are discussed.

Keywords:

Mental Workload, Human Robot Collaboration, Trust, Mood, Quantitative Analysis, Monte Carlo Simulation, Generalised Estimating Equations, Inertial Motion Capture

ACKNOWLEDGEMENTS

I feel like the longer your PhD takes the longer that the acknowledgements end up being. This PhD has affected not just my life but also those around me who have provided invaluable support over the last seven years.

Sarah, you gave me this opportunity to achieve a PhD, and you have been with me on this journey from the beginning, through all changes that this PhD has taken, from the purely physical to the introduction of psychology. Thank you for fighting for my opportunity to include, what was always the important part to me, the psychology. Thanks to you I have been able to explore so many areas that I don't think I would have had the opportunity to anywhere else. I have also been able to stretch my understanding and to grow intellectually while gaining so much independence, and confidence. Thank you for putting up with my mangled sentences and your feedback on language and grammar. It is finally finished, and you must be as relieved as I am.

John Thrower, thank you for your support in "science!", the early experiments, your insightful questions and for all of your time helping me get the mocap suit to work. Ellis Haddow thank you for co-ordinating the suit and the robot during the psychological and physical experiments, and for being the bridge between John leaving and James arriving. Your knowledge and presence meant that that transition went so smoothly and enabled the last two sets of experiments to come together well. James Fowler, thank you for your support during the final two experiments completed to finish data collection for this thesis.

Richard Lilley and Tracksys thank you so much for the loan of the heart rate monitor. It allowed me to take the psychological and physical experiment in a direction I had not thought I would be able to. Thank you for the training of its use and the discussion and help during the process of figuring out just which HRV statistic was the most appropriate. This enabled me to develop in a new area of psychology that I had not had the opportunity to in my undergraduate and master's studies.

The success of the experiment completed in Chapter 5 wouldn't have been possible without the help of Airbus and the AMRC. Thank you, Jose Gonzalez Domingo, for finding a process that we could redesign and conducting the technical task analysis; and Adrian Hirst, Tom Smith, and Phil Kitchen thank you for your programming skills. Without your help, I wouldn't have gotten to the CAD integration stage and this thesis wouldn't have felt quite complete. Thank you.

Thank you, Wil Baker, you were there at the start and were full of encouragement that I would achieve the PhD and by the end successfully develop an understanding of where we fit in manufacturing. You were right that I would get there in the end and that all I needed was a little bit of time.

Cyril Jaksic and Iveta Eimontaite, you arrived at the end and got me through the last slog, and I am forever grateful for our chats, laughs, and your encouragement. Thank you both so much for your stats and experimental methodology wisdom. Trying to figure out the analysis would have taken so much longer than it did with both of your brains also working to figure out how to get an equation for a repeated measures experiment. Thank you so much for your reviewing and writing help, I think I figured out some of the commas in the end. When the world has settled down a trip to hang out in person needs to be organised.

Mom and Dad, it is finally done. Thank you for your patience with me as I moaned and complained and fought to finish it. Thank you, Mom, for your reviews, this should be the last time I need your help reviewing my work for my education. I think you might be as relieved about that as I am. Mom, Dad, Kirst and Abi thank you for your support over the years of doing this. Kirstin turns out the floppy hat was harder to come by for me than I ever expected. Thank you so much for that advice on the mountainside, it helped keep this journey a good adventure rather than a not good adventure. You are starting on your own PhD journey; I wish you well little sister and I am here for you if things get rubbish and look forward to seeing you getting your own floppy hat!

Thank you to all the friends who have been with me from the beginning of the journey and those who have joined along the way. I cannot tell you how much

your support, prayers, and cheering on has meant and I look forward to celebrating with all of you.

And finally, I guess a thank you is required for Covid-19, without it and the world grinding to an absolute halt I would never have finished writing this thesis on time. Furloughing afforded me the time and space to do the final little bits of analysis and to immerse myself completely in the write up.

TABLE OF CONTENTS

ABSTRACT.....	i
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	vii
LIST OF FIGURES	ix
LIST OF TABLES	xi
LIST OF EQUATIONS.....	xiii
LIST OF ABBREVIATIONS.....	xiv
PUBLICATIONS	19
1 INTRODUCTION	20
1.1 Summary of the Research Problem.....	20
1.2 Research Aim, Objectives, and Contribution to Knowledge.....	23
1.3 Thesis Overview	24
2 LITERATURE REVIEW.....	27
2.1 CAD and DHM	27
2.1.1 The Advantages of CAD and DHM	31
2.1.2 The Limitations of CAD and DHM	32
2.2 Introducing the Psychological to DHM CAD Programs	39
2.2.1 Mental Workload	42
2.2.2 Human Robot Collaboration	60
2.3 Aim and Objectives.....	67
3 EXPERIMENT: INVESTIGATING THE EFFECTIVENESS OF MOTION CAPTURE TECHNOLOGY	68
3.1 Method	69
3.1.1 Apparatus.....	69
3.1.2 Design	73
3.1.3 Research Ethics.....	74
3.1.4 Participant	74
3.1.5 Procedure.....	74
3.1.6 Analysis.....	86
3.2 Results	88
3.2.1 Whole Body Assessment.....	88
3.2.2 Individual Segment Analysis.....	89
3.2.3 Summary of Findings.....	103
3.3 Discussion.....	106
4 EXPERIMENT: PREDICTIVE COGNITIVE AND PHYSICAL INTERACTION	111
4.1 Method	112
4.1.1 Research Ethics.....	112
4.1.2 Participants	112
4.1.3 Design	112

4.1.4 Apparatus.....	115
4.1.5 Materials.....	120
4.1.6 Procedure.....	121
4.1.7 Analysis.....	129
4.2 Results	134
4.2.1 Descriptive Statistics.....	134
4.2.2 Model Development.....	138
4.3 Discussion.....	139
4.3.1 Study Limitations.....	142
4.3.2 Conclusion	143
5 STUDY: DEVELOPMENT AND INTEGRATION OF A HUMAN ROBOT COLLABORATION EQUATION INTO DHM	145
5.1 Development of a Human Robot Collaboration Equation	146
5.1.1 Method	146
5.1.2 Analysis.....	155
5.1.3 Results	158
5.1.4 Conclusion	164
5.2 Assessment of the Accuracy of the CAD Integrated Human Robot Collaboration Trust Speed Equation	165
5.2.1 Method	165
5.2.2 Results	176
5.3 Discussion.....	177
5.3.1 Conclusion	180
6 DISCUSSION AND CONCLUSION.....	183
6.1 Discussion.....	183
6.1.1 Objective 1: Investigating the Effectiveness of Motion Capture Technology.....	183
6.1.2 Objective 2: Establishing the Combined Parameters of Physical Strain and Mental Workload	186
6.1.3 Objectives 3 & 4: Establish Human-Robot Interaction Equation/s for Industrial Human Robot Collaboration and Assessment of the Predictive Accuracy of the HRC Equation/s after Integration into DHM ..	190
6.2 Contribution to Knowledge	192
6.3 Conclusion	195
REFERENCES	197
APPENDICES.....	219
Appendix A BVH Reference Table.....	219
Appendix B REBA Assessment Worksheet.....	220
Appendix C Investigating the Effectiveness of Motion Capture Technology.....	221
Appendix D Predictive Cognitive and Physical Interaction.....	222
Appendix E Development and Integration of a Human Robot Collaboration Trust Speed Equation into DHM.....	234

LIST OF FIGURES

Figure 1-1: Graphical Illustration of the Structure of the Thesis	26
Figure 3-2: Correct Anatomical Placement of the IGS-180-I Sensors (Synertial, 2011).....	71
Figure 3-1: Animazoo (Synertial) Motion Capture Suit Software Options	72
Figure 3-3: Goniometry of the Shoulder	77
Figure 3-4: Goniometry of the Elbow	78
Figure 3-5: Goniometry of the Wrist.....	78
Figure 3-6: Goniometry of the Hip in a Prone Position.....	79
Figure 3-7: Goniometry of the Knee	80
Figure 3-8: Goniometry of the Ankle.....	81
Figure 3-9: Adapted Goniometry of the Neck.....	82
Figure 3-10: Adapted Goniometry of the Trunk (Chertman, et al., 2010).....	83
Figure 3-11: IGS-Bio Output	85
Figure 3-12: Right Knee Edited BVH Data and Unedited BVH Data.....	86
Figure 3-13: Left Knee Edited BVH Data and Unedited BVH Data	87
Figure 4-1: Laboratory Layout.....	115
Figure 4-2: Labelled Mock-up Identifying Structural Components.....	116
Figure 4-3: Diagram of a Labelled Cut-Section of a Wing (Howe, 2004)	117
Figure 4-4: Bolt Sealant Task.....	118
Figure 4-5: Front and Side Calibration Pictures within AutoCal Software (Synertial, 2009)	122
Figure 4-6: Participant Working In the OH Sat Condition While Completing the Duct Installation Task	124
Figure 4-6: OH Stood and Sat Working Conditions	125
Figure 4-7: Neutral Working Condition	125
Figure 4-9: Layout of Components for Duct Installation Tasks	126
Figure 4-10: Duct Installation Task Steps: a) Blast Tube Inserted and Secured with Three Bolts, b) Installed Cradle Assembly, c) Installed Duct 1 and Cradle Assembly Tightening, d) Duct 1 and 2 Alignment	127
Figure 4-11: Jubilee Clip Placement and Tightening	128

Figure 5-1: Attachment of the Fittings to the End of a Pipe	148
Figure 5-3: Experiment Robot and Layout.....	149
Figure 5-4: Total Trust Scores Across Robot Program Speeds	161
Figure 5-5: Experiment Robot and Layout.....	169
Figure 5-5: Experiment Layout with Representative Slat Start (a) and Finish Positions (b).....	171
Figure 5-7: Participant Completing the Bolt Tightening Task under the Frame	172
Figure 5-8: End Output from Simulation Run	174

LIST OF TABLES

Table 2-1: DELMIA and Jack Comparison	30
Table 3-1: Alpha, Beta, and Complex Body Segments (Synertial, 2012).....	71
Table 3-2: Range of Movement Measured Per Joint	76
Table 3-3: Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs	89
Table 3-4: Right Shoulder Comparison of the Values Obtained with the Goniometer with the Values Obtained Via the Six Outputs	90
Table 3-5: Left Shoulder Comparison of the Values Obtained with the Goniometer with the Values Obtained Via the Six Outputs.....	91
Table 3-6: Right Elbow Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs.....	92
Table 3-7: Left Elbow Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs.....	93
Table 3-8: Right Wrist Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs.....	94
Table 3-9: Left Wrist Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs.....	95
Table 3-10: Neck Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs	96
Table 3-11: Trunk Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs	97
Table 3-12: Right Hip Comparison of the Values Obtained With the Goniometer with the Values Obtained via the Six Outputs	98
Table 3-13: Left Hip Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs.....	99
Table 3-14: Right Knee Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs.....	100
Table 3-15: Left Knee Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs.....	101
Table 3-16: Right Ankle Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs.....	102
Table 3-17: Left Ankle Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs.....	103

Table 3-18: Agreement between Synertial Outputs and Goniometer across Whole Body and Individual Segments.....	105
Table 4-1: Minimum, Maximum, Mean and SD for Mental Workload and Postural Analysis.....	137
Table 5-1: Independent Variable Conditions: Robot Speeds.....	150
Table 5-2: Trust in Industrial Human Robot Collaboration Scale.....	152
Table 5-3: Scoring Scheme for Each Trust Statement.....	156
Table 5-4: Scoring Scheme for Mood	157
Table 5-5: Means and SDs for Trust in Human-Robot Collaboration across Dimensions	159
Table 5-6: NASA-TLX Overall Workload Scores and SDs across Robot Program Speeds.....	162
Table 5-7: UMACL Means and SDs across Robot Program Speeds	163
Table 5-8: Descriptive Statistics and Confidence Interval Data.....	177

LIST OF EQUATIONS

(4-1)..... 138

(4-2)..... 138

(4-3)..... 139

(5-1)..... 173

LIST OF ABBREVIATIONS

ACT-R	Adaptive Control of Thought-Rational
AMBR	Agent-Based Modelling and Behaviour Representation
AMRC	Advanced Manufacturing Research Centre
ANOVA	Analysis of Variance
ANSUR	Anthropometric Survey of US Army Personnel
BPS	British Psychological Society
BVH	Biovision Hierarchical
CAD	Computer Aided Design
CUREC	Cranfield University Research Ethics Committee
CURES	Cranfield University Research Ethics System
DELMIA	Digital Enterprise Lean Manufacturing Interactive Application
DHM	Digital Human Modelling
DV	Dependent Variables
ECG	electrocardiogram
EEG	electroencephalography
ERP	Event related brain potentials
fMRI	functional magnetic resonance imaging
fNIRS	functional near infrared spectroscopy

GDPR	General Data Protection Regulation
GEE	Generalised Estimating Equations
GUI	Graphical User Interface
HF	High Frequency
HR	Heart Rate
HRC	Human Robot Collaboration
HRV	Heart Rate Variability
HSE	Health and Safety Executive
IA	Intelligent Automation
IBI	Inter-beat Interval
IMPRINT	Improved Performance Research Integration Tool
IMU	Inertial Measurement Unit
ISO	International Organization for Standardization
IT	Information Technology
IV	Independent Variable
MAC	Manual Handling Assessment Charts
MDF	medium-density fibreboard
mocap	Motion Capture
MSD	Musculoskeletal Disorders

NASA-TLX	National Aeronautical and Space Administration Task Load Index
NIOSH	National Institute for Occupational Safety and Health
OH	Overhead
OWAS	Ovako Working Posture Analysis System
PC	Personal Computer
PET	Positron Emission Tomography
PFC	Prefrontal cortex
PPE	Personal Protective Equipment
RAPP	Risk Assessment of Pushing and Pulling
REBA	Rapid Entire Body Assessment
RMSSD	Root Mean Square of the Differences between consecutive inter-beat intervals
RSA	Respiratory Sinus Arrhythmia
RULA	Rapid Upper Limb Assessment
SA	Situation Awareness
SAGAT	Situation Awareness Global Assessment Technique
Sammie CAD Ltd.	System for Aiding Man-Machine Interaction Evaluation (Sammie) CAD
SART	Situation Awareness Rating Technique
SD	Standard Deviation

SDNN	Standard Deviation of Normal to Normal inter-beat intervals
SEREC	Science and Engineering Research Ethics Committee
SWAT	Subjective Workload Assessment Technique
UK	United Kingdom
UMACL	UWIST Mood Adjective Checklist
US	United States
UWIST	University of Wales Institute of Science and Technology

PUBLICATIONS

CONFERENCE PAPERS

Johnson, T. L., & Fletcher, S. R. (2018, June). The limitations of using only CAD and DHM in design relating to high value manufacturing. In *Contemporary Ergonomics and Human Factors 2015: Proceedings of the International Conference on Ergonomics & Human Factors 2015, Daventry, Northamptonshire, UK, 13-16 April 2015* (p. 329). CRC Press.

Johnson, T., & Fletcher, S. (2014). A Computer Software Method for Ergonomic Analysis Utilising Non-Optical Motion Capture. *Contemporary Ergonomics and Human Factors*, 93-100.

JOURNAL PAPERS

Johnson, T. L., Fletcher, S. R., Gonzalez Domingo, J., Jaksic, C., Hirst, A., Smith, T., Kitchen, P., Morgan, B., Dodd-Hughes, K., & Carle, F. (In Progress). Development and Integration of a Human Robot Collaboration Rule into DHM.

Johnson, T. L., & Fletcher, S. R. (In Progress). The predictive effect of physical strain on mental workload

TECHNICAL REPORTS

Johnson, T. L., Fletcher, S. R., Gonzalez Domingo, J., Jaksic, C., Hirst, A., Smith, T., Kitchen, P., Morgan, B., Dodd-Hughes, K., & Carle, F. (2019). ICHORD: 'Integrating the Cognitions of Human Operators in digital Robot Design' Feasibility study for the EPSRC Connected Everything Network+

Johnson, T. L., & Fletcher, S. R. (2016). Exploratory Ergonomic Investigation of Working Postures Adopted for Completion of Airbus A320 In-tank Sealing Processes

1 INTRODUCTION

1.1 Summary of the Research Problem

There is currently a shift towards increased digitisation within manufacturing (Peruzzini et al., 2020; Romero et al., 2016; Turk et al., 2021). This partially involves digitally modelling human interactions using computer aided design (CAD) programs, a process that is referred to as digital human modelling (DHM). DHM capabilities in CAD programs are used within proactive ergonomics to design and monitor products, workstations, and new technology being introduced into manufacturing working environments, including robotics (Peruzzini et al., 2020; Turk et al., 2021). This process involves modelling the impact manufacturing products and environments have on those who work in and with them, by ideally identifying and predicting the effect of new and current processes on both the physical and psychological functions of the human to enhance the design of a process or to effectively redesign a process (Sundin & Örtengren, 2006; Yang et al., 2004). Despite the ideal definition of DHM including both physical and psychological tools, at present, CAD programs with integrated DHM tools are capable of assessing the physical impact of manufacturing environments on workers or operators. Presently, there are no measures available to assess psychological impacts. The ability to understand and predict the effect of complex manufacturing systems on an operator's mental abilities or psychological perceptions, particularly when changing a task or introducing new technology, is vital for ensuring the introduction or adaption of safe and effective systems (Young et al., 2015). If psychological effects are not fully understood post-implementation assessments by human factors specialists and psychologists would be required to obtain advice for remedial design or task changes. By addressing these factors prior to implementation, the need for intervention and redesign is either reduced or eliminated. This reduces the cost for companies and helps them to ensure the psychological and physical health and safety of their workforce.

A range of psychological assessment tools, models, and some predictive cognitive modelling programs are available. However, computer models of these

behaviours and processes have not been integrated into DHM CAD programs. Reasons for this include the lack of accessibility of the current cognitive modelling programs, and their inability to present information in a form that is easily implemented into DHM CAD programs (Duffy, 2012; Laughery et al., 2012). This thesis looked to identify a means of introducing psychological assessment tools into DHM. First by identifying suitable constructs that would address current manufacturing challenges, and then secondly by identifying a way to measure psychological change based on current assessment methods used by design engineers. By introducing psychological concepts in this way, it is likely that the need for training may be minimised and the concepts will be provided in easy to understand form, thereby, addressing the limitations of current tools as identified by Duffy (2012) and Laughery et al. (2012).

Two psychological constructs were identified within this thesis as most suitable for integration into DHM CAD programs, due to their criticality for ensuring effective working conditions, for operators within manufacturing. The first was to identify a means of predicting mental workload. The second was to identify the psychological factors affected by collaborative robotics, and to assess whether they could be used to develop a predictive model that could be integrated into a DHM CAD program.

Developing a means for assessing mental workload in DHM CAD programs is critical because excessive mental workload has been found to be one of the greatest contributors to human error in many systems (Laughery et al., 2012). Mental workload is a psychological construct concerned with cognitive processes and the relationship between task demand and psychological resources (Parasuraman et al., 2008; Tattersall, 2000). The lack of psychological assessment methods within CAD becomes a problem where tasks or workstations are redesigned to mitigate physical strain for an operator but contribute to greater psychological strain or workload, which could lead to increases in error rate and possibly production time (Laughery et al., 2012). This effect may be further exacerbated where work systems become more complex, particularly with the move towards integrating more technology into

manufacturing (Romero et al., 2016; Turk et al., 2021). Providing design engineers with a means to assess workload during their decision-making process may reduce the need for additional post design assessments and remedial actions. To do this, effective ways of measuring mental workload requires identification along with a means to integrate it in to DHM CAD programs. There is a small but important body of literature that has researched the concurrent effect of physical and psychological strain (Mehta & Parasuraman, 2013) and found that working in physical strain has an effect on mental workload capacity (DiDomenico & Nussbaum, 2008; Marras & Hancock, 2014; Shaikh et al., 2012). Therefore, this thesis looked to take this understanding and identify whether a statistical relationship could be identified between physical strain and mental workload. The identification of this relationship would provide a means of integrating a measure of mental workload into a DHM CAD program. The intended application for this assessment would be during the design and redesign of workstations or large structures, with the aim for this information to feed into digital based task design and monitoring programs.

One of the technologies being introduced into the manufacturing sector is large high-payload collaborative robots. These are robots that are able to lift heavy loads while working in collaboration with a person. This is a change from robots that could only run safely behind physical barriers, to those that can work in proximity to operators without the need for the same level of separation. They are being implemented, or adapted into collaborative systems, as a means of reducing some of the physical strain, inherent to manufacturing tasks, on operators. However, it is anticipated that this change may contribute to negative psychological reactions from operators, due to the closer working conditions (Arai et al., 2010; Grahn et al., 2018; Koppenborg et al., 2017; Parasuraman & Riley, 1997). As a result, a means of assessing this impact would be advantageous to design engineers, and robot integrators, particularly as the rate of implementation of collaborative robotics is increasing within manufacturing (IFR, 2021). Information of this type would be of benefit to design engineers who are looking to understand how a redesign of a process with the implementation of

collaborative robots may affect operators and could be implemented in digital based task design and monitoring programs.

1.2 Research Aim, Objectives, and Contribution to Knowledge

The aim of this research was to identify and establish the psychological parameters that will improve the capability of digital human modelling for manufacturing system design.

To satisfy the aim four objectives were set:

The first objective was to identify an effective and accurate means of capturing a person's movement. This was essential in order to conduct the experiment to meet the second objective where a means of measuring human strain was required. This objective was met in Chapter 3.

The second objective was to establish the combined parameters of physical strain and mental workload. The intention of this objective was to be able to use these combined parameters to establish an equation that could be integrated into a DHM CAD program, which would provide design engineers with a means of assessing workload. This objective was met in Chapter 4.

The third objective was to establish human-robot interaction equations for industrial human robot collaboration. The intention of this objective was to ascertain whether human-robot interaction equations could be established for trust, workload, and mood. This objective was met in Chapter 5 Section 5.1.

The fourth objective was to assess the reliability of the predictive accuracy of the equation identified to meet Objective 3 after integration into DHM. The experiment conducted to meet the third objective, although two of the concepts could have been taken forward project limitations meant one had to be identified: trust, therefore, the trust speed equation was assessed for its reliability. This objective was met in Chapter 5 Section 5.2.

The completion of the four objectives met the aim of this research and provided principle contributions to:

- A clearer understanding of the current capabilities and limitations of CAD programs that contain DHM. These were not clear within the literature, and the information presented on the various company websites were not comprehensive. This review has identified clear gaps and limitations with the software and highlighted areas for improvement.
- An understanding of the effectiveness of the Synertial IGS-180i inertial motion capture suit, as well as its effectiveness for recording work completed in occluded and confined working conditions. The results indicate specific limbs where the suit does not effectively capture motion, providing the developers of the system with a targeted means of improving it.
- The establishment of the workload and physical strain equations contribute to the current limited understanding of the interaction between the two as identified within the literature. The equations provide a starting point for further work validating and refining this interaction.
- The development of the trust speed equation and integration into a CAD program provided supporting evidence for the importance of speed in the human robot collaboration (HRC) trust relationship. This contributes to our current knowledge of how the introduction of this relatively new technology may affect those required to work with it.
- The reliability assessment of the integrated trust speed equation in the CAD program provides evidence for the ability to identify psychological equations and include them in DHM CAD programs.

1.3 Thesis Overview

This thesis has been structured in the following order:

Chapter 2 will present the literature review which covers the state of the art of CAD and DHM, including limitations and directions for improvement. It will then progress on to discussing how to integrate psychological processes into CAD, before detailing these psychological processes and the most suitable means of measuring them for creating rules or equations for integration into a CAD program. As the process for developing a mental workload equation required

assessment in relation to physical strain, postural assessment tools and means of capturing motion are additionally covered.

Chapter 3 will present the experiment detailing the assessment of the effectiveness of an inertial motion capture suit. The Chapter will cover details of the methodology, analysis, results and a discussion detailing the impact of the results for the experiment presented in Chapter 4, limitations with the study, and areas for future research.

Chapter 4 will present an experiment identifying the predictive effect of physical strain on mental workload. The methodology will be presented, in addition to the analysis, results and discussion. The results will show three possible equations and will discuss their suitability for integration into a DHM CAD program. The discussion will cover how the results fit in the context of other research, limitations of the study, and the next steps for future research.

Chapter 5 will be split into two parts, the first will present an experiment which will identify psychological equations that could be used to facilitate human robot collaborative interactions. This first half provides the methodology, and results. These results will identify an equation which will be taken forward to the reliability phase. This reliability assessment will be presented in the second part of Chapter 5 where details of the method used and the reliability results will be provided. The final part of this Chapter will be a discussion covering both experiments, and will detail how the results fit in the context of the literature, limitations of the experiments, and areas for future research.

Chapter 6 discusses the findings from across the experiments detailed within this thesis, discussing the progress made towards each objective and how they meet the aim. It will cover the limitations of the research and areas for future research before providing a summary conclusion.

Figure 1-1 illustrates the structure of the thesis, presenting in which Chapter each of the objectives were met.

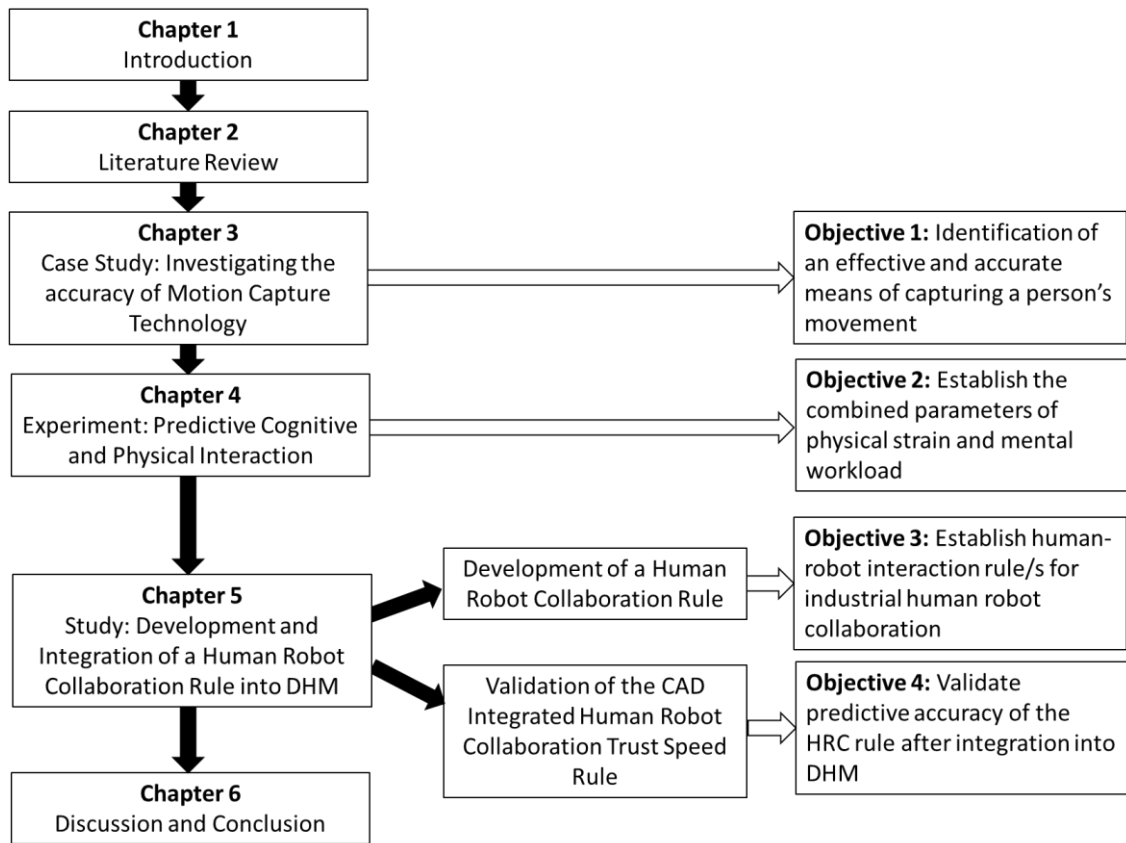


Figure 1-1. Graphical Illustration of the Structure of the Thesis

2 LITERATURE REVIEW

Industry 4, the label given to the newest and current industrial revolution is bringing an increase in digitisation to manufacturing, with the intention of increasing productivity and flexibility within the manufacturing sector. Part of this involves a move towards using the digital human modelling (DHM) capabilities in computer aided design (CAD) programs to design and monitor products, workstations, and new technologies introduced into manufacturing working environments (Peruzzini et al., 2020; Romero et al., 2016; Turk et al., 2021). One aspect of this process is used for proactive ergonomics which involves modelling the impact manufacturing products and work environments will and do have on workers. The intention of this modelling is to identify and predict the effect of new and current processes and improve working conditions and, thereby, productivity. At present, the focus of CAD programs with integrated DHM tools is the physical impact of manufacturing environments on workers; psychological impacts, however, are ignored. The following literature review details the capabilities and limitations of DHM tools available in CAD programs and provides direction for improvement.

2.1 CAD and DHM

There are a range of CAD programs used within manufacturing for product and environmental design and monitoring. CAD programs are used to create digital environments in an attempt to identify and design out any problem areas prior to the final build. CAD programs are mainly concerned with the physical attributes of a design and have no means off assessing the designs impact on the final user. DHM can be an added tool kit to a CAD program, or a standalone program. Where DHM tools are used within a CAD program the CAD design can be changed to improve the design for human interaction. Where DHM is available in a standalone program a CAD design can be imported and assessed but not changed. DHM provides digital tools which enable assessment of how a design may affect the final user.

DHM tools are used to identify how a person may interact with the design, whether it is an environment or a component, and assess the extent to which they will be able to do so without experiencing a negative physiological impact. These programs are meant to holistically model the complex functions of the human body, including both the physical and psychological (Sundin & Örtengren, 2006) and to design out human related issues including performance and/or safety with a simulation or virtual environment (Yang et al., 2004). DHM tools range in complexity from a limited set of human-like manikins with repositionable limbs to human-like manikins with a large choice of anthropometric size options and various activity analysis tools.

The combined use of DHM and CAD is becoming increasingly popular with improvements in digital design technology making virtual environment renderings increasingly realistic (Peruzzini et al., 2020; Romero et al., 2016). This increased capability financially benefit companies as they reduce the number of physical mock-ups that are required to review the effectiveness of a design. Indeed, mock-ups would often be required for each change to a design, which can be costly. Additionally, the software removes the need for large participant pools to interact with the mock-ups in the early stages of the design process. Further aiding this increased use is the developed awareness that poor ergonomics design can have a negative impact on worker health (injuries and illness) (Broberg, 1997) and workplace performance (Wang & Lau, 2013).

The development of CAD programs containing a DHM capability (DHM CAD programs) has included the integration of activity analysis tools which have become a common technique for predicting, identifying and preventing physical risk within manufacturing (Chiu et al., 2012; Cimino et al., 2009; Estefani, 2011; Gomez, 2014; Judt et al., 2016; Lockett et al., 2014). The first type of tool is referred to as postural analysis assessment. It consists of a means of analysing the posture and orientation of body parts to identify and assess poor working postures. The second tool is force analysis and details the extent to which individuals should be able to manually move a product without contributing to the development of musculoskeletal disorders (MSDs). Finally, energy expenditure

is sometimes included in these toolsets. It identifies the likelihood of an individual having the energy to complete a task or the level of fatigue they may experience after completing the task. Effective utilisation of DHM CAD programs in the design stage has meant that companies have seen an optimisation of work conditions and output (Battini et al., 2011; Cimino et al., 2009; Colombini & Occhipinti, 2006), and improvement in worker attitudes and satisfaction (Das et al., 2007; Drury, 2000). However, despite these capabilities, there are limits to what the available programs are capable of modelling.

Not all CAD programs contain DHM capabilities, while some DHM tools available are not integrated into a particular CAD program. Standalone DHM programs require a digital design to be uploaded to the DHM program for analysis, the manikin can be adapted and changed within this environment but often the product or workstation cannot be changed. These tools include SantosHuman (SantosHuman Inc), the Anybody Modelling System (AnyBody Technology A/S), and the System for Aiding Man-Machine Interaction Evaluation (Sammie) CAD (Sammie CAD Ltd.).

DHM CAD programs enable a greater ease of use and effectiveness, for designers as they are able to redesign components within the same program as the DHM assessment is completed. These CAD programs with DHM capabilities enable a more holistic assessment of the suitability of a design before it is constructed and, therefore, are the focus of this literature review. Only two CAD programs were identified with full DHM and activity analysis tools: the Digital Enterprise Lean Manufacturing Interactive Application (DELMIA) application for CATIA (Dassault Systèmes) and Siemens' Tecnomatix Jack (Siemens PLM). The available tools and differences between the two programs are detailed in Table 2-1, with information acquired from online sources including pamphlets from Siemens and Dassault Systèmes, email correspondence with suppliers of the programs, and program trials.

Table 2-1: DELMIA and Jack Comparison

Ergonomic tools		DHM Application	
		DELMIA	Jack
Force analysis	NIOSH Lifting equations	✓	✓
	Snook and Ciriello lifting and lowering tools	✓	
	Torque, load and sheer for tools	✓	
	Lower back analysis		✓
	Force solver		✓
	Manual material handling		✓
	Static strength prediction		✓
Energy expenditure	Metabolic equation	✓	✓
	Fatigue analysis		✓
Task analysis	MTM-1		✓
Postural Analysis	RULA	✓	✓
	OWAS		✓
Anthropometrics	Populations	American	✓ - ANSUR_NHANES and North American AutoWorkers anthropometric databases
		Canadian	✓ - Canadian Land Forces
		French	
		Japanese	
		Korean	
		German	✓
		Taiwanese	
		Chinese	✓
		Indian	✓
	Tools	Clash detection	✓
		Inverse kinematics	✓

2.1.1 The Advantages of CAD and DHM

The introduction of DHM and ergonomic assessment capabilities in CAD enabled the easy variation of anthropometric and biomechanical parameters of virtual manikins (a digital representation of a human) (Paul & Wischniewski, 2012), and activity analysis in a format familiar to designers (Broberg, 1997). This has helped to improve design and usability of products and work systems (Paul & Wischniewski, 2012) by allowing designers to identify the impact of their designs on those required to interact with them. This has included how their designs may be used and how they may cause potential negative physical effects, including the risk of developing MSDs.

MSDs refer to inflammatory and degenerative conditions that affect the nerves, tendons, ligaments, muscles, and supporting structures of the body (Punnett & Wegman, 2004; Waters, 2004). The disorders can include nonfatal and/or non-traumatic injuries to the upper extremities, neck, back, trunk, and lower extremities (Rodrick et al., 2016). These injuries can be caused by bending, twisting, awkward working postures, whole body vibration, materials handling, forceful movements, lifting, and manual handling (moving objects) (Health and Safety Executive [HSE], 2019; Kilbom, 1997; Punnett & Wegman, 2004; Van Nieuwenhuysse et al., 2006; Waters, 2004). These activities are inherent to tasks required in manufacturing and are the reason for the high incidence of work-related MSDs in manufacturing (HSE, 2019). Not only do MSDs have obvious adverse effects on the worker but they also place financial strain on companies, with sick leave, and the possible need to retrain or reskill those individuals who can no longer perform their job as a result of these injuries (Dunning et al., 2010).

DHM CAD programs vary in the number of anthropometric scales and activity analysis tools, as can be seen in Table 2-1. Anthropometrics enable the scaling of digital human representations, using pre-existing anthropometrics data for a specific population group. This ensures that designs take into account the size variations of the population/s that will be using them. An example of this may be ensuring that a newly developed workstation is adapted to suit a specific population range. For example, a workstation designed for a Dutch population

may lead to the development of MSDs for a population that has a comparatively smaller stature, such as Chinese or Japanese populations. Daniell and colleagues (2012) detail the unsuitability of using western anthropometric surveys for informing design decisions for Chinese populations as a result of reviewing local Chinese populations. Therefore, designs need to be made to fit the population that will be using them, or at least be adjustable to suit differing populations. The populations available within DHM CAD programs enable this, to an extent.

Activity analysis tools cover a range of assessment methods that are used to identify the likelihood of a task resulting in MSDs for the individual completing a task. The range of activity analysis tools enables an assessment of the interaction a person may have with the design and the possibility of developing MSDs as a result. This is an advantage in manufacturing environments where conducting some of these analyses in person, such as postural analysis, is not always possible due to visual occlusion of body parts. Within a digital environment, this is not a problem as it is possible to remove parts of a digital structure to view how the manikin interacts with it. A further advantage of using a DHM CAD program is its ability to record and ergonomically assess the activity of the manikin frame by frame (Johnson & Fletcher, 2015). This makes assessment much faster than traditional forms of analysis and allows the designer to identify exact problem areas throughout the interaction with a new design.

2.1.2 The Limitations of CAD and DHM

Although the intention of DHM tools is to be able to model the complex physical and psychological functions of the human body (Sundin & Örtengren, 2006), and despite the advantages of DHM CAD programs, there are, however, limitations with DHM tools. The limitations are found in DHM CAD programs ability to effectively measure physical strain and a complete absence of information on the psychological impact of a design on operators. A consequence of this is that these tools are not able to mitigate all of the negative impacts of new designs on the operators that will be interacting with them.

2.1.2.1 Physical Limitations of DHM CAD and Improvement Options

2.1.2.1.1 Anthropometrics

Firstly, although a range of anthropometric populations are available, as seen in Table 2-1; they are limited in the types of pre-set manikins available. South American and African populations are completely missing from the programs. Although some European populations can be found, there is no inclusion of any populations from the United Kingdom (UK) which is of particularly interest for this thesis, as it is where this research is being carried out. This makes designing for populations that are not represented within these programs difficult. Additionally, the readily available populations within DHM CAD programs are often drawn from military populations, such as the Canadian Land Forces and the Anthropometric Survey of US Army Personnel (ANSUR), which do not represent the typical user populations (Parkinson & Reed, 2010). This means that there is likely to be a design mismatch between the digital manikin and the intended user, and may contribute to the development of MSDs.

Research is being conducted to expand and gather up-to-date anthropometric characteristics from a range of countries including specific working populations: Chilean workers (Castellucci et al., 2019), United States (US) firefighters (Hsiao et al., 2014), Chinese males (Daniell et al., 2012) and Chinese, Japanese, Korean, and Taiwanese individuals (Lin et al., 2004). Although these data sets may be available from the researchers they are unlikely to have been integrated into the DHM programs, which means they are not readily available for designers to use. It should be noted that it is possible to adapt the anthropometric data within both Jack and DELMIA, however, this requires designers to identify and interpret the relevant anthropometric data. This also requires an awareness of how to manually make changes to the manikins, and create manikins that fit the percentiles of interest. This can be a time-consuming process in relation to the ease of use for the preprogramed populations and, therefore, may not be possible due to time pressures and levels of understanding.

2.1.2.1.2 Positioning

There are limitations to the physical positioning of manikins within Jack and DELMIA. These include being able to move the joints of the manikins in ways that people are unable to and a lack of effective clash detection. Indeed, the degrees of freedom in the manikin's joints do not always represent those of humans, making it possible to position a manikin in such a way that would cause pain to a person (Reyterou, 2009). Consequently, it becomes possible to make a task look feasible that would actually lead to the development of MSDs (Helander & Khalid, 2012).

A further limitation is that the clash detection is not effective on some parts of the manikin's physiology. An example of this can be seen in DELMIA v5 where the clash detection for the webbing between the fingers does not exist, however, improvements have been made with DELMIA v6 and clash detection now works in that area. Jack on the other hand has no clash detection. This prevents the accurate representation of true space and obstruction limitations, which would require an awareness of the manikin in relation to the design to ensure that reach limits are accurate and that the manikin is not positioned within a design. If a designer lacks this awareness they may be expecting individuals to be able to interact with a structure in a physically strained way, increasing the risk of developing MSDs.

2.1.2.1.2.1 Vibration

Vibration has been found to contribute to the development of MSDs, and guidelines have been produced by the International Organization for Standardization (ISO). These guidelines cover whole body, and hand and arm vibration exposure. The ISO 10816 (ISO, 2018) guidelines provide regulations regarding exposure to vibration including length of time and vibration frequency. There are currently no measurement of vibration within the DHM CAD programs as can be seen in Table 2-1. This is potentially because vibration can greatly differ across tools and machinery; as a result, it would not be possible to implement a simple equation to account for it. The inclusion of vibration in CAD models, although important, would require extensive work to cover the most

common tools and machines. Therefore, it may not be possible to include within DHM CAD programs.

2.1.2.1.3 Physical Analysis

The assessment tools available for physical analysis within DHM programs often cover postural analysis. Long-lasting static or constrained postures, or dynamic postures with high repeated cycle times and little or no rest within a task (Smith, 1996), which involve bending, twisting and/or awkward working postures all of which can contribute to the development of MSDs (HSE, 2019; Kilbom, 1997; Punnett & Wegman, 2004; Van Nieuwenhuysen et al., 2006; Waters, 2004). Postural analysis tools look to identify whether individuals adopt these specific postures during their work activities.

There are a range of postural analysis methods and tools available which are detailed extensively within literature (Albin, 2013; Bao et al., 2009; Chiasson et al., 2012; Janowitz et al., 2006; Juul-Kristensen et al., 1997; Karhu et al., 1977; Keyserling et al., 2010; Li & Buckle, 1999; Roman-Liu, 2014). Despite this range of tools, the number of postural assessment tools available in the DHM CAD programs of interest is limited to two: Rapid Upper Limb Assessment (RULA) (McAtamney & Corlett, 1993) and Ovako Working Posture Analysis System (OWAS) (Karhu et al., 1977), as can be seen in Table 2-1. RULA was designed to assess full body postures with a focus on the upper body in workplaces where work-related upper limb disorders are reported (McAtamney & Corlett, 1993). Consequently, it focuses on the upper extremities, head and neck, and ignores extension of the back and leg bending, along with unilateral weight distribution. The OWAS assesses the flexion and extension of the neck and only flexion of the back not its extension. The arms are assessed but not the lower arms nor wrists, and the legs are fully assessed including unilateral weight distribution (Karhu et al., 1977). The lack of whole body assessment in RULA and OWAS, makes these two methods unsuitable for assessing all of the types of postures adopted in manufacturing (Chiasson et al., 2012; Pillastrini et al., 2007), and particularly aerospace manufacturing where extreme postures requiring back and neck extension are seen (Pillastrini et al., 2007). Therefore, the tools available

within Jack and DELMIA may be suitable for some manufacturing environments but are unsuitable for aerospace manufacturing and may produce evaluations that do not take into account the whole risk for the operator and, consequently, leave key physical costs missed.

The limited number of postural analysis tools is of concern because poor working postures have been identified as a contributing factor to the development of MSDs (Punnett & Wegman, 2004; Roman-Liu, 2014; Van Nieuwenhuysen et al., 2006), and are of particular interest within aerospace manufacturing, for addressing extreme postures. A solution to this would be to include whole body assessment methods such as the Rapid Entire Body Assessment (REBA), into DHM CAD programs, which accounts for these postural positions. REBA would be more appropriate than other methods for aerospace manufacturing as it assesses the awkward postures often observed in this industry. These postures include trunk flexion and extension, knee flexion and unilateral weight distribution, flexion of the upper arms, lower arms, and the wrists, and flexion and extension of the neck (Hignett & McAtamney, 2000). REBA was designed to fill the gaps that RULA left, with the aim of assessing unpredictable postures in work places where the entire body was likely to see strain (Hignett & McAtamney, 2000). Research has supported this aim showing that REBA can score and identify awkward postures that other methods of postural assessment miss or do not rate as highly as expected for risk of MSD development (Chiasson et al., 2012). This indicates that REBA would be the most suitable tool for DHM in aerospace manufacturing environments.

A limitation of postural assessment methods is that they quantify and score through weighting and magnitudes of interactions which are largely hypothetical (Kilbom, 1997; Li & Buckle, 1999; Winkel & Westgaard, 1992). This means that it has not yet been possible to objectively assess whether the boundaries between scores accurately account for the damage resulting from poor postures. For this reason, the methods usually assess risk of MSD and not absolute indicators of damage to a person's musculoskeletal system (Roman-Liu, 2014). Although research has found a link between posture and MSD, it has not yet been possible

to link particular postures to specific damage (Smith, 1996). Objective measures of working conditions would aid in the development of a more effective understanding of the interactions between poor working postures and MSDs.

2.1.2.1.4 Summary of DHM CAD Physical Limitations and Improvement Options

The limitations detailed within this section can be addressed through an update of the software programs and integration of readily available data or tools. This would include the integration of additional analysis tools, more comprehensive up to date, and varied anthropometric populations, and more valid and reliable limits on body positioning. If these changes are not integrated then an awareness of these limitations helps to mitigate against them, and put in place alternative solutions including: manually inputting anthropometric data, visually assessing postures using whole body tools where appropriate, and editing the joint limits within the DHM CAD programs.

2.1.2.2 Psychological Limitations of DHM CAD

Although there are options for improving the physical aspects of DHM CAD capabilities, the lack of any psychological assessments are more difficult to address. Within ergonomics, the assessment of the psychology and mental processes of operators focuses on how a situation is appraised and an individual's associated reactions to humans and other elements within a system. These processes can include factors such as perception, memory, information processing, reasoning and motor response as they affect interactions among humans and other elements of a system (Karwowski, 2012; Wogalter et al., 2012). They can additionally include factors such as beliefs, traits, and states or moods. When in any situation an individual progresses through each of these processes, from isolating and taking in information using the senses in the attention stage, to perceiving and interpreting information at the perception stage. Decisions regarding how to react to this information is the filtered through memory, reasoning, current state or mood and different forms of information processing before actions are decided which may lead to a motor response (Eysenck, 1983). Individuals have limits on their capacity within each of these processes, and when capacity is achieved, they contribute to errors and

mistakes. To avoid these errors and mistakes, which can be costly for a company, the intention is often to ensure that the individual works at an optimal level. This is often identified through cognitive and psychological assessment tools. Although tools and models are available for assessing the psychological impact of work on the individual, they have not yet been integrated into DHM CAD programs.

The lack of any psychological assessment methods within CAD becomes a problem where tasks or workstations are redesigned to mitigate MSDs but contribute to greater psychological strain or workload. Additionally, where work systems become more complex, particularly with the move towards integrating more technology into manufacturing, a means of predicting the likely associated psychological impact is required. Within aerospace manufacturing, the ability to assess this psychological impact would be greatly advantageous as new and novel technology is now being designed in such a way that it can be introduced to this sector, where it was not able to before. Providing design engineers with a means of assessing an individual's psychological response to what has been designed, and the impact on an individual's cognition enables them to make more effective design decisions and reduce the negative impact for human operators. The availability of psychological response data integrated into digital platforms would additionally benefit digital based task design and monitoring programs.

Computer models of cognition include programs such as: the IMPRINT (Improved Performance Research Integration Tool) (Archer & Adkins, 1999) which allows the modelling of tasks and human performance analysis. The ACT-R (Adaptive Control of Thought-Rational) cognitive architecture (Anderson & Lebiere, 1998) which models the steps of cognition through a production rules model that coordinates retrieval of environmental and memory information (Laughery et al., 2012; Leiden et al., 2007). Another project designed to model cognition and behaviour is AMBR (Agent-Based Modelling and Behaviour Representation) which was designed to examine multitask management, dynamic priority setting, and attention management, with the particular intention to identify operator overload and inform effective information management strategies (Tenney et al.,

2006). These tools are difficult to gain access to for use because they are often designed for military use and have not been integrated into DHM CAD programs. Reasons given for this are that they are not yet sufficiently accurate and, consequently additional analysis is required to effectively assess cognition (Laughery et al., 2012). Duffy (2012) explains that an additional reason for this lack of integration is the form of information presentation, where DHM uses visualisations to present information, human performance modelling does not and, therefore, the information does not lend itself to easy implementation into DHM CAD programs for design engineers, and would require extensive training.

If psychological assessments can effectively be established and integrated into DHM CAD programs, they would provide design engineers with a means of holistically assessing the impact of their designs. Additionally, its inclusion would financially benefit companies who would not need to pay for remedial redesigns. This information could additionally be utilised in software programs used to design and monitor tasks. Therefore, the aim of this thesis is to identify whether it is possible to effectively integrate human psychological into DHM. This integration would improve the capability of digital human modelling for manufacturing system design.

- **Aim:** Identify and establish psychological parameters that will improve the capability of digital human modelling for manufacturing system design

2.2 Introducing the Psychological to DHM CAD Programs

This section intends to identify the most relevant psychological constructs for integration into DHM CAD programs. Not all psychological constructs can be addressed within this thesis due to the broad nature of psychology. In cognitive engineering and human factors, there is a specific focus on how mental processing is affected by complex systems (Parasuraman et al., 2008), particularly within manufacturing. The ability to understand and predict the effect of complex manufacturing systems on an operators mental abilities, particularly when changing a task or introducing new technology, is vital for ensuring the introduction or adaption of safe and effective systems (Young et al., 2015). In many cases, new technology or task design changes intended to improve working

conditions, reduce or alleviate the development of MSDs, or decrease cognitive demand, have been found to do the opposite. Therefore, this section will look to explain the different changes adopted within manufacturing to reduce MSDs and, for each, assess what psychological aspects are relevant. Historically, automation has been introduced to reduce the physical load on operators, and has led to unforeseen increases in mental workload in operators and contributed to higher and costly error rates (Parasuraman & Riley, 1997). Finding a ways to be able to model these psychological reactions during the design stage provides the opportunity to prevent unintended negative consequences. The following subheadings of this thesis will look to identify which psychological measures might be more suitable for integration into DHM CAD programs.

Aerospace manufacturing currently requires the completion of highly physical craft-based tasks and, as a result, has high instances of MSDs. Design and research engineers look to address this in two main ways with design changes to the physical component being worked on or with the introduction of robotics or automation (Judt et al., 2016). Where a component is changed to address MSD risk, the task the operator is required to complete may change, this is likely to result in changes in mental workload for an operator. As a result, mental workload has been identified as an appropriate cognitive psychology measure for introduction into DHM CAD.

Workload is an overarching cognitive psychology construct and is defined as “the cost incurred by a human operator to achieve a particular level of performance” (Hart & Staveland, 1988 p. 140), with mental workload a dimension of workload. Mental workload is associated with the amount of mental and perceptual activity required for a task (Hart & Staveland, 1988). Workload is complex with people having multiple processing resources, each with a limited capacity. When these resources are exceeded there are two levels of effects: the first are acute effects which are the more immediate and direct effects upon a task, such as mistakes, or errors (DiDomenico & Nussbaum, 2011). The second are chronic effects which are the more long-term results of managing high workload and not only see an impact on performance but also on health and well-being (Hockey, 1997).

The inclusion of mental workload within DHM CAD programs is further justified by the finding that excessive mental workload placed on the human operator is one of the greatest contributors to human error in many systems (Laughery et al., 2012). Mitigation of this impact is required, particularly in aerospace engineering as it is a safety critical industry and errors can have catastrophic consequences. Providing design engineers with a means of assessing workload during their decision-making process, regarding the design of a workstation or structure that an operator will be working on, may reduce the need for additional post-design assessments and remedial actions. Additionally, this information can be to design operators' tasks and to predict performance identifying when breaks or job rotation would be best implemented. To do this, effective means of measuring mental workload required identification and a method to integrate it in DHM CAD programs.

The second way that risk of MSD development is being addressed within aerospace manufacturing is through the introduction of large high-payload collaborative robots. This is a change in manufacturing from previously dangerous robots that were secured behind physical barriers for safety, to the introduction of robots that can safely work in proximity to operators without the need for the same level of separation. These robots that can safely work in closer proximity to operators, or collaborate with operators on their tasks are referred to as collaborative robots. The interaction between a human operator and collaborative robot is generally referred to as human robot collaboration (HRC). It is expected that HRC will have a psychological reaction for those operators required to work with robotics (Arai et al., 2010; Grahn et al., 2018; Koppenborg et al., 2017; Parasuraman & Riley, 1997). Therefore, the ability to understand and address human psychological reactions to the integration of collaborative robotics will be investigated within this thesis. By including this information into DHM CAD programs integrators will be able to test and set up systems that promote faster adoption and use of collaborative robotics. Furthermore, where tasks are managed by programs this information can be used to design work tasks and predict operator capacity for effectively completing their work. To do so, the most suitable measures for experimentation to develop psychological equations that

can be integrated into DHM CAD will be explored below. The criteria for these measures included those that were not solely related to a task. This is because the task is not always known during the design process and would, therefore, make the tool unusable for some processes of design or redesign. These measures had to have a numerical output as this enables the ability to measure the change on the relevant dependent variables (DV) as the independent variables (IV) are manipulated. The resulting application of statistical analyses, whether regressions or general estimating equations, can then be utilised to identify whether a relationship exists between the identified factors, and what the resulting equation looks like. These equations can then be integrated into a DHM CAD program. The following two subheadings (2.2.1 and 2.2.2) provide details of current research and the direction taken by this thesis to identify, isolate, and predict mental workload and perceptions of collaborative robotics for introduction to DHM CAD programs.

2.2.1 Mental Workload

Mental workload has to do with the cognitive processes and relationships between task demands, the psychological resources, and the effort available to meet these task demands (Parasuraman et al., 2008; Tattersall, 2000). The main determinants of mental workload can be characterised into two types of demands: task based, and processing based. Task demands include task difficulty, task priority, and situational factors. Task demands have an effect on the level of arousal (wakefulness) for an operator, at an intermediate level they provide enough stimuli to ensure optimum performance, however, if these conditions sit at either end of the arousal spectrum, a decrease in performance is seen (Yerkes & Dodson, 1908). Whereas the supply of attentional or processing resources relates to functions that support information processing such as perceiving, updating memory, planning, decision-making, and response processing. The focus of this literature review and thesis is on overload, where arousal sits above the optimum threshold, rather than underload where it sits below.

As mentioned, computer models of cognition and workload such as IMPRINT and ACT- R are not readily available and their data is not easily understandable by

design engineers. To integrate a mental workload assessment into DHM CAD programs, a way of predicting workload changes required identification. IMPRINT and ACT-R assess cognitive capacity based on task-based assessments. This is not suitable where a redesign takes place and the associated task is not yet fully known, or where the tasks may have a level of flexibility to them. Furthermore, the intention was to associate cognitive changes with a process already measured within DHM CAD programs, as a means to more easily introduce the workload assessment to current DHM CAD programs.

A process that is already measured in DHM CAD programs is physical strain through postural assessments; therefore, research was conducted into whether interaction effects had been found between mental workload and the physical strain associated with poor working postures. The intention of this was to identify a theoretical foundation for associating changes in physical strain with a measured change in mental workload. The reason for focusing on assessments of postural analysis is because this type of assessment is well known and used by design engineers (Judt et al., 2016). Therefore, it provided a good reference point for use. Within manufacturing, and specifically aerospace manufacturing, most tasks operators are expected to complete are physically and cognitively demanding. This means that measuring mental workload based on physical demand would be suitable for most tasks completed within this field, and overcomes the need for a task-based mental workload modelling as seen with current cognitive models.

There is a small but important body of literature that has researched the concurrent effect of physical and psychological strain (Mehta & Parasuraman, 2013) and found that working in physical strain has an effect on mental workload capacity (DiDomenico & Nussbaum, 2008; Marras & Hancock, 2014; Shaikh et al., 2012). Within the literature the reason for this is because they draw on the same underlying energy resource, which as with all energy resources will be finite until refilled through rest or refuelling (Kahneman, 1973; Norman & Bobrow, 1975; Van Acker et al., 2018). Mehta and Parasuraman (2013) argue that further work is required to fully understand this interaction effect as it is critical to ensuring

work is organised in such a way that operators are able to work effectively, particularly where expected to utilise both resources. For example, Marras and Hancock (2014) detail the theoretical effect of a body's biochemical or physical state and how it impacts information processing capacity. Their paper suggested that physical strain has a negative effect on cognitive capacity, reducing it not only due to a need to manage discomfort while maintaining performance on a task, but additionally, over a longer term because of fatigue. Shaikh et al. (2012a; 2012b) completed research looking at the impact of physical stressors on workload. They compared working with arms above shoulder height to working in a more physically neutral posture and found that working with higher arms (high physical strain) leads to significantly higher mental workload as measured with the National Aeronautics and Space Administration's Task Load Index (NASA-TLX). DiDomenico and Nussbaum (2008) found similar significant effects of physical strain on the Mental Demand workload dimension of the NASA-TLX. Their study required participants to lift and place boxes of varying weights (no load to high load), while completing arithmetic equations. Perceived Mental Demand was greater with higher loads. These findings indicate that it should be possible to find a predictive effect of physical strain on mental workload, which could be modelled and integrated into CAD DHM programs. The results would additionally contribute to improving the understanding of the interaction between cognitive capacity and physical strain outlined by Marras and Hancock (2014).

To effectively establish whether there is a measurable interaction between physical strain and workload, the appropriate methods of measuring both are required. The following subheading details measurement techniques for both including identification of suitable measures for the experiments detailed within this thesis together with the research considerations required prior to experimental progress.

2.2.1.1 Measuring Mental Workload

Within the literature, there are three distinct means of measuring mental workload: through performance, subjectively through questionnaires, and

objectively through physiological measurements. Each of these methods will be detailed and their suitability for achieving the aim of this thesis explored.

2.2.1.1.1 Performance Measures

Mental workload assessment looks to identify whether individuals are working within tolerable limits, in terms of their information processing capacity. Performance measures assess mental workload through identifying performance changes as demands change. These include, amongst others, the number of errors, and the length of time taken to complete a task. Individuals have a finite capacity for dealing with task demands, so if the number of errors or time to complete a task increase from one condition to the next, it is assumed that mental workload has increased for the participant. If the task demands continue to rise it can eventually lead to the individual being unable to complete their task. When a single task is assessed like this, it is referred to as a primary task assessment. There are limitations with this type of assessment, individual factors have an effect on the extent to which an individual will choose to perform to their full capacity. Although individuals do have a finite capacity to deal with the demands of a task, they have been found to protect their cognitive resources which makes identifying mental workload capacity difficult (Mehrabian, 1977; Tsang & Vidulich, 2006).

An experimental means of identifying excess capacity is with a secondary task measure. A second task is added to the primary task, and the individual completing the tasks is asked to maintain performance or to work to the best of their abilities on the primary task. Therefore, as demand changes in the primary task, performance is expected to change in the secondary task (Tsang & Vidulich, 2006). Mental workload is then measured based on performance changes, with capacity being understood based on secondary task decrements. When using this technique, care has to be taken to ensure that the same cognitive resources are drawn on in both the primary and secondary task. Multiple resource theory explains that individuals have a number of processing resources or channels which are each limited in their capacity (Wickens, 2002). These resources relate to:

- verbal or spatial information coding
- visual and auditory perception
- cognitive activity or processing (memory and decision making)
- response strategies (manual and verbal)

If both primary and secondary tasks do not draw on the same resource, no or limited interference will be found in the performance measures for the primary task (Allport et al., 1972). Tasks interfere to the extent they compete for the same resources, with the resultant outcome being performance decrements or increased task time.

A difficulty with these measures is that changes in performance may be a result of changes in an individual's motivation which can see a shift in performance either in the negative with an increase in errors and time, or positive with a decrease in both. Kahneman's (1973) and Hockey's (1997) research detailing the role of the compensatory control model shows that individuals have a capacity to protect performance despite demands. Additionally cognitive supply is modulated by individual differences including skill and expertise, which can be used to protect performance. If an individual is motivated to protect their performance, an accurate or measured change in mental workload capacity is unlikely to be found. As a result, performance measures are usually supported with surveys to gain a subjective understanding of the mental workload experienced by the individual. Although performance measures provide a numerical output, they are associated with time or number of errors, therefore, are task specific, and so are not suitable for use when looking to establish an equation that can be used to model mental workload within DHM CAD programs.

2.2.1.1.2 Subjective Measures

Subjective measures require individuals to quantify their experience of workload, usually on scales after a task has been completed. There are a range of scales available for assessing mental workload, however there are two predominant scales that have been shown to have high validity and reliability (Colle & Reid, 1998; Nygren, 1991; Tsang & Vidulich, 2006; Young et al., 2015) they are:

- NASA-TLX (Hart & Staveland, 1988)
- Subjective Workload Assessment Technique (SWAT) (Reid & Nygren, 1988)

The NASA-TLX is comprised of six scales used to measure six principal dimensions of workload: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. The Mental Demand dimension is a measure of mental workload. Individuals being assessed are presented with a scale for each dimension and asked to mark the most suitable point from very low to very high, or perfect to failure depending on the dimension. It can be used in either its traditional or raw form. The traditional form involves weighting the workload dimensions based on their relative importance before rating the scales, while in the raw version the weighting process is not completed. The use of the raw form has been validated as effective as the traditional form as a measure of workload (Bittner et al., 1989; DiDomenico & Nassbaum, 2008; Hendy et al., 1993; Moroney et al., 1992; Nygren, 1991).

The SWAT workload assessment uses three scales to assess workload: Time Load, Mental Effort Load, and Psychological Stress Load. Mental Effort Load is a measure of mental workload. The first step of the assessment is the scale development where cards containing the dimensions are sorted into rank order reflecting an individual's perception of increased workload (Rubio et al., 2004). Based on the rankings, a conjoint analysis is used to produce a scale with intervals. After completing a task, workload is assessed using these scales, this is referred to as event-scoring. Finally, each of the three dimension ratings are converted to a numeric score between 0 and 100 using the interval scale developed at the beginning of the process (Tsang & Vidulich, 2006). As with the NASA-TLX, research has found that a simpler form of the SWAT can be effectively used to assess workload. This simplified version requires individuals to score the three SWAT dimensions, and has been found to be as effective at measuring workload as the full version (Luximon & Goonetilleke, 2001; Verwey & Veltman, 1996).

Although subjective assessments provide an experiential measure of an individual's workload, critiques have been levelled against them particularly regarding the data collection process. Subjective measures are usually completed at the end of a task and there is some concern that primacy and recency effects may influence the results. This is where an individual more readily remembers something from the beginning (primary) and or end (recency) of an event rather than in the middle. Therefore, there is a concern that an individual will be unable to effectively assess their workload levels as a whole throughout an activity. Research suggests that so long as assessments are administered immediately after trial completion, these effects should be minimised (Tsang & Vidulich, 1994). An additional way to overcome these effects is using physiological measures (see Section 2.2.1.1.3), which can be measured in conjunction with the subjective measures.

Identification of the most effective measure to be used within this thesis to establish a predictive equation for workload is required. Comparisons have been conducted between the two methods and research has found that although both instruments show similar sensitivity, the NASA-TLX has been shown to be slightly more sensitive when looking at dimension validity (Hill et al., 1992; Nygren, 1991; Rubio et al., 2004). This suggests that the NASA-TLX is the more appropriate method as this thesis is looking to measure mental workload and, therefore, the Mental Demand dimension will be used within the experiments detailed within this thesis. Furthermore, Nygren's (1991) review of both assessment methods found the NASA-TLX to support a general predictive model for experienced workload whereas the SWAT was more sensitive to individual differences. As this thesis is looking to establish a model for integration into DHM CAD programs, the development of a general model using the NASA-TLX is the more suitable option.

2.2.1.1.3 Physiological Measures

There are three main physiological measures used to measure mental workload: cardiovascular, ocular, and measures of brain activity. Under each of these three measures are different ways of assessing mental workload. Each measure assesses an individual's level of arousal with an increase of arousal being

associated with an increase in mental workload. A discussion of each of these three types of measures is detailed below.

2.2.1.1.3.1 Brain Activity

There are a number of ways in which brain activity can be measured to identify levels of mental workload. Some of which are suitable for use in applied settings and others are not, such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), which both use large machines that require an individual to be inserted into them and do not allow for free movement while the scan is taking place. Therefore, they are unsuitable for use within the experiments detailed within this thesis.

Another measure of brain activity is electroencephalography (EEG), which is a method of recording the electrical activity of the brain. It produces spectral and event based measures. Mental workload can be measured against frequency and amplitude changes. Amplitude changes are measured using event related brain potentials (ERPs) (Berka et al., 2007). Mental workload has been found to be measurable in both frequency and amplitude changes in brain waves measured by EEGs (Berka et al., 2007; Henelius et al., 2009; Wilson et al., 1994). Both the frequency and ERP measures are susceptible to motion artefacts and electrical noise (Vidulich & Tsang, 2012). Artefacts are errors in the data, and motion artefacts are a result of the person moving during data collection. Electrical noise, on the other hand, is caused by interference from electrical devices working in proximity to the sensors or recorder. This susceptibility makes EEGs unsuitable for experiments where participants will be moving, including in applied settings. This, in addition to a lack of access to the technology and software, makes EEG recordings unsuitable for use within this thesis.

One technique that is more suited to an applied setting where individuals are required to complete physical tasks is functional near infrared spectroscopy (fNIRS), as it does not restrict the individual wearing the device (Foy & Chapman, 2018). FNIRS uses near infrared light to record changes in the concentration of oxygenated and deoxygenated haemoglobin from a resting baseline measure. FNIRS has been used to successfully measure mental workload by assessing

these changes in oxygenation in the prefrontal cortex (PFC) (Causse et al., 2017; Foy & Chapman, 2018; Molteni et al., 2008). However, a lack of access to this type of device has meant that it was not possible to use this measure of mental workload within this thesis.

2.2.1.1.3.2 Ocular

Measurement of various types of ocular activity have been correlated with mental workload changes. This includes changing blink rates, pupil diameter, and length of gaze fixation, which are all recorded using eye tracking glasses or electrodes attached above and below the eyes. Research indicates that blink rate is context dependent and varies both due to mental workload and the type of information being processed. Higher blink rates have been found to be associated with higher mental workload (Bucks et al., 1994; Veltman & Gaillard, 1998; Wilson et al., 1994; Yang et al., 2012), although, when mental workload increases but the task is visual, blink rates have been found to decrease (Wilson et al., 1994). Pupil diameter has also been found to be sensitive to mental workload changes, increasing with higher cognitive demands (Recarte & Nunes, 2000; 2003). However, care has to be taken to ensure no change in brightness levels have occurred during the experimental conditions as this interferes and impacts pupil diameter (Recarte & Nunes, 2000). Finally, gaze fixation duration has been found to decrease with increased mental workload (Foy & Chapman, 2018; Recarte & Nunes, 2000; Yang et al., 2012). Although access to eye tracking glasses was possible within the research conducted in this thesis, their inability to capture the use of peripheral vision made them unsuitable. Some overhead experimental tasks used within this thesis may have seen participants using their peripheral vision which would have led to data loss.

2.2.1.1.3.3 Cardiovascular

Electrocardiograms (ECG) are used to measure the cardiovascular system and specifically electrical activity of the heart (Mulder et al., 2004). An ECG provides the heart rate (HR) signal of an individual over a period of time, this includes the number of HR peaks, and provides information on heart rate variability (HRV) which can be referred to as the inter-beat intervals (IBI). IBI is the time between

HR peaks, therefore, as the rate of HR peaks increase the IBI decreases. This data can be analysed in the time or frequency domain. Frequency data is transformed using filters and has time requirements for recording, often recording needs to be over 5 minutes. These transformations are referred to as low frequency, mid frequency and high frequency (HF). High frequency is also referred to as respiratory sinus arrhythmia (RSA) and requires a measurement of respiration to validate HR data (Miyake et al., 2009). Time domain measures include IBI, and standard deviation of normal-to-normal intervals (SDNN), normal to normal intervals are another name for IBIs. The third time domain measure is: root mean square of the differences between consecutive inter-beat intervals (RMSSD) (Henelius et al., 2009; Miyake et al., 2009).

HRV is generally used rather than HR within research because it has better statistical properties and is, therefore, more useful for analysis (Bucks, 2001). When looking to identify the most suitable HRV domain for use, data collection time and the ability to record respiration has to be considered. Low frequency, SDNN, and RMSSD data require a minimum data collection of 5 minutes (Henelius et al., 2009; Shaffer & Ginsberg, 2017). HF/RSA and IBI both require a minimum of one minute of data, although as already stated the HF/RSA data has to be validated with respiration measures (Shaffer & Ginsberg, 2017).

ECG experimental procedures involve the cardiovascular system being recorded in each condition in an experiment and then compared to identify differences. During task performance individuals expend mental effort which has been found to be reflected in HR and HRV, as mental effort increases, HR has been found to increase and consequently HRV decreases (Henelius et al., 2009; Miyake, 2001; Miyake et al., 2009; Mulder et al., 2004). Veltman and Gaillard (1998) measured mid and high frequency, and IBI HRV data from pilots in a flight simulator at four levels of manoeuvring/pursuit difficulty. In a comparison between the three measures, they found IBI to be more diagnostic compared to the frequency domains. Henelius and colleague's (2009) research supports this, they found IBI to be the best performing HRV metric for classification ability when assessing cognitive demand at different task load levels. These studies indicate that the

most suitable HRV measure for developing an equation of the interaction between physical and mental workload is the IBI metric.

A limitation of using HRV to measure cognitive demand is that its results have been found to be susceptible to physical strain (Backs et al., 1994; DiDomenico & Nussbaum, 2011). However, Tattersall and Hockey (1995) suggested that this effect was minimised if physical workload was consistent. DiDomenico and Nussbaum (2011) were unable to discern whether manipulating physical strain affected mental strain using HRV and, as a result, recommended the use of a subjective measure in conjunction with a physiological measure. Veltman and Gaillard (1998) who suggest the use of multiple techniques to measure mental workload echo this recommendation. By using multiple measures, differences in the results from the different measures may reveal areas where energy is being conserved rather than fully utilised, or identify the activation of Hockey's compensatory model (1997). Furthermore, this also helps to overcome some of the limitations inherent in each of the techniques.

2.2.1.1.4 Summary

HRV, and particularly IBI, will be assessed within this thesis alongside the NASA-TLX. The reason for doing so is to counter limitations of primacy and recency effects, as HRV is measured across the whole task and provides an objective measure of mental workload. The NASA-TLX is measured because it provides a subjective understanding of the participant's perception of the task and the Mental Demand dimension is directly associated with mental workload. Furthermore, it was the only effective objective measurement device available within the laboratory. While eye-tracking glasses were available and could have been used to record ocular data they do not capture the use of peripheral vision which has led to data loss in previous use. As mental workload will be measured against physical demand which will be manipulated, there may be an interaction effect on the HRV measures. However, it is intended that the physical demand will be consistent within each condition and it is expected that this effect will be minimised.

2.2.1.2 Measuring Physical Strain

As physical strain was identified as the most suitable factor to predict mental workload against, two things were required: a means of measuring, recording or capturing physical strain, and a means of quantifying it. The intention was to use pre-existing methods of quantifying physical strain found in DHM CAD programs, which are limited to those identified in Table 2-1. The context for this thesis is aerospace manufacturing and the postures observed in this industry tend to be extreme. As has been identified in Section 2.1.2.1.3, the postural analysis tools available within DELMIA and JACK are not suitable for application in this sector. Although RULA, which is the postural analysis tool in DELMIA, is not suitable for use within the context of this thesis, REBA is suitable because it accounts for the extreme postures observed within aerospace manufacturing. REBA is an extension of RULA, with both scoring postures in a similar way. They both identify the angles between body segments to classify the level of risk (Klippert et al., 2012; Hamilton et al., 2011). Therefore, as RULA has been integrated into DELMIA, it stands to reason that it would be possible to integrate REBA too. As a result, REBA was identified as the most suitable measure for assessing the postures captured within the experiments conducted within this thesis. To be able to assess the postures to identify levels of physical strain a means of capturing and recording posture required identification. The means of doing this is motion capture (mocap) of which there are various forms and the most suitable for the experiments conducted within this thesis required identification.

2.2.1.2.1 Motion Capture

Traditionally, postural analysis is conducted via observations where an assessor visually identifies the worst posture based on their knowledge of the scoring system in use and scores it. Although inter-rater reliability has been shown to be acceptable when using these methods (Bao et al., 2009; Chiasson et al., 2012; Dockrell et al., 2012; Keyserling et al., 2010), the assessor relies on memory and attention, which adds a level of subjectivity to the scoring and, consequently, reduces the reliability of the method. A means of countering the limited reliability has been the use of video recording to capture a person's motion and their postures (Chiasson et al., 2012; Li & Buckle, 1999). This allows assessment from

the same viewpoint by two or more assessors; two cameras are usually used to prevent bias caused by visual perspective. Multiple cameras may be used if the individual being assessed is working in a visually obscured environment. However, this method is not always effective within manufacturing environments where an operator may adopt postures in dynamic, occluded and confined working conditions. The result of which, is that part of their body may be completely out of sight. A key consequence of this occlusion is that estimates about the exact position of body parts may be required (Chang & Wang, 2007), and critical postures may be missed (Klippert et al., 2012).

The use of mocap technology to record a person's motion and postures would increase the reliability of postural analysis (Dockrell et al., 2012) and the identification of physical strain. There are two systems for capturing human motion: non-optical and optical (Hamilton, et al., 2011). Non-optical systems use mechanical, magnetic, or inertial systems, whereas optical systems use cameras and in some cases markers attached to the body. The reason for their effectiveness is because they provide a digitised recording of the motion of an individual which enables motion data to be uploaded into DHM CAD programs. This enables ergonomists and design engineers to take advantage of the DHM and human activity analysis features within DHM and for virtual environments to be used to investigate the impact of current designs on postures (Johnson & Fletcher, 2014). This can be a significant advantage to design engineers particularly as it would address some of the positioning limitations mentioned in Section 2.1.2.1.2. This form of analysis would enable faster activity analyses than currently possible; all frames can be scored by the DHM CAD program and areas of high risk of MSD can be identified. However, suitable tools would need to be introduced that cover whole body assessments, such as REBA, to ensure they are effective for assessing all manufacturing activities. Moreover, where this analysis and data capture is accurate enough it may provide objective rather than correlational insight into the factors contributing to the development of MSD.

2.2.1.2.1.1 Optical Motion Capture

Marker-based optical systems use cameras and in some cases markers attached to the body. These markers are used by the computer program to better identify the body part being recorded and helps to prevent data loss due to visual limitations. One such system is the Vicon system (Vicon Motion Systems Ltd UK), which is considered the gold standard in mocap. This is because it uses cameras and specialised software that accurately pick up the x-, y-, z-coordinates of the markers placed on the individual being recorded (Dutta, 2012) and are, therefore, able to more accurately capture the person of interest. An alternative optical system that is increasingly being used to capture postures is a markerless system, specifically Microsoft's Kinect (Bonnechere et al., 2014; Chang et al., 2011; Schmitz et al., 2014; Xu & McGorry, 2015). Optical systems like traditional cameras are limited because they required line of sight for effective recording. They are, therefore, unsuitable for environments or experiments where line of sight cannot be ensured, which are used within the experiments in this thesis.

2.2.1.2.1.2 Non-Optical Motion Capture

Non-optical mocap provides a 360° interactive visual output of the individual being captured which resolves the perspective and occluded view problems seen with optical forms of mocap. There are a range of non-optical systems available, one of which is the magnetic mocap system like those supplied by the Ascension Technology Corporation (Waterloo) which do not require line of sight (Hamilton et al., 2011), and enable the capture of occluded postures. Electromagnetic systems involve the placement of sensors tethered to a computer attached to the limb segments of interest on an individual. As the sensors move through field lines created by a transmitter, a computer is able to map the location and movement of the sensors (Hamilton et al., 2011). Although there is no occlusion problem with this method, the individual is tethered to a computer via the attached sensors (Dyer et al., 1995) which may encumber movement. Furthermore, they are highly susceptible to magnetic and electronic interference and require carefully controlled lab environments for effective use (Dyer et al., 1995). This makes them unsuitable for capturing data where it is not possible to reduce ferrous metal and electrical interference, as is the case within manufacturing

environments. These limitations made this system unsuitable for use within the experiments detailed within this thesis.

Mechanical mocap systems, such as the Gypsy-7 (Animazoo), use exoskeletons attached to the body, so that as the individual moves, so do the articulated parts which measure the relative motion. These suits do not require line of sight, however, they can be very bulky. As a result, they were deemed impractical for use, particularly as the tasks within the experiments detailed in this thesis required some movement in and out of a structure with limited space.

Inertial systems use sensors to capture data, which are referred to as inertial measurement units (IMU). Each IMU contains gyroscopes and magnetometers which measure positional data of the body segments of the individual wearing the mocap suit or individual sensors. This positional data is transmitted wirelessly using a transmitter to a computer where biomechanical models and sensor fusion algorithms are then applied to the captured data. The application of these models and algorithms enable data to be viewed in real time and recorded. The sensors are small, which reduces interference in movement, and can be attached to a Lycra suit that can be worn over clothing, or alternatively sensors can be held in place with wrapping. As this system is non-optical, movement can be captured in occluded conditions which is an advantage over optical systems. This makes inertial mocap suitable for capturing motion and postures in manufacturing environments. As a result, inertial mocap suits are seeing use in research within manufacturing environments for capturing working postures (Ferguson et al., 2011; Fletcher et al., 2018). These advantages made an inertial system most suitable for use within the experiments detailed within this thesis. However, as will be explained below, this technology was relatively new and questions regarding its accuracy required investigation prior to use.

2.2.1.2.1.3 Inertial Motion Capture Accuracy

There are known factors that limit the accuracy of inertial mocap suits, they include errors associated the type of joint being captured by a sensor, and environmental factors. These errors are referred to, within the literature, as artefacts, and cover all errors introduced to recorded data. There are four specific

types of artefacts associated with inertial mocap. The first are clothing artefacts, which are caused by the sensors moving with clothing rather than the body. The second are tissue artefacts these are where there is no bone for the sensor to be secured against and, therefore, the sensor moves without its segment moving. This is particularly an issue on the thighs and upper arms where there are limited solid spaces to securely hold a sensor (Cappozzo et al., 1996; Cutti et al., 2005). The third type of artefact are skin artefacts, this is where the sensor moves on loose skin instead of the body (Cloete & Scheffer, 2008; Synertial, 2011). The final type of artefact is kinetic, which is where the sensor continues to move even though the segment has come to a stop, this artefact results from the momentum that has developed from a fast movement or instances where the participant is not connected to the ground (jumping) (Cutti et al., 2006; Damgrave & Lutters, 2009; Lebel et al., 2017). Therefore, not only do artefacts affect accuracy but the type of joint being measured will also have an effect on the accuracy of the data captured by the mocap suit. The reduced weight and size of sensors, the use of a Lycra suit with strapping, and specific placement of the sensors on the body help to reduce the effect of all four of these artefacts.

Despite being able to account for the artefacts mentioned above, errors still remain. Studies have been conducted to investigate what factors other than the artefacts detailed above affect the accuracy of IMUs as sensors, and mocap suits. Accuracy studies have been conducted in controlled conditions and have demonstrated that the environment can affect the quality of orientation data and accuracy of IMU mocap (Brodie et al., 2008; Cutti et al., 2006; Damgrave & Lutters, 2009; Lebel et al., 2017; Roetenberg, et al., 2007; Wong et al., 2015). A primary environmental factor is that of magnetic fields, which cause drift to occur in the sensor's orientation and a consequent reduction in accuracy (Wong et al., 2015). Drift results from the sensitivity of the magnetometers in IMUs, to small changes in the magnetic field. This sensitivity causes in errors in the data captured by the sensors in the horizontal plane (Damgrave & Lutters, 2009). Compensation algorithms have been developed to reduce this effect (Battini et al., 2014; Klippert et al., 2012; Roetenberg, et al., 2007). However, Roetenberg and colleagues (2007) noted that although a filter algorithm may be effective in

the situation it was designed for, these situations are usually where magnetic disturbances are fixed and, therefore, the algorithms may be ineffective in workplaces with moving ferrous metal parts. Overall, despite these factors, the studies investigating the accuracy of IMUs have found inertial mocap to be close enough to Vicon measurements to be suitable for clinical measurement (Lebel et al., 2017; Schall et al., 2016). Accuracy and suitability have been assessed using the 5° acceptable difference recommended by McGinley and colleagues (2009), from their systematic review of inter-session and inter-assessor reliability of gait analysis. As long as the error is 5° or under, the measurement technique is considered acceptable for clinical measurement.

Most of the studies assessing the accuracy of IMUs vary from looking at individual sensors attached to solid objects (Brodie et al., 2008; Cutti et al., 2006), to two or three sensors attached to mock-ups replicating body segments (Brennan, et al., 2011). Alternatively, they assess one to three sensors attached across a joint on real people (Roetenberg et al., 2007; Schall et al., 2016). By conducting controlled experiments like this, researchers have been able to develop an understanding of the accuracy of the individual sensors and have identified good levels of accuracy for IMUs except when near ferrous metals (Roetenberg et al., 2007), or when the sensor was exposed to too much movement or high velocity (Brodie et al., 2008; Thies et al., 2007). However, an understanding of the effect of how IMUs interact across a whole body is still required.

Some accuracy assessments have been completed with a whole mocap suit, although the focus has remained on specific segments. In other cases where a whole suit has been evaluated for accuracy a dynamic activity was assessed and strapping rather than a Lycra suit were used. Cloete and Scheffer's (2008) study focused on benchmarking a MOVEN Xsens suit for clinical gait analysis and assessed only on the hip, knee and ankle joints. Their study found that when comparing an inertial mocap MOVEN Xsens suit with a Vicon system, there was a similar level of accuracy in the measurements for flexion and extension of the knee and hip, however, the ankle joint angles were found to be less accurate. Kim and Nussbaum (2013) compared the Xsens MVN with the Vicon system, and

found that the Xsens system varied from the Vicon system by 28% during peak kinematic movement. Lebel et al. (2017) compared a Synertial IGS-180 suit with a Vicon system, within their study participants were asked to complete a variety of tasks including sit-to-stand, walk, and turn. Their results found good levels of accuracy, although accuracy varied depending on the task performed and sensor considered, with the head and ankle showing the greatest difference between the two mocap systems. Hand and arm scores were not presented within this paper. Furthermore, strapping rather than Lycra suits were used by Kim and Nussbaum (2013) and Lebel et al. (2017). Strapping can lead to greater artefacts than are found with the Lycra suit (Synertial, 2012), and may explain some of the 28% difference between the Xsens MVN suit and the Vicon system in Kim and Nussbaum's (2013) study. The difficulty with whole body dynamic assessments is that it is challenging to identify where variation occurs with regard to specific body segments, and it is difficult to isolate the effect of movement or velocity, which are known to negatively affect the accuracy of IMUs, from the data.

Although inertial mocap was identified as the most suitable means of assessing physical strain in the experiments detailed within this thesis, the accuracy of the available mocap suit required identification prior to use. To this end, a static whole body accuracy assessment was required before it could be used with confidence.

2.2.1.3 Interaction Effects

The previous sections have identified the most suitable measures for assessing the effect of physical strain on mental workload, however, there may be a difficulty with just measuring the change in mental demand as affected by physical strain (DiDomenico & Nussbaum, 2011). DiDominico and Nussbaum (2011) found that the influence of physical workload on HRV was indeterminable from mental workload, and they could not identify whether HRV levels were associated with mental workload or a reflection of workload that was comprised of physical demand. Their recommendation for future experiments was to ensure that a subjective measure was recorded in addition to the physiological measure, particularly one with a level of diagnosticity. Both SWAT and the NASA-TLX provide this ability to diagnose where demand is experienced, as research has

shown that individuals are capable of being able to effectively discern between the different dimensions of workload. DiDomenico and Nussbaum (2008) found that when using a single type of physical task, individuals using subjective assessments (NASA-TLX) were able to distinguish and evaluate between physical and mental workload when concurrently exposed to both. However, extensive validation work may be required for expanding the equation to encompass different types of tasks.

2.2.1.4 Summary

From the literature review detailed above, two objectives can be identified to achieve the aim of identifying and establishing psychological parameters that will improve the capability of digital human modelling for manufacturing system design. One objective is to provide design engineers with a means of assessing workload in DHM CAD programs. The most effective means of doing this has been identified as establishing the combined parameters of physical strain and mental workload (Objective 2). However, prior to any measurement of physical strain using inertial mocap, an assessment of its effectiveness and accuracy is required (Objective 1). The following sections address the second way of improving DHM CAD programs that was identified in Section 2.2, this involves identifying and measuring psychological reactions to collaborative robotics that can be integrated into DHM CAD programs.

- **Objective 1:** Identify an effective and accurate means of capturing a person's movement
- **Objective 2:** Establish the combined parameters of physical strain and mental workload

2.2.2 Human Robot Collaboration

In addition to redesigning environments and assessing risk of MSD in virtual environments, the second way that risk of MSD development is being addressed within aerospace manufacturing is through the introduction of large high-payload collaborative robots. This is a change in manufacturing from robots that previously ran behind physical barriers, to those that can work in proximity to

operators without the need for the same level of separation (Fletcher et al., 2019; Gopinath et al., 2018). Current working conditions within aerospace manufacturing require individuals to perform physically and cognitively demanding tasks concurrently, it is expected that the introduction of collaborative robotics is likely to further increase the psychological and cognitive demand on individuals (Arai et al., 2010; Grahn et al., 2018; Koppenborg et al., 2017; Parasuraman & Riley, 1997). A consequence of this is likely that operators may end up mentally overloaded which may lead to increased errors and longer task times. To be able to understand and address human psychological reactions to the integration of collaborative robotics and integrate this understanding into DHM CAD programs, the identification of the psychological factors impacted is required. The introduction of these equations will provide design engineers and systems integrators with an understanding of the factors that require consideration during set up and installation of collaborative robots into factories. Additionally, they may form the basis of the development of adaptive automation systems that increase the number of tasks completed by an automated system or collaborative robot as the human operators workload increases (Vidulich & Tsang, 2012). This information is currently lacking and a contributing factor to the delay in the effective development of these systems.

There are psychological factors that have been identified within the literature as valuable in predicting human-system performance particularly in complex systems, automation, and robotics. These factors include: workload, trust in the robot, situation awareness, and mood (Grahn et al., 2018; Koppenborg et al., 2017; Parasuraman et al., 2008; Parasuraman & Riley, 1997; Vidulich & Tsang, 2012). Details of each of these interactions can be found below, including the identification of the most effective means for assessing each and establishing their suitability for developing an equation, or rule, that can be integrated into DHM CAD programs. No research has been found that has looked to establish a model for predicting the effect of interacting with a collaborative robot. Therefore, the equations or models are required to be developed from the ground up, incrementally, and require association with different measurable aspects of a robot in the first instance. Therefore, behavioural changes need to be measured

against objective variables associated with collaborative robotics that can be fully controlled and measured.

2.2.2.1 Mental Workload

In the context of HRC, the assessment of mental workload (for a definition and methods of analysis, see Section 2.2.1) is concerned with whether the act of interacting with or being in proximity to a collaborative robot increases mental workload levels for the individual. As operators are likely to increasingly work with and in close proximity to robots, it is possible that there will be an increase in monitoring demand, and anticipation and planning on how to interact and respond to the robot. A consequence of this is likely to be an increase in cognitive demand for the operator. If interaction leads to greater workload, the acute (increased error rates and task time, etc.) and chronic effects (physiological long-term health conditions) of excessive workload may be observed (DiDomenico & Nussbaum, 2011; Hockey, 1997). Understanding the effect of introducing HRC on mental workload, in both the short and longer term, is vital for building a safe and effective system (Vidulich & Tsang, 2012).

The means of measuring mental workload have already been detailed in Section 2.2.1.1. Although the literature recommended using multiple measures to assess workload, it was not possible in the studies presented in the last experimental Chapter of this thesis, as physiological measurement tools were unavailable for use at the time of experimentation. Therefore, the NASA-TLX was used on its own.

2.2.2.2 Trust

Trust in automation and robotics is inherently tied to its reliability (Vidulich & Tsang, 2012). A system's ability to perform reliably, particularly where an individual works in close contact with it, as is expected with HRC, is critical to ensuring an effective working relationship between the robot and the person. Other factors have also been found to affect trust in robotics including the size, shape, sounds and movement of the robot (Charalambous et al., 2016). Similarly to workload, trust has an optimal level and any departure from it leads to negative outcomes. Too much trust leads to complacency and errors while too little leads

to errors and increased mental load (Parasuraman et al., 2008), both of which increase the chance of errors. An additional consequence of a lack of trust is the system not being used (Lee & Moray, 1992), which undermines the benefits anticipated from implementation. Therefore, it is essential to understand how HRC affects trust and to be able to design collaborative environments that engender trust.

Trust in automation and robotics is typically assessed through subjective measures. Examples of surveys used to measure trust include those developed by Jian and colleagues (2000), and Muir (1994). Both scales provide emotive questions that explore trust as it relates to suspicion, confidence, familiarity and integrity. These types of scales provide an overview of trust in automation and robotics that are relevant to the systems that were being used prior to the currently highly collaborative systems. The recent development of high payload collaborative robotic systems has brought a new working experience for operators within manufacturing, as a result, there are limited measures for assessing trust in this instance. The closer working conditions seen with HRC require a trust measure directly relevant to these conditions. The perceived risk to safety is likely higher under those conditions, because the distance between the individual and the robot along with the visually obvious protective physical barriers are no longer required. A subjective measure of trust that provides a diagnostic output is most suitable for application within manufacturing, particularly for design engineers looking to design and implement new systems, as it allows for adjustments that ensure effective implementation and interaction with the system. A measure that provides an element of diagnosticity is particularly of importance for identifying a relationship or equation that can be integrated into DHM CAD programs for design of HRC systems.

A review of the literature provided generic trust scales or those developed specifically for other contexts, with a single survey identified for measuring trust in HRC within manufacturing designed by Charalambous and colleagues (2016): The Trust in Industrial Human-Robot Collaboration Scale. This scale measures the dimensions and factors specific to people's trust with large-scale industrial

robots and HRC (Charalambous et al., 2016). Within this survey, trust is measured across three dimensions: motion and pick-up speed, safe co-operation, and robot and gripper reliability. The survey items and dimensions provide a numerical output that can be used to create an equation for integration into a DHM CAD program, and its dimensions provide diagnostic potential in predicted trust in HRC. Therefore, it would be a suitable tool for identifying an equation that can be integrated into DHM CAD programs.

2.2.2.3 Mood

A person's mood affects their perception of an environment, the agents within the environment, and the person's judgment and associated behaviours (Lee, 2006; Tomlinson & Meyer, 2009). Mood primes judgement with positive moods leading to positive judgements and negative moods leading to negative judgements. Research has shown that an individual's mood affects information processing (Lee, 2006) and has an early-stage interaction effect with trust (Stokes et al., 2010). Therefore, it is critical to identify whether aspects or features of HRC induce negative moods to prevent a negative effect on judgment, and consequently, a negative attitude to a collaborative robot. Although a negative effect on trust may be overcome with time and exposure (Stokes et al., 2010), the time required to do so may contribute to increased errors or task time within manufacturing and an associated increase in costs. Therefore, an understanding of the effect of mood on the perception of a collaborative robot at first introduction is required. This would enable the robot to be set up and introduced in such a way that it does not trigger or contribute to negative moods.

Mood is generally measured using subjective measures. There are several widely used tools available, that can be distinguished by whether they measure two or three factors associated with mood. The Stress Arousal Check List (Mackay et al., 1978), Tension and Effort Stress Inventory (Svebak, 1993), and the Positive and Negative Affect scale (Watson et al., 1988) are two factor measures of mood, providing insight into energetic (wakefulness) and tense (stress) arousal. The University of Wales Institute of Science and Technology (UWIST) Mood Adjective Checklist (UMACL) (Matthews et al., 1990) on the other hand is a three factor

measure of mood providing data on energetic and tense arousal, and hedonic tone (capacity to experience pleasure). The three factor model enables a more diagnostic investigation of how HRC impacts mood and was, therefore, the measure of mood used within the experiments detailed in this thesis. Additionally, it provides a factor of frustration, which gives insight into a further negative mood that may be associated with interacting with a collaborative robot. This measure was identified as most suitable for assessing whether HRC has a measurable effect on mood because it provides a broader assessment of the factors that mood is comprised of.

2.2.2.4 Situation Awareness

Situation awareness (SA) is a person's mental model of the current state of their environment, all incoming information about the environment is processed and used to inform decision-making and associated actions (Endsley, 2012). Effective SA is acquired by ensuring accurate and complete perception, comprehension and projection of the state of an environment or of a factor within an environment (Endsley, 2012). The introduction of automation or robotics that take over part of a decision making process may contribute to a loss of awareness and, consequently, understanding regarding the tasks being completed by a robot or automated system (Parasuraman et al., 2000). Within HRC, maintaining effective SA is key to preventing errors and accidents (Grahn et al., 2018). This requires human operators knowing where a collaborative robot is at any point in time, understanding its intended task steps, and knowing its working speed (Grahn et al., 2018).

There are several methods used for measuring SA including: task performance, freeze probe techniques, real-time performance measurement, and subjective measures. Performance-based measures are used to assess how system changes may affect SA. This may include changes to how a machine presents information to an individual, a person's ability to effectively identify an alert or information and effectively understand it (Vidulich & Tsang, 2012). If an individual is unable to identify or understand an alert or the robots intent, then the system has not effectively contributed to the person's awareness of the situation.

Freeze probe techniques look to assess an individual's SA at a specific point during a task. The Situation Awareness Global Assessment Technique (SAGAT) developed by Endsley (1988) is a freeze probe procedure that requires a task to be unexpectedly stopped, and the individual being assessed to answer questions to assess their knowledge of the situation. Their responses are compared to the true situation and this comparison is used to determine the SAGAT score. This method has been critiqued as highly intrusive (Salmon et al., 2006; Vidulich & Tsang, 2012) and difficult to use in an applied or workplace settings where unexpectedly stopping a task is not possible.

Real-time performance assessments were developed to overcome the disruptive nature of the freeze-probe techniques, individuals are asked questions about aspects of the task they are completing as they complete it (Salmon et al., 2006; Vidulich & Tsang, 2012). Time taken to respond, in addition to accurate information provision, are used to assess effective SA (Salmon et al., 2006). This information can then be used diagnostically to inform design changes.

The most commonly used subjective measure of SA is the Situation Awareness Rating Technique (SART) developed by Taylor (1990). SA is assessed on three dimensions: attentional demands, attentional supply and understanding (Vidulich & Tsang, 2012), the scores associated with each dimension are combined to produce an overall SA score (Salmon et al., 2006).

Situation awareness was not measured within the experiments detailed within this thesis because, in order to establish equations for the other factors: trust, mood, and mental workload, extraneous variables had to be controlled for. This meant that the factors that may affect situation awareness were not manipulated. However, as the equations that are developed within the experiments detailed in this thesis are validated in a more ecologically valid environment, with a greater number of inputs or additional variables, work on predicting situation awareness should be considered.

2.2.2.5 Summary

Two further objectives to meet the aim of identifying and establishing psychological parameters that will improve the capability of DHM for manufacturing system design can be identified from the literature review above. One is to identify whether it is possible to establish psychological equations for reactions to HRC (Objective 3). Although four key psychological factors were identified as likely to affect a person's reaction to collaborative robotics, three factors were deemed suitable due to the ability to establish numeric outputs from their measures of assessment: mental workload, trust, and mood. The research conducted to address Objective 3 was part of a wider project with programmers who integrated the established equation into CAD, this provided the opportunity to assess the reliability of the equation established to meet Objective 3 (Objective 4).

- **Objective 3:** Establish human-robot interaction equation/s for industrial human robot collaboration
- **Objective 4:** Assess the predictive accuracy of the HRC equation after integration into DHM

2.3 Aim and Objectives

From the literature review four objectives have been identified to achieve the aim of identifying and establishing psychological parameters that will improve the capability of digital human modelling for manufacturing system design. One way is to providing design engineers with a means of assessing mental workload in DHM CAD programs. The most effective means of doing this has been identified as establishing the combined parameters of physical strain and mental workload (Objective 2). However, prior to any measurement of physical strain using inertial mocap an assessment of its effectiveness and accuracy is required (Objective 1). The third objective involves establishing human robot interaction equation/s for industrial HRC that can be integrated into DHM CAD programs (Objective 3) and assessing the predictive accuracy of the equation/s after integration into a DHM CAD program (Objective 4). Each of these objectives have been achieved in the following experimental Chapters.

3 EXPERIMENT: INVESTIGATING THE EFFECTIVENESS OF MOTION CAPTURE TECHNOLOGY

Four objectives for this thesis have been identified to address the aim of identifying and establishing psychological parameters that will improve the capability of digital human modelling for manufacturing system design. Objective 1 requires the identification of an effective means of capturing a person's movement particularly for use in occluded shop-floor work. This was partially met in the literature review, where the ways of capturing postures were identified, as a result, inertial mocap suits were identified as suitable for capturing an individual's movement particularly during occluded work seen within aerospace manufacturing. The relative youth of this technology means that there is limited literature on the accuracy of inertial mocap. Although studies have assessed mocap suits for dynamic accuracy (Cloete & Scheffer, 2008; Lebel et al., 2017), they have not provided a joint-by-joint assessment of whole body static accuracy. Therefore, an assessment of the static effectiveness and accuracy of the mocap suit and its associated software outputs was required. Due to technology access limitations an assessment of accuracy was not possible, however, an assessment of the suit's effectiveness was.

This Chapter describes an experiment reviewing the effectiveness of a mocap suit, which has been specifically designed for industrial fieldwork. The apparatus used in this experiment was detailed prior to the Design Section (Section 3.1.2) because this provides a clearer understanding of the tools described within the Design Section. Each limb was measured at 5° increments to ensure the repeatability of the angles measured in the graphical user interfaces (GUI) and to identify where deviations may occur. A position for best measurement was identified and replicated for each measurement of a particular joint, and this was replicated across the body.

3.1 Method

3.1.1 Apparatus

3.1.1.1 Measuring Equipment

Two measurement tools were used in this study, a physical measurement tool and an electronic one.

3.1.1.1.1 Protractor Goniometer

The physical placements of body segments were measured using a 12 ½" double-armed protractor goniometer. A protractor goniometer was identified as the most appropriate tool for providing the objective measurement to compare with the data captured using a mocap suit. Double-armed protractor goniometers are standard direct objective method for measuring body segment placement and range of motion in joints (Hamilton et al., 2011). This method of measurement has been used in clinical settings and has well-described reliability and validity (Clarkson, 2000). Additionally, this method has been used in a study to measure the accuracy of a kinetic model using an optical mocap system (Cook et al., 2007), supporting its suitability for this study.

It should be noted that comparing the inertial mocap suit with the Vicon mocap system would have been a more effective means of assessing accuracy, as it is considered the gold standard in mocap (Dutta, 2012). However, it was not possible to gain access to a Vicon system for this study, and before any further experiments could be conducted within this PhD the mocap suits accuracy required review. Therefore, due to the literature regarding reliability and validity of the protractor goniometer as an effective measure of joint angles (Clarkson, 2000; Cook et al., 2007) this method was identified as most suitable. Although this method has been utilised to assess accuracy, because it is not the gold standard in measurement it can provide an assessment of effectiveness, but not accuracy.

3.1.1.1.2 Screen Protractor

The angles recorded by the mocap suit and presented in the Synertial programs were measured using a screen protractor program (v.4., Iconico Inc.), as the GUIs

have no internal measurement options, and the measurement needed to be consistent with the protractor goniometer. Its use has been documented within clinical studies to measure angles onscreen (Cuff & Pupello, 2012; Johnson & Goodman, 2006; Ravenell et al., 2011).

3.1.1.2 Hardware

The joint data was electronically captured using a wireless full body inertial mocap suit (IGS-180-I, Synertial Ltd.) that records body segment and joint movements, with no restriction to motion. The system provides a real-time data transmission to a computer program that provides a visual output in a software program called IGS-Bio (v2.0 ALPHA, Synertial Ltd).

As detailed in Section 2.2.1.2.1.3 of the literature review, the accuracy of inertial mocap suits is affected by artefacts as well as the type of joint being recorded by a sensor. The reduced weight and size of sensors, the use of a Lycra suit with strapping, and specific placement of the sensors on the body (see Figure 3-2) help to reduce the effect of these artefacts, although some remain. As a result, each sensor is able to achieve one of three levels of accuracy: Alpha, Beta and Complex (Synertial, 2012). Alpha segments have a high level of data accuracy expected under almost every condition and range of movement. Beta segments can only obtain the same level of expected accuracy as Alpha ones if they are used to capture a limited range of movement, or if the sensors are placed on segments using rigid plates fasted at both ends of the segment to be unaffected by tissue artefacts. Finally, Complex segments are unable to achieve high levels of data accuracy because they generally encompass multiple bones working together in various axes but are measured using a single sensor, which will measure more of the displacement incurred by the bone nearest to it. Table 3-1 shows the level of accuracy or effectiveness expected for each body part measured using a mocap suit in this study, according to the documentation provided by supplier Synertial.

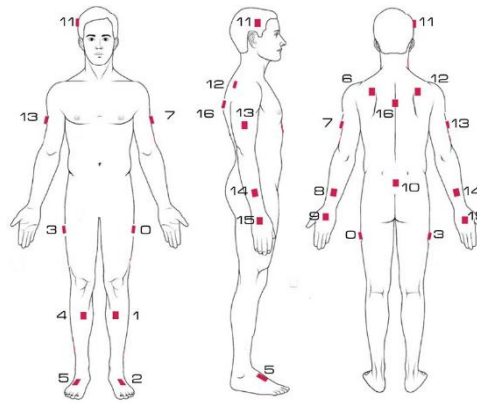


Figure 3-2: Correct Anatomical Placement of the IGS-180-I Sensors (Synertial, 2011)

Table 3-1: Alpha, Beta, and Complex Body Segments (Synertial, 2012)

Type of Segment	Body Part Measured
Alpha	Neck
	Elbow
	Wrist
	Knee
	Ankle
Beta	Upper arm
	Upper leg
Complex	Spine

There are three commercially available inertial mocap suits: the Synertial suit, the Xsens MVN Link (Xsens Technologies B.V.), and the Rokoko Smartsuit Pro (Rokoko Technologies). These suits have reduced size IMUs and provide a Lycra suit with strapping to minimise possible artefact effects on measurement. Additionally, the wires connecting the IMUs run under the suit, which prevents these wires from catching on external structures while the individual is carrying out an activity. This makes a suit the most suitable form of mocap for manufacturing where operators may be moving in and out of large structures. The Synertial IGS-180-I suit was chosen due to its availability, because it has been

developed for manufacturing environments and has reinforced connections to cope with the more strenuous movements that may be encountered within a manufacturing environment.

The Synertial IGS-180-I suit has 17 IMU sensors (Synertial Ltd.), which are used to capture changes in position, generating the numerical and visual output data. Synertial have laboratory-tested the resolution of the IMUs used in an IGS-180-I, which are < 1 degree for yaw and < 0.2 degree for pitch and roll. However, this does not guarantee accurate data collection due to the remaining artefacts as detailed above (Synertial Ltd., 2011).

3.1.1.3 Software

The inertial mocap system saves recorded data into biovision hierarchical (BVH) files, which provide Euler angles (x-y-z planes) between each joint, this data is translated into visual outputs and can be observed in two GUIs developed by Synertial: IGS-Bio and AnimaView (v.11.6., Synertial Ltd.). Both software programs are similar in that they provide a visual 360° interactive output, however, where a wireframe and solid manikin can be viewed in AnimaView, IGS- Bio only provides a solid manikin output. A diagram of these distinctions is shown in Figure 3-1 below.

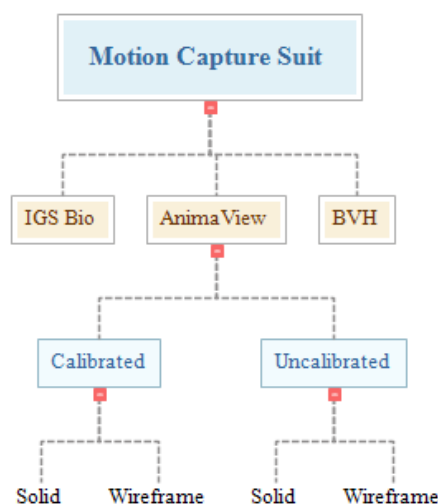


Figure 3-1: Animazoo (Synertial) Motion Capture Suit Software Options

Calibrated files are created through the application of an Actor File to the motion files in AnimaView (v.11.6., Synertial Ltd.), these Actor Files are created in a program provided by Synertial called AutoCal (v11.6., Synertial Ltd.). Motion files are calibrated because the mocap suits measure limb orientation displacement and maps these angles onto a model. This model can be either a standard model produced within the program or a model of the individual within the suit. The use of the standard model bypasses the need for segment positional data at the start of data collection, a consequence of which is that if the individual wearing the suit does not have the same physical dimensions as the model, the output will not accurately show limb placement because the limb lengths will be incorrect. However, a model of the individual wearing the suit can be acquired before data collection, and then the data can be mapped to the model after recording. The instructions and information sheets provided with the Synertial mocap suit state that the application of recorded data to a model of the recorded individual should not have an effect on the angles captured using the suit (Synertial, 2012). The calibration process requires two photographs be taken of the participant standing inside a cube shaped jig, both right side on and face on. The photos are then uploaded to AutoCal (v11.6., Synertial Ltd.) where an Actor File can be created to extract accurate segment lengths.

3.1.2 Design

The study looked to investigate the effectiveness of a mocap suit by comparing angles as measured using a protractor goniometer (standard direct objective method for measuring body segment placement within ergonomics [Hamilton et al., 2011]) with the angles electronically recorded by the suit's software. Such a comparison would enable an investigation of whether there was a difference between the three software programs and their outputs. There are five GUI output options and a text-based BVH file output. Therefore, the experiment's independent variables (IV) were the joint angles recorded by the mocap suit, with the dependent variables (DV) were the corresponding angles measured in each of the six available output options, the output options are presented in Figure 3-1.

The IV is the angles of flexion and extension created for fourteen of the participant's body segments. These angles were created using a protractor goniometer. Each body segment was placed at 5° intervals within the sagittal plane, and the angle was electronically recorded using an inertial mocap suit. The six output options (Figure 3-1) include two graphical software outputs that show a manikin replicating the posture captured by the mocap suit: IGS-Bio (condition 1) and AnimaView. AnimaView has 4 visual output options: Calibrated Solid (condition 2), Calibrated Wireframe (condition 3), Uncalibrated Solid (condition 4), and Uncalibrated wireframe (condition 5). The calibrated formats refer to the application of the Actor File to the recorded motion data. Henceforth, the calibrated data is referred to as Actor File data. The angles created in these programs were measured using an onscreen protractor. The sixth condition is the BVH file (numerical data output of the mocap suit).

3.1.3 Research Ethics

The research was performed in accordance with the regulations of the Science and Engineering Research Ethics Committee (SEREC) at Cranfield University and conducted in accordance with the Cranfield Research Integrity Policy. Details of this can be found in the description of the procedure in Section 3.1.5, and SEREC approval can be found in Appendix C.1.

3.1.4 Participant

A single male volunteer aged 29 and 186.2cm tall took part in this study. The participant's stature is at the 96 percentile from the anthropometric tables for the age range 19 to 65 years (Pheasant & Haslegrave, 2006).

3.1.5 Procedure

3.1.5.1 Set Up

The participant was fully informed of the process of data collection and their right to withdraw at any point. Once their informed consent was obtained, the participant was asked to put the mocap suit on. The researcher checked the placement of the sensors to ensure they matched the recommended placement

presented in Figure 3-2. Where placement was incorrect, the sensors were shifted and the strapping tightened to help reduce skin and clothing artefacts.

Two photos were taken of the participant for the creation of the Actor File, to enable the manikins to be calibrated from the standard manikin to the participant's limb length. The participant stood inside a calibration jig wherein front and right side photos were taken from a camera placed on a tripod positioned at waist height and 3 meters away from the jig. The camera was lined up with the jig central to the focal point of the camera on the horizontal plane, and with an equidistant gap between the rear and front upright poles of the jig as outlined in the Synertial user guide (Synertial, 2011). The photos were later uploaded to a personal computer (PC) for the Actor File to be created in AutoCal (v11.6). The Actor File for the participant was applied to the recorded data during data analysis, rather than prior to recording, to be able to identify its effect on joint measurements.

3.1.5.2 Data Capture

After taking the photos for AutoCal, the suit was switched on and aligned. Each of the participant's fourteen joints (Table 3-2) were individually positioned at 5° flexion and extension increments using a manual goniometer and recorded in IGS-Bio. The participant had a slightly reduced range of motion in their left wrist, explaining the difference in the highest measurement between the left and right wrist, 60° and 70° respectively.

Table 3-2: Range of Movement Measured Per Joint

Joint measured		Range of motion (measured in degrees)
Neck		-70 to 55
Trunk		-30 to 60
Right	Shoulder	-60 to 180
	Elbow	40 to 180
	Wrist	-55 to 70
	Hip	80 to 200
	Knee	50 to 180
	Ankle	90 to 140
Left	Shoulder	-60 to 180
	Elbow	40 to 180
	Wrist	-55 to 60
	Hip	80 to 200
	Knee	50 to 180
	Ankle	90 to 140

3.1.5.2.1 Goniometry by Segment

Each joint position was recorded separately for ease of measurement and was saved onto the standard Actor File provided by the IGS-Bio GUI. Standard measurement techniques as detailed in Clarkson (2000) were used for all segments except the neck and trunk due to the need to be able to compare the goniometer measurements with the GUI and BVH outputs. These measurement techniques are detailed below; the placement of the arms and centre of the goniometer were maintained for each of the 5° increments across the segments.

3.1.5.2.1.1 Arms

The joints in both arms were measured in flexion and extension. For the shoulder, this involved placement of the axis of the goniometer at the lateral aspect of the centre of the humeral head with the arm at the side. The stationary arm of the goniometer was held parallel to the midaxillary line of the trunk and the moveable arm parallel to the longitudinal axis of the humerus. Figure 3-3 provides an image

showing the physiological landmarks used as reference points to align the goniometers axis and arms for the shoulder.

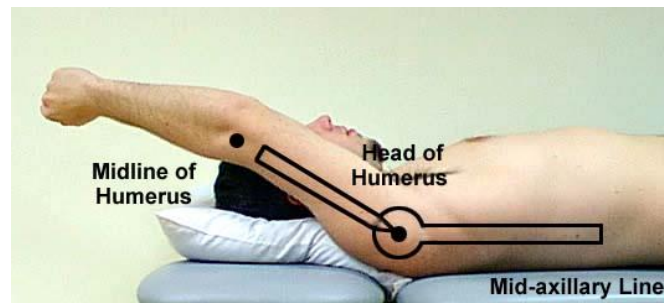


Figure 3-3: Goniometry of the Shoulder¹

Measurement of the elbow involved placing the goniometer's axis on the lateral epicondyle of the humerus, the stationary arm ran parallel to the longitudinal axis of the humerus pointing to the tip of the acromion process, and the tip of the moveable arm pointed towards the radial styloid. Figure 3-4 provides an image showing the physiological landmarks used as reference points to align the goniometers axis and arms for the elbow.

¹ Image retrieved from:
<https://www.scranton.edu/faculty/kosmahl/courses/gonio/upper/pages/shoulder-flex.shtml>

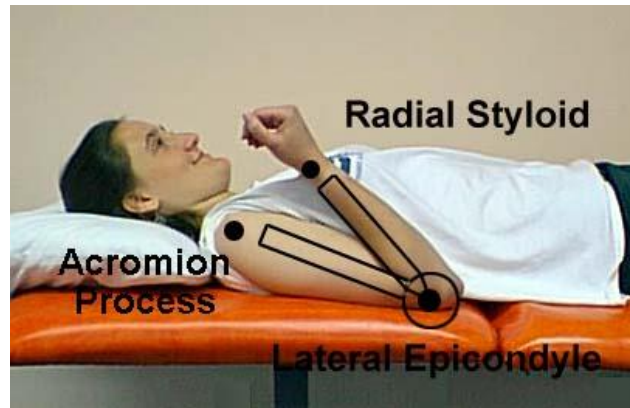


Figure 3-4: Goniometry of the Elbow²

The wrist was measured with the forearm neutral on a table and the palm facing in. The axis of the goniometer was placed over the styloid process of the radius, with the stationary arm parallel to the longitudinal axis of the radius, and the moveable arm ran parallel to the longitudinal axis of the index finger. Figure 3-5 provides an image showing the physiological landmarks used as reference points to align the goniometers axis and arms for the wrist.



Figure 3-5: Goniometry of the Wrist³

² Image retrieved from:

<https://www.scranton.edu/faculty/kosmah/courses/gonio/upper/pages/elbow-flex.shtml>

³ Image retrieved from: <https://otassessments.wordpress.com/events-list/wrist/>, labels added by author.

3.1.5.2.1.2 Legs

The joints in both legs were also measured in flexion and extension. Hip flexion required the participant to start in a supine position and in a prone position for extension. The axis of the goniometer was placed over the greater trochanter of the femur, with the stationary arm parallel to the midaxillary line of the trunk and the moveable arm parallel to the longitudinal axis of the femur. Figure 3-6 provides an image showing the physiological landmarks used as reference points to align the goniometers axis and arms for the hip.

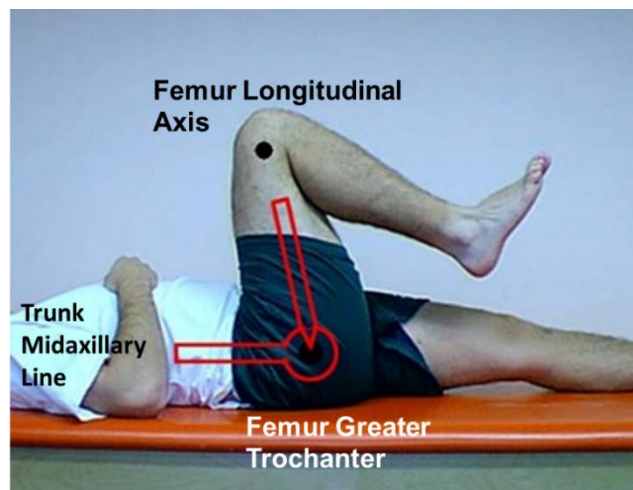


Figure 3-6: Goniometry of the Hip in a Prone Position⁴

Measurement of the knee required the axis to be placed over the lateral epicondyle of the femur, the stationary arm ran parallel to the longitudinal axis of the femur, pointing towards the greater trochanter and the movable arm ran parallel to the longitudinal axis of the fibula. Figure 3-7 provides an image showing the physiological landmarks used as reference points to align the goniometers axis and arms for the knee.

⁴ Image retrieved from: <https://quizlet.com/106018468/basic-practical-4-goni-flash-cards/>, labels added by author.

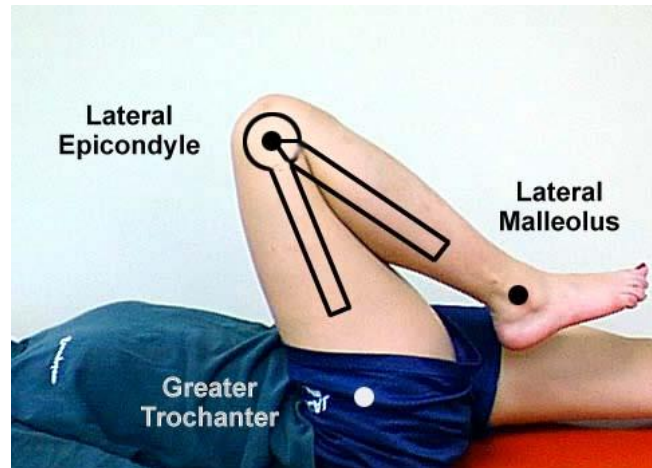


Figure 3-7: Goniometry of the Knee ⁵

Ankle dorsiflexion and plantarflexion was measured by placing the goniometer's axis just below the lateral malleolus, with the stationary arm parallel to the longitudinal axis of the fibula and the moveable arm was parallel to the fifth metatarsal. Figure 3-8 provides an image showing the physiological landmarks used as reference points to align the goniometers axis and arms for the ankle.

⁵ Image retrieved from:

<https://www.scranton.edu/faculty/kosmah/courses/gonio/lower/pages/knee-flexion.shtml>

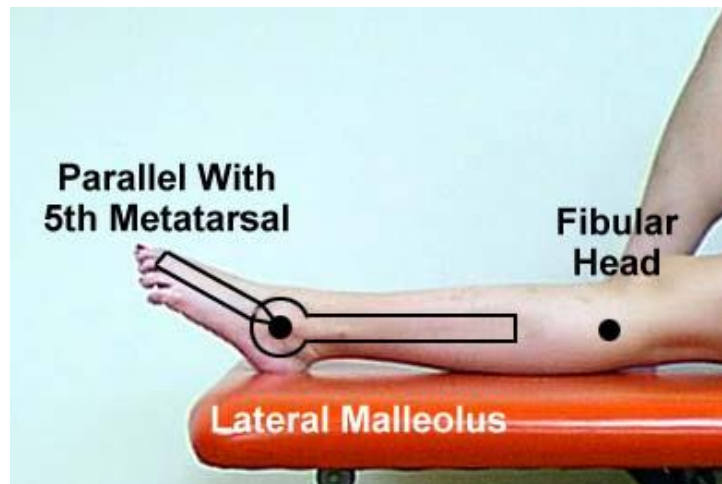


Figure 3-8: Goniometry of the Ankle⁶

3.1.5.2.1.3 Neck and Trunk

Traditional methods were not followed for the neck and the trunk due to the need to reflect the Synertial BVH output. The Synertial BVH output interpolates the necks movement from a sensor on the head and a sensor on the chest, and provides an angle of extension and hyperextension based on movement of the head in relation to the chest sensor from the calibration pose, which does not match traditional goniometry methods. The traditional measurements require that cervical extension and hyperextension of the neck be measured by placing the axis of the goniometer at the tragus of the ear, the stationary arm held parallel to the ground and the moveable arm in line with the base of the nose (Clarkson, 2000). The lack of physical landmarks like the ear in the mocap outputs additionally makes the replicability of traditional measurement methods for the neck impossible. Therefore, due to the need for a reference limb in the Synertial GUIs, the stationary arm was parallel to the neck and the moveable arm ran through the midline to the top of the head. Figure 3-9 provides an image showing the adapted landmarks used as reference points to align the goniometers axis and arms for the neck.

⁶ Image retrieved from:
<https://www.scranton.edu/faculty/kosmahl/courses/gonio/lower/pages/ankle-dor.shtml>



Figure 3-9: Adapted Goniometry of the Neck⁷

The traditional method for measuring trunk flexion and extension employs the use of a tape measure down the middle of the back down the spine. This gives a centimetre measurement of how much the spine extends or flexes but does not provide an angle, which the mocap suit gives. Therefore, the goniometer was used to measure forward bending, similarly to the way it was used on the hips with the axis placed over the greater trochanter of the femur, the stationary arm parallel to the midaxillary line of the trunk and the moveable arm parallel to the longitudinal axis of the femur. Figure 3-10 provides an image showing the physiological landmarks used as reference points to align the goniometers axis and arms for the trunk.

⁷ Image retrieved from: <https://headandcspine.weebly.com/manual-muscle-testingrange-of-motion.html>, goniometer outline added by author

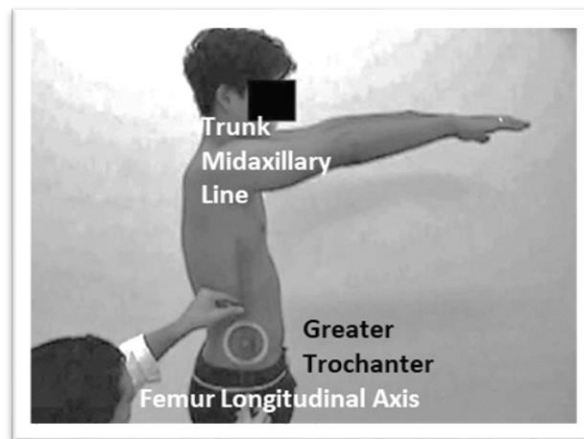


Figure 3-10: Adapted Goniometry of the Trunk (Chertman, et al., 2010)

3.1.5.2.2 Frame Isolation

The mocap suit software records postures over a period of time rather than at a single point in time or frame. Consequently, a single point in time, or frame, has to be identified from the recording so that the same point could be measured in all GUIs. During the recording phase of data collection, the participant held their segment immobile at the correct angle for three seconds, leading to any frame within these three seconds being suitable for use. Each recording was reviewed to ensure no movement had occurred before identifying a suitable frame. The identified frame was used to measure the recorded angle in each of the five GUI outputs and isolated from the BVH text file.

3.1.5.2.3 Actor File Creation

The participant's Actor File was applied to the data during the analysis stage to identify differences from the standard Actor File angles provided by the mocap software. To apply the Actor Files and create a calibrated manikin an Actor File was created in AutoCal (v11.6) using the two photographs taken in the set up processes. In order to do so, the photographs were uploaded to the AutoCal and the markers for the calibration jig were then placed at each point of the cube to provide a reference for the representative skeleton that was used to create the calibrated Actor File. The representative skeleton has moveable joints and points

which were placed on the body of the participant in the photograph, creating the representative model of the participant.

3.1.5.3 Graphic User Interface and BVH data collection

None of the GUIs provides a means to measure the manikin; as a result, a computer-generated protractor (v.4.0, Iconico, Inc) was used to measure the angles in each of the visual outputs. To ensure the repeatability of the angles measured in the two GUIs and their multiple output options, a position for best measurement was identified and replicated for each measurement of a particular joint. The placement of the computer generated protractor for each body segment reflected the placement of the goniometer during the data capture phase. The following paragraphs will describe the specific placement for each joint in the GUIs.

The ends of the protractor ran to the tip of each segment to ensure that the protractor arms were as close to the midline as possible. The shoulder, elbow, and hip data required the manikin to be aligned through the chest with a side on view, either right or left depending on which joint was measured. As can be seen in Figure 3-11, the chest was comprised of three diamond shapes. To align them, the widest point at the base and the three points at the top of the trunk were lined up so that the points overlapped. During recording of the wrist, the forearm rested on the table and movement of the wrist was recorded from the neutral position, accordingly, a top-down view was used for the wrist. To measure the neck, the two shoulder segments were aligned with one another. The knee required the two thigh segments to be aligned in a side view. Measurement of the ankle involved lining up two sections that comprise the foot in the GUIs and then measuring the movement of the ankle from the side view.

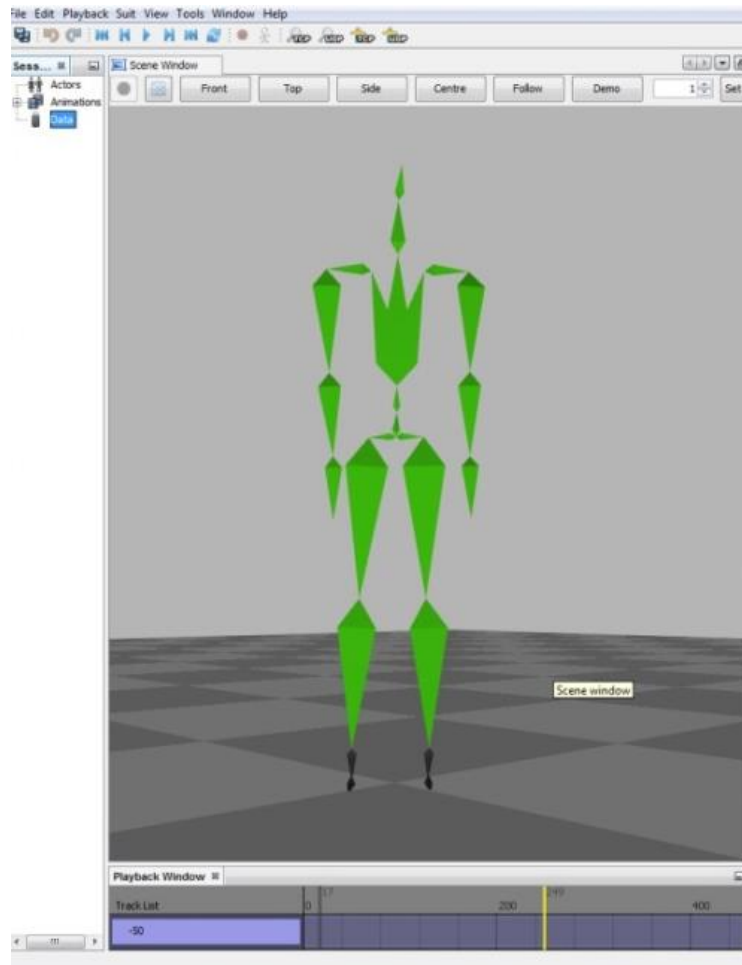


Figure 3-11: IGS-Bio Output

The manikin's trunk is made of three sections, these sections were aligned and one arm of the protractor was placed through the top point of the middle diamond in the chest section (the central diamond shape in Figure 3-11) of the manikin, to account for all of the movement seen in the trunk. The second arm of the protractor was placed at the top end of the thigh.

BVH files are the numerical output of the suit and are created by Synertial through the use of a kinematics algorithm which is used to generate the 3D manikin seen in the GUIs. The corresponding angles in the BVH formats were isolated using the frame numbers and the BVH output column identified in the reference table found in Appendix 6.3 Appendix A. The reference table provides the segment and

joint being measured in addition to identifying what each axis of the Euler angles are measuring.

3.1.6 Analysis

Prior to statistical analysis of the data, the differences between the measurement outputs were graphically compared to the goniometry measurements to identify any unexpected variations or outliers in the measured data. As a result of this review it was identified that the BVH data for the knees required additional processing due to the way it is presented. The BVH numerical output appears to reduce as the knee progresses through 90° to 180° so, rather than increasing from 90° to 180° as the leg extends, it reduces back towards 0° . Figure 3-12 and Figure 3-13 show the BVH output for the right and left knees and the corrected output against the goniometer measurements. Therefore, it appears as though the knee in a 90° position is considered 90° in the BVH files and then, as the lower leg moves either towards the thigh or away from the thigh, the angles decrease. To correct this for comparative purposes, the knee BVH outputs from 90° to 180° were subtracted from 180° . This calculation was applied to all BVH measurements that corresponded with the 95° to 180° goniometer measurements for both the left and right knee.

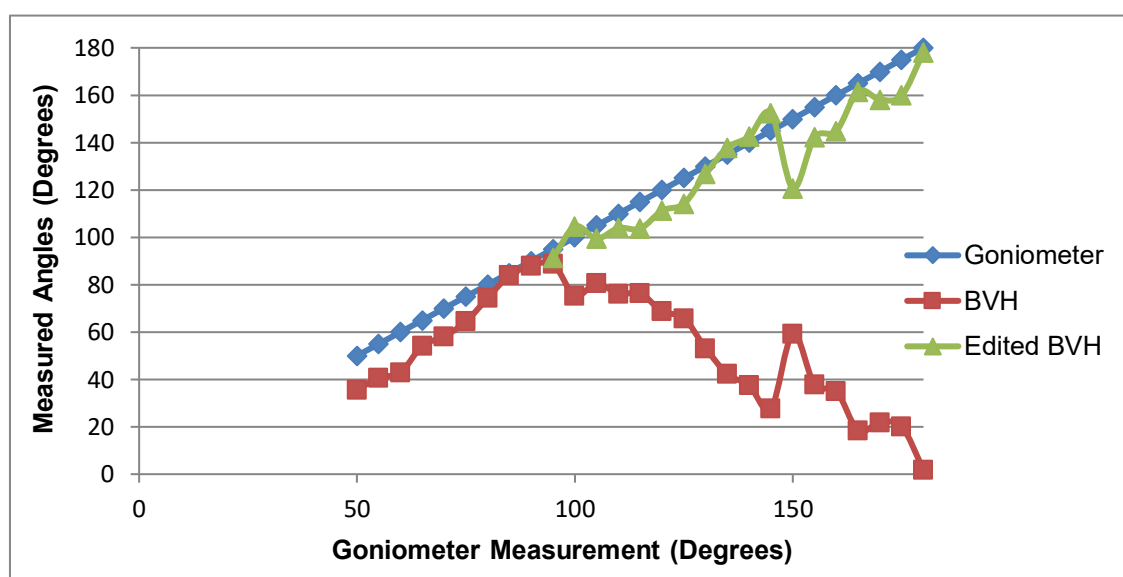


Figure 3-12: Right Knee Edited BVH Data and Unedited BVH Data

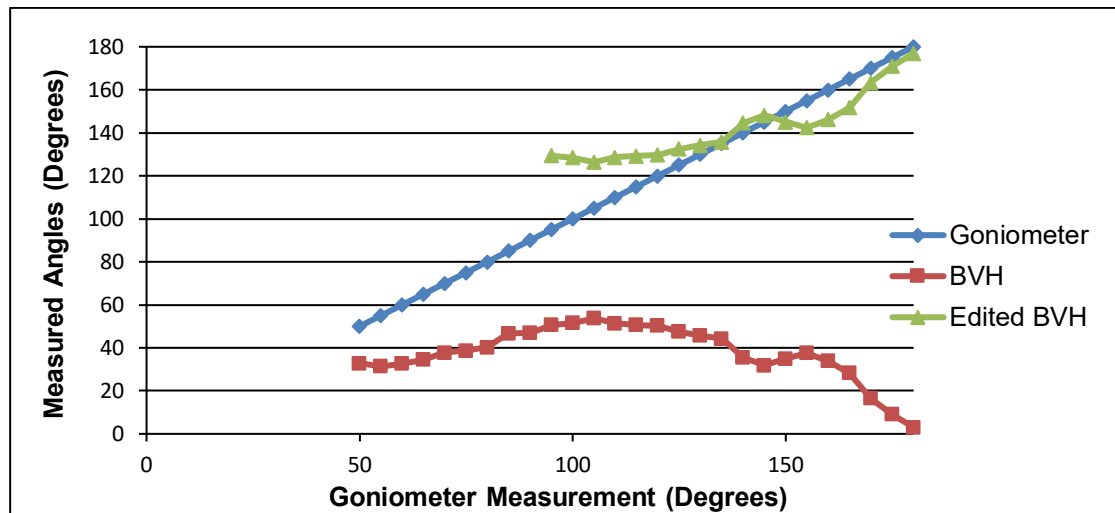


Figure 3-13: Left Knee Edited BVH Data and Unedited BVH Data

The study looked to investigate whether the measurement of the angles created by an individual wearing the mocap suit are the same whether said measurement is done using a protractor goniometer or provided by one of the five available Synertial outputs, measured using a computer generated protractor, or the BVH files. To assess this, an inter-rater reliability assessment was completed to identify whether there was a good level of agreement between the different measurements. The data was assessed for normality using skew and kurtosis values and were identified as normally distributed. Therefore, Pearson's product-moment correlations were applied to the data to identify the inter-rater reliability between the Synertial outputs and the goniometer measurements. Correlations allow identification of whether two datasets co-vary. A high correlation indicates that said datasets present the same pattern across the different angles. However, it is possible that two identical patterns (perfect correlation) are not aligned but rather have a systematic difference. An example may be that the BVH systematically overestimates every angle by 2° compared to the goniometer scores. For this reason, a measure of systematic bias is required. The Bland Altman plot provides such a measure (Ranganathan et al., 2017). The test provides a value representing the systematic difference. A value significantly different from zero indicates the presence of a systematic bias. For this reason,

the value obtained for each angle was tested against zero using one-sample t-tests. In order to control for the increased chance of Type I error induced by the numerous tests, a Bonferroni correction was applied, lowering the significance threshold to .008. The following section presents the results for the whole body and individual body segments.

3.2 Results

3.2.1 Whole Body Assessment

Pearson's product-moment correlations were run between the goniometer and each dataset and found strong significant positive correlations when comparing the GUI measurements with the goniometer placements individually (IGS-Bio: $r(375) = .960$, $p < .001$; AnimaView Actor File solid: $r(375) = .966$, $p < .001$; AnimaView Actor File Wireframe: $r(375) = .967$, $p < .001$; AnimaView solid: $r(375) = .955$, $p < .001$; AnimaView Wireframe: $r(375) = .955$, $p < .001$; BVH: $r(375) = .791$, $p < .001$) (Evans, 1996).

Although the correlations suggest that all six measures vary similarly to the goniometer, they do not report on how close the measures are. In order to test for this, Bland-Altman plots were conducted and provided the mean differences. Each mean difference was then tested against zero using a one-sample t-test to identify whether the difference was significant. This provided information on whether the identified systematic bias was significantly greater than zero. Table 3-3 presents these results in addition to the standard deviation (SD) of the mean difference and the upper and lower bounds. These results show the extent to which the angles identified in the outputs differed from the goniometer measurements.

Table 3-3: Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Lower Bound	Upper Bound	<i>t</i> value	<i>p</i> value
IGS-Bio	377	1.219	19.246	-36.503	38.940	1.229	.220
AnimaView Solid	377	1.414	20.482	-38.723	41.558	1.341	.181
AnimaView Wireframe	377	1.464	20.387	-38.494	41.422	1.395	.164
AnimaView Actor File Solid	377	7.650	18.188	-27.998	43.298	8.167	< .001
AnimaView Actor File Wireframe	377	7.631	18.115	-27.875	43.137	8.179	< .001
BVH	377	9.386	43.157	-75.201	93.973	4.223	< .001

The results show a significant systematic difference for both calibrated AnimaView outputs and for BVH from the goniometer measurements. This means that despite all three methods varying similarly to the true value, as shown by the correlations, they systematically misestimate the value of the angle. The systematic differences of the AnimaView outputs without the Actor Files, and the IGS-Bio outputs were not significant, which suggests that they have both a good level of congruence and agreement with goniometer measurements. When looking at the mean difference, IGS-Bio has the lower mean difference, and SDs and both bounds are closer to zero than the AnimaView outputs, suggesting that it is likely to be the most suitable output to use. The upper and lower bounds show that the systematic bias of 95% of the IGS-Bio scores vary between -36.503 and 38.940 from the goniometer measurements, which indicates that additional forms of observation may be required to record postures of individuals.

3.2.2 Individual Segment Analysis

3.2.2.1 Shoulder

3.2.2.1.1 Right

Pearson's product-moment correlations of the right shoulder's angle of flexion and extension between the goniometer and all of the datasets found very strong positive correlations as defined by Evans (1996), when comparing the output

measurements with the goniometer placements (IGS-Bio: $r(47) = .988, p < .001$; AnimaView Actor File solid: $r(47) = .992, p < .001$; AnimaView Actor File Wireframe: $r(47) = .992, p < .001$; AnimaView solid: $r(47) = .990, p < .001$; AnimaView Wireframe: $r(47) = .991, p < .001$; BVH: $r(47) = .937, p < .001$).

Table 3-4 provides the results from right upper arm Bland-Altman assessments along with the SD of the mean difference and the upper and lower bounds. Despite the correlations suggesting that all six measures vary similarly to the goniometer, the Bland-Altman plots identified that the difference between the conditions and the goniometer were significant and, therefore, identified a systematic bias significantly greater than zero between all measurements and the goniometer. This provides evidence for a lack of agreement between the goniometer and the mocap outputs and suggests that additional forms of mocap are used alongside the suit for assessing shoulder movement.

Table 3-4: Right Shoulder Comparison of the Values Obtained with the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Lower Bound	Upper Bound	<i>t</i> value	<i>p</i> value
IGS-Bio	49	34.963	48.163	129.362	-59.436	5.082	< .001
AnimaView Solid	49	14.053	26.374	65.746	-37.641	3.730	.001
AnimaView Wireframe	49	14.131	26.273	65.626	-37.364	3.765	< .001
AnimaView Actor File Solid	49	22.466	21.910	65.408	-20.477	7.178	< .001
AnimaView Actor File Wireframe	49	22.428	21.771	65.098	-20.242	7.211	< .001
BVH	49	34.963	48.163	129.362	-59.436	5.082	< .001

3.2.2.1.2 Left

Pearson's product-moment correlations for the left shoulder found very strong positive correlations, as defined by Evans (1996), when comparing the GUI measurements with the goniometer placements individually (IGS-Bio: $r(47) = .990, p < .001$; AnimaView Actor File solid: $r(47) = .982, p < .001$;

AnimaView Actor File Wireframe: $r(47) = .981$, $p < .001$; AnimaView solid: $r(47) = .987$, $p < .001$; AnimaView Wireframe: $r(47) = .987$, $p < .001$), and a strong correlation between the goniometer placements and the BVH data ($r(47) = .781$, $p < .001$) (Evans, 1996).

Table 3-5: Left Shoulder Comparison of the Values Obtained with the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Lower Bound	Upper Bound	<i>t</i> value	<i>p</i> value
IGS-Bio	49	5.731	15.022	35.174	-23.713	2.670	.010
AnimaView Solid	49	12.451	23.584	58.676	-33.775	3.695	.001
AnimaView Wireframe	49	12.439	23.416	58.335	-33.458	3.718	.001
AnimaView Actor File Solid	49	18.746	17.579	53.201	-15.708	7.465	< .001
AnimaView Actor File Wireframe	49	18.394	17.516	52.726	-15.937	7.351	< .001
BVH	49	39.403	56.717	150.567	-71.762	4.863	< .001

Table 3-5 presents the results from the left shoulder Bland-Altman plots along with the SD of the mean difference, the upper and lower bounds, and the t-tests to identify whether the systematic differences were significantly different from zero.

A lack of significance was found in the difference between the IGS-Bio and goniometer measurements. This suggests that for the left upper arm the IGS-Bio GUI is the more suitable tool for assessing flexion and extension in the left shoulder. The Bland-Altman plots for the remaining five outputs revealed a systematic bias significantly greater than zero from the goniometer measurements, consequently, revealing their lack of suitability for effective measurement of the left shoulder when compared to the goniometer.

3.2.2.2 Elbow

3.2.2.2.1 Right

Pearson's product-moment correlations between the goniometer and all mocap suit measurements, for the right elbow, indicate very strong significant positive correlations, as defined by Evans (1996) (IGS-Bio: $r(27) = .988$, $p < .001$; AnimaView solid: $r(47) = .990$, $p < .001$; AnimaView Wireframe: $r(27) = .991$, $p < .001$; AnimaView Actor File solid: $r(47) = .992$, $p < .001$; AnimaView Actor File Wireframe: $r(27) = .993$, $p < .001$; BVH: $r(27) = .828$, $p < .001$).

Table 3-6: Right Elbow Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Lower Bound	Upper Bound	t value	p value
IGS-Bio	29	-17.263	6.929	-3.683	-30.843	-13.418	< .001
AnimaView Solid	29	-17.825	6.813	-4.471	-31.179	-14.089	< .001
AnimaView Wireframe	29	-17.643	6.677	-4.557	-30.730	-14.230	< .001
AnimaView Actor File Solid	29	12.032	8.525	28.741	-4.677	7.600	< .001
AnimaView Actor File Wireframe	29	12.139	8.391	28.585	-4.306	7.791	< .001
BVH	29	-36.963	28.258	18.422	-92.348	-7.044	< .001

As can be seen in the outputs from the Bland-Altman plots for all right elbow measurements in Table 3-6, the t-tests revealed that the difference between all mocap measurements and the goniometer measurements are all significantly different from zero and have, therefore, a systematic bias. Consequently, this suggests that none of the outputs effectively match the goniometer and that other forms of posture capture should be used. The IGS-Bio, both AnimaView datasets, and the BVH files systematically overestimate the angles, whereas both AnimaView Actor File outputs systematically underestimate the angles. The AnimaView Actor File files have the lowest mean difference suggesting that they

might be the closest to the goniometer measurements. Further testing would be required to identify whether they are significantly different from one another.

3.2.2.2.2 Left

Pearson's product-moment correlations between the goniometer and all mocap suit measurements, for the left elbow, indicate very strong correlations as defined by Evans (1996) (IGS-Bio: $r(27) = .996$, $p < .001$; AnimaView solid: $r(27) = .996$, $p < .001$; AnimaView Wireframe: $r(27) = .995$, $p < .001$; AnimaView Actor File solid: $r(27) = .996$, $p < .001$; AnimaView Actor File Wireframe: $r(27) = .996$, $p < .001$; BVH: $r(27) = .932$, $p < .001$).

Bland-Altman plots were then applied to the datasets, and t-tests were used to investigate the level of agreement between the goniometer and the mocap suit measurements. This data can be found in Table 3-7. As all of the measurements are systematically significantly different from zero, this suggests that there is no agreement between the measurements. Therefore, this indicates that the mocap suit outputs cannot be used to reliably assess the left elbow.

Table 3-7: Left Elbow Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Lower Bound	Upper Bound	t value	p value
IGS-Bio	29	-29.662	5.463	-18.955	-40.370	-29.240	< .001
AnimaView Solid	29	-29.571	5.285	-19.213	-39.930	-30.132	< .001
AnimaView Wireframe	29	-29.160	5.456	-18.466	-39.854	-28.780	< .001
AnimaView Actor File Solid	29	7.108	4.715	16.349	-2.134	8.118	< .001
AnimaView Actor File Wireframe	29	7.157	4.656	16.282	-1.969	8.278	< .001
BVH	29	-41.522	21.734	1.078	-84.122	-10.288	< .001

3.2.2.3 Wrist

3.2.2.3.1 Right

Pearson's product-moment correlations of the right wrist angles between the goniometer and all mocap suit measurements indicate very strong correlations as defined by Evans (1996) (IGS-Bio: $r(24) = .995$, $p < .001$; AnimaView solid: $r(24) = .993$, $p < .001$; AnimaView Wireframe: $r(24) = .993$, $p < .001$; AnimaView Actor File solid: $r(24) = .988$, $p < .001$; AnimaView Actor File Wireframe: $r(24) = .988$, $p < .001$; BVH: $r(24) = .995$, $p < .001$). This indicates that the mocap measurements co-vary with the goniometer measurements.

The Bland-Altman tests were then used to identify the level of agreement between the goniometer measurements and those from the other outputs. The results are presented in Table 3-8 and show that the measures are all systematically significantly different from the goniometer and do not provide a confident measure of the right wrist in any of the outputs.

Table 3-8: Right Wrist Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Lower Bound	Upper Bound	<i>t</i> value	<i>p</i> value
IGS-Bio	26	19.239	10.869	40.542	-2.065	9.025	< .001
AnimaView Solid	26	18.149	11.415	40.523	-4.225	8.107	< .001
AnimaView Wireframe	26	18.126	11.419	40.506	-4.255	8.094	< .001
AnimaView Actor File Solid	26	9.162	11.097	30.913	-12.589	4.210	< .001
AnimaView Actor File Wireframe	26	9.172	11.098	30.923	-12.580	4.214	< .001
BVH	26	18.338	13.365	44.533	-7.857	6.996	< .001

3.2.2.3.2 Left

Pearson's product-moment correlations of the left wrist angles between the goniometer and all mocap suit measurements indicate very strong correlations (Evans, 1996) (IGS-Bio: $r(22) = .983$, $p < .001$; AnimaView solid: $r(22) = .982$, p

< .001; AnimaView Wireframe: $r(22) = .982, p < .001$; AnimaView Actor File solid: $r(22) = .983, p < .001$; AnimaView Actor File Wireframe: $r(22) = .983, p < .001$; BVH: $r(22) = .973, p < .001$). This indicates good reliability between each of the mocap suit outputs and the goniometer measurements.

Bland-Altman plots and t-tests were conducted to identify the level of agreement between the mocap outputs and the goniometer measurements. These results are found in Table 3-9, and show that the mocap measurements are significantly different from the goniometer measurements and, therefore, despite their good reliability they do not have a good level of agreement. This indicates that alternative methods of mocap should be used alongside the mocap suit.

Table 3-9: Left Wrist Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Lower Bound	Upper Bound	t value	p value
IGS-Bio	24	-15.283	8.078	0.551	-31.117	-9.268	< .001
AnimaView Solid	24	-15.712	8.476	0.902	-32.325	-9.081	< .001
AnimaView Wireframe	24	-15.719	8.487	0.916	-32.354	-9.073	< .001
AnimaView Actor File Solid	24	-32.158	6.708	-19.011	-45.306	-23.486	< .001
AnimaView Actor File Wireframe	24	-32.135	6.699	-19.006	-45.265	-23.501	< .001
BVH	24	-13.625	14.961	15.699	-42.949	-4.461	< .001

3.2.2.4 Neck

Pearson's product-moment correlations between the goniometer and all mocap suit measurements, for the neck angles, indicate very strong correlations (Evans, 1996) (IGS-Bio: $r(24) = .992, p < .001$; AnimaView solid: $r(24) = .989, p < .001$; AnimaView Wireframe: $r(24) = .989, p < .001$; AnimaView Actor File solid: $r(24) = .988, p < .001$; AnimaView Actor File Wireframe: $r(24) = .989, p < .001$; BVH: $r(24) = .990, p < .001$). These results illustrate that the measurements strongly

co-vary. However, to understand how well they agree, Bland-Altman plots and t-tests were applied to the data. The results of these analyses are presented in Table 3-10, showing that all outputs except for the IGS-Bio measurements are not significant which suggests a good level of agreement between the measurements. Consequently, the mocap suit can be used with a level of satisfactory confidence when analysing flexion and extension within the neck, excluding the IGS-Bio output.

Table 3-10: Neck Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Upper Bound	Lower Bound	<i>t</i> value	<i>p</i> value
IGS-Bio	26	-7.707	12.921	17.618	-33.031	-3.041	.005*
AnimaView Solid	26	-6.640	14.343	21.473	-34.752	-2.360	.026
AnimaView Wireframe	26	-6.687	14.470	21.674	-35.047	-2.356	.027
AnimaView Actor File Solid	26	.293	14.357	28.433	-27.848	.104	.918
AnimaView Actor File Wireframe	26	.446	14.353	28.577	-27.685	.158	.875
BVH	26	-6.675	14.667	22.072	-35.423	-2.321	.029

*Bonferroni correction: significance= .008

3.2.2.5 Trunk

Pearson's product-moment correlations of the trunk angles found very strong significant correlations between each of the output measurements and the goniometer measurements, as defined by Evans (1996) (IGS-Bio: $r(17) = .978$, $p < .001$; AnimaView solid: $r(17) = .986$, $p < .001$; AnimaView Wireframe: $r(17) = .986$, $p < .001$; AnimaView Actor File solid: $r(17) = .986$, $p < .001$; AnimaView Actor File Wireframe: $r(17) = .986$, $p < .001$; BVH: $r(17) = .965$, $p < .001$) (Evans, 1996). This strong correlation indicates a close level of co-variance.

Agreement between the output measurements and the goniometer measurements was then assessed using the Bland-Altman plots and t-tests to

test whether the mean differences were significantly different from zero. These results are presented in Table 3-11 and show that the only mocap output not systematically significantly different from zero is the BVH files. Therefore, the results suggest that only the BVH files might be used with confidence. It should be noted that the mean difference is 7.291° and, consequently, it is possible that any assessment using the BVH output may be mismeasured by that much. This may have a negative effect if postural analysis is being conducted and the trunk posture is close to a scoring boundary. This would be of concern if it leads to underscoring of a posture and the need for remedial action is not identified.

Table 3-11: Trunk Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Upper Bound	Lower Bound	t value	p-value
IGS-Bio	19	-6.953	8.379	9.471	-23.377	-3.617	.002
AnimaView Solid	19	-13.604	7.974	2.024	-29.232	-7.437	< .001
AnimaView Wireframe	19	-13.568	8.034	2.178	-29.315	-7.362	< .001
AnimaView Actor File Solid	19	-7.228	7.769	8.000	-22.456	-4.055	.001
AnimaView Actor File Wireframe	19	-7.255	7.807	8.047	-22.557	-4.050	.001
BVH	19	7.291	15.703	38.068	-23.486	2.024	.058

Note: Bonferroni correction: adjusted significance threshold = .008

3.2.2.6 Hip

3.2.2.6.1 Right

To identify the inter-rater reliability between the mocap outputs and goniometer measurements for the right hip, Pearson's product-moment correlations were applied to the hip data. A strong correlation (Evans, 1996) was identified between the BVH and goniometer measurements (BVH: $r(23) = .688$, $p < .001$), and very strong correlations (Evans, 1996) were found between the remaining output measurements and the goniometer measurements (IGS-Bio:

$r(23) = .986, p < .001$; AnimaView solid: $r(23) = .990, p < .001$; AnimaView Wireframe: $r(23) = .990, p < .001$; AnimaView Actor File solid: $r(23) = .989, p < .001$; AnimaView Actor File Wireframe: $r(23) = .989, p < .001$).

The Bland-Altman plots and t-tests (Table 3-12) identified that the BVH and goniometer measurements agree with one another. However, all other outputs had mean differences that were significantly different from zero, and cannot be used with confidence for mocap for postural assessments of the right hip as they systematically misestimate the angles of extension and flexion.

Table 3-12: Right Hip Comparison of the Values Obtained With the Goniometer with the Values Obtained via the Six Outputs

	N	Mean Difference	SD	Upper Bound	Lower Bound	t value	p-value
IGS-Bio	25	15.231	8.188	31.279	-0.817	9.301	< .001
AnimaView Solid	25	10.070	5.695	21.232	-1.092	8.841	< .001
AnimaView Wireframe	25	10.106	5.694	21.266	-1.055	8.873	< .001
AnimaView Actor File Solid	25	4.717	5.988	16.454	-7.020	3.939	.001
AnimaView Actor File Wireframe	25	4.737	5.983	16.464	-6.990	3.958	.001
BVH	25	7.797	26.799	60.323	-44.729	1.455	.159

3.2.2.6.2 Left

The results of the Pearson's product-moment correlations for the left hip angles show that all mocap output measurements for the left hip were very strongly correlated with the goniometer measurements (IGS-Bio: $r(23) = .994, p < .001$; AnimaView solid: $r(23) = .994, p < .001$; AnimaView Wireframe: $r(23) = .994, p < .001$; AnimaView Actor File solid: $r(23) = .993, p < .001$; AnimaView Actor File Wireframe: $r(23) = .993, p < .001$; BVH: $r(23) = .882, p < .001$) (Evans, 1996). This indicates that the mocap measurements all vary in a similar way to the goniometer.

Bland-Altman plots and t-tests were then applied to the data to identify the degree to which the output and goniometer measurements agreed. As can be seen in Table 3-13, the mean differences for IGS-Bio, both uncalibrated (Actor File) AnimaView output measurements and the BVH outputs were not significantly different from zero, which suggests that they might agree with the goniometer measurements. IGS-Bio has the smallest mean difference suggesting that it is closest to the goniometer. Despite this, the upper and lower bounds indicate that the score can vary between roughly $\pm 13^\circ$. Both calibrated outputs had mean differences that were significantly different from zero, which indicates a lack of agreement with the goniometer measurements.

Table 3-13: Left Hip Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Upper Bound	Lower Bound	<i>t</i> value	<i>p</i> -value
IGS-Bio	25	0.337	6.550	13.175	-12.502	0.257	.799
AnimaView Solid	25	3.240	6.437	15.857	-9.378	2.516	.019
AnimaView Wireframe	25	3.192	6.329	15.597	-9.213	2.522	.019
AnimaView Actor File Solid	25	-4.483	6.893	9.028	-17.993	-3.252	.003
AnimaView Actor File Wireframe	25	-4.396	6.922	9.172	-17.964	-3.175	.004
BVH	25	-2.247	17.717	32.478	-36.971	-0.634	.532

Note: Bonferroni correction: adjusted significance threshold = .008

3.2.2.7 Knee

3.2.2.7.1 Right

Pearson's product-moment correlations indicated that all mocap output measurements, for the right knee, were very strongly correlated (Evans, 1996) with the goniometer measurements (IGS-Bio: $r(25) = .963$, $p < .001$; AnimaView solid: $r(25) = .981$, $p < .001$; AnimaView Wireframe: $r(25) = .981$, $p < .001$;

AnimaView Actor File solid: $r(25) = .981$, $p < .001$; AnimaView Actor File Wireframe: $r(25) = .981$, $p < .001$; BVH: $r(25) = .981$, $p < .001$). These correlations indicate a good level of inter-rater reliability and covariance. Bland-Altman plots and t-tests were then applied to the data to identify the extent to which the outputs agreed with the goniometer measurements (Table 3-14). These analyses suggest that only the IGS-Bio output measurements agree with the goniometer, and might be used with some confidence for measuring the right knee extension.

Table 3-14: Right Knee Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Upper Bound	Lower Bound	t value	p-value
IGS-Bio	27	-3.321	10.756	17.761	-24.403	-1.604	.121
AnimaView Solid	27	6.679	8.459	23.259	-9.900	4.103	< .001
AnimaView Wireframe	27	6.715	8.424	23.226	-9.796	4.142	< .001
AnimaView Actor File Solid	27	12.910	8.485	29.542	-3.721	7.906	< .001
AnimaView Actor File Wireframe	27	12.977	8.568	29.771	-3.818	7.869	< .001
BVH	27	7.752	7.898	23.233	-7.729	5.100	< .001

3.2.2.7.2 Left

Pearson's product-moment correlations revealed that all of the mocap output measurements, for the left knee, were very strongly correlated (Evans, 1996) with the goniometer measurements indicating that the measurements have good inter-rater reliability and co-vary with the goniometer (IGS-Bio: $r(25) = .985$, $p < .001$; AnimaView solid: $r(25) = .994$, $p < .001$; AnimaView Wireframe: $r(25) = .994$, $p < .001$; AnimaView Actor File solid: $r(25) = .994$, $p < .001$; AnimaView Actor File Wireframe: $r(25) = .994$, $p < .001$; BVH: $r(25) = .922$, $p < .001$).

Bland-Altman plots and t-tests were applied to the data to identify whether the mean differences were significant from zero (Table 3-15). The BVH data was the

only output in agreement with the goniometer measurements. The rest of the output measurements had mean differences that were significantly different from zero. Consequently, when looking to analyse mocap data for the left knee, the BVH output provides the closest measurement to the goniometer.

Table 3-15: Left Knee Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Upper Bound	Lower Bound	t value	p-value
IGS-Bio	27	15.493	6.864	28.946	2.039	11.728	< .001
AnimaView Solid	27	4.193	4.592	13.194	-4.808	4.744	< .001
AnimaView Wireframe	27	4.155	4.575	13.121	-4.812	4.719	< .001
AnimaView Actor File Solid	27	7.614	4.799	17.021	-1.792	8.244	< .001
AnimaView Actor File Wireframe	27	7.629	4.800	17.037	-1.780	8.258	< .001
BVH	27	7.405	21.608	49.755	-34.946	1.781	.087

Note: Bonferroni correction: adjusted significance threshold = .008

3.2.2.8 Ankle

3.2.2.8.1 Right

Pearson's product-moment correlations found very strong positive correlations for the right ankle angles, as defined by Evans (1996), between all mocap outputs and the goniometer measurements (IGS-Bio: $r(9) = .973$, $p < .001$; AnimaView solid: $r(9) = .959$, $p < .001$; AnimaView Wireframe: $r(9) = .959$, $p < .001$; AnimaView Actor File solid: $r(9) = .957$, $p < .001$; AnimaView Actor File Wireframe: $r(9) = .959$, $p < .001$; BVH: $r(9) = .958$, $p < .001$). This indicates that the different outputs and the goniometer measurements vary similarly in how they measure the different angles.

Table 3-16: Right Ankle Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Upper Bound	Lower Bound	<i>t</i> value	<i>p</i> -value
IGS-Bio	11	12.381	4.114	20.444	4.318	9.982	< .001
AnimaView Solid	11	10.618	4.743	19.915	1.322	7.425	< .001
AnimaView Wireframe	11	10.644	4.699	19.855	1.433	7.512	< .001
AnimaView Actor File Solid	11	15.116	4.825	24.573	5.660	10.391	< .001
AnimaView Actor File Wireframe	11	14.987	4.710	24.218	5.756	10.554	< .001
BVH	11	5.531	4.961	15.255	-4.193	3.697	.004*

Note: Bonferroni correction: adjusted significance threshold = .008

Although the correlations indicated a similarity in measurement, Bland-Altman plots and t-tests were applied to the data to identify the extent to which the outputs agreed with the goniometer measurements. All means for each output measurement were significantly different from zero and do not agree with the goniometer measurements (Table 3-16). This indicates that if looking to assess the right ankle, alternative or additional forms of mocap should be used.

3.2.2.8.2 Left

Pearson's product-moment correlations, for the left ankle, found very strong positive correlations (Evans, 1996) between all mocap outputs and the goniometer measurements (IGS-Bio: $r(9) = .981$, $p < .001$, AnimaView solid: $r(9) = .975$, $p < .001$, AnimaView Wireframe: $r(9) = .980$, $p < .001$, AnimaView Actor File solid: $r(9) = .987$, $p < .001$, AnimaView Actor File Wireframe: $r(9) = .986$, $p < .001$, BVH: $r(9) = .972$, $p < .001$). This means that the Synertial measurements have a very strong level of inter-rater reliability, and varied in a very similar way to the goniometer.

Table 3-17: Left Ankle Comparison of the Values Obtained With the Goniometer with the Values Obtained Via the Six Outputs

	N	Mean Difference	SD	Upper Bound	Lower Bound	t value	p-value
IGS-Bio	11	16.677	5.743	27.934	5.421	9.631	< .001
AnimaView Solid	11	18.384	5.283	28.739	8.028	11.540	< .001
AnimaView Wireframe	11	18.368	5.011	28.190	8.546	12.157	< .001
AnimaView Actor File Solid	11	22.421	2.716	27.744	17.098	27.379	< .001
AnimaView Actor File Wireframe	11	22.401	2.814	27.916	16.886	26.405	< .001
BVH	11	13.504	4.903	23.115	3.894	9.135	< .001

The Bland-Altman plots and t-tests applied to the left ankle data reveal a significant systematic lack of agreement between the goniometer and mocap output measurements, as shown in Table 3-17. This suggests that the mocap suit cannot be used with confidence for measuring flexion and extension of the left ankle. Consequently, alternative or additional forms of mocap should be used for these measurements. The smallest mean difference was measured for the BVH files, however, the upper and lower bounds vary from 24° to 3°, which is close to over half the ankles range of movement for this participant.

3.2.3 Summary of Findings

To summarise the findings detailed above, Table 3-18 presents the instances of agreement between the Synertial outputs and the goniometer measurements for the whole body and each segment. The Synertial outputs and goniometer measurements were considered in agreement when a significant correlation and non-significant Bland-Altman test were found. For the body as a whole, IGS-Bio and the uncalibrated AnimaView outputs may be suitable for capturing flexion and extension from an individual wearing the mocap suit. However, there was no agreement between any of the Synertial outputs and the goniometer measurements for the right shoulder, both elbows, both wrists and both ankles. For the joints where agreement was found, there was variation across the

Synertial outputs, and it was not the same between left and right measurements of the same type of joint. It should be noted that besides the neck, no other segments had agreement between the Actor File, AnimaView output measurements and the goniometer.

Table 3-18: Agreement between Synertial Outputs and Goniometer across Whole Body and Individual Segments

		IGS-Bio	AnimaView Solid	AnimaView Wireframe	AnimaView Actor File Solid	AnimaView Actor File Wireframe	BVH
Whole Body		✓	✓	✓			
Neck			✓	✓	✓	✓	✓
Trunk							✓
Shoulder	Right						
	Left	✓					
Elbow	Right						
	Left						
Wrist	Right						
	Left						
Hip	Right						✓
	Left	✓	✓	✓			✓
Knee	Right	✓					
	Left						✓
Ankle	Right						
	Left						

3.3 Discussion

The aim of this study was to investigate the effectiveness of an inertial mocap system, and to identify its suitability for use in future experiments. Both the mocap suit as a whole and individual joints have been assessed for level of agreement between the Synertial outputs and the goniometer measurements. The level of agreement varied between joints and between Synertial outputs. This suggests a lack of reliability in the suit's ability to effectively capture and record angles of flexion and extension across the joints, when compared to goniometer measurements. These findings suggest that when using the mocap suit, an alternative form of mocap should be used alongside it, to provide context to the mocap output and a confirmatory check on the effectiveness of the suit. This is important because mocap remains the only means of measuring postures in occluded spaces and provides some insight into the postures of individuals within those conditions. It is recommended that further work be done to improve the effectiveness of inertial sensors, so that they can be used with confidence within occluded environments, where they are needed most.

As detailed in the literature review (Section 2.2.1.2.1.3) and method (Section 3.1.1.2), there are known limitations to the accuracy of the suit due to the type of joint being captured by a sensor, and environmental factors. These errors are referred to within the literature as artefacts, and cover all errors introduced to recorded data. Within inertial mocap, there are four specific types of artefacts that result from the type of joint being captured by a sensor as have been described in the Section 2.2.1.2.1.3. The first are clothing artefacts, which are caused by the sensors moving with clothing rather than the body; the second are tissue artefacts, which are caused by a lack of a bone for a sensor to be placed against, this results in the sensor moving without their body segment. The third type of artefact are referred to as skin artefacts and are a result of the sensors moving with loose skin rather than the wearer's body. The last type of artefact is kinetic, which is where the sensor continues to move because of momentum despite the wearer being stationary. The reason this experiment recorded static angles was

to counter kinetic artefacts; however, the other factors may have affected the effectiveness of the suit.

The manufacturer provides information regarding the level of accuracy expected from the suit. Table 3-1 shows the type of accuracy that is expected for each joint by Synertial; the neck, elbows, wrists, knees, and ankles are described as being Alpha segments where a high level of accuracy was anticipated under almost every condition and range of motion (Synertial, 2012). A reason for this is that the associated sensors are placed and tightened against a bone, which should minimise the risk of tissue artefacts. Therefore, it was expected that for these joints, agreement between the Synertial outputs and the goniometer would be found. However, the only joint where this was the case was the neck where agreement was found for all but the IGS-Bio output. A single output measurement agreement was found for both knees: IGS-Bio for the right knee and BVH for the left knee. The expected similarity was not found for the remaining Alpha joints (wrists and ankles). It should be noted that other researchers have found the head and ankle joints to show the greatest difference in their investigations of the accuracy of inertial mocap suits (Cloete & Scheffer, 2008; Lebel et al., 2017). Although the head and neck results are not consistent with Lebel and colleagues (2017) findings, the ankle results support them. It is possible that skin or clothing artefacts had an effect on the elbow, knee and ankle joints. For the wrists, the gloves were strapped directly and securely onto the hand, so clothing and skin artefacts could not have been the cause of the lack of agreement unless the glove had come loose during recording. This is unlikely as care was taken throughout the data collection process to ensure that all straps remained secure.

The shoulder and hip are considered to be able to achieve Beta levels of accuracy by the manufacturer (Synertial, 2012). This means that accuracy can only be achieved with the use of rigid plates fasted at both ends of the segment to compensate for the tissue artefact that affects these measurements. This technique was not used as the intended use for the mocap suit is within manufacturing environments where metal plates attached to the upper arm and leg may hinder movement in and out of structures. Therefore, this may explain

why none of the Synertial output measurements agreed with the goniometer measurement for the right shoulder. However, there was agreement between at least one Synertial output for the left shoulder (IGS-Bio), and right hip (BVH), and four for the left hip (IGS-Bio, AnimaView Solid and wireframe, and BVH).

It was anticipated that as Synertial considers the trunk a Complex segment, it would not be possible to find agreement between any of the outputs and the goniometer. However, agreement was found between the BVH and goniometer measurements. This indicates that when stationary, it is possible to gain a potentially effective measure of trunk flexion and extension. Despite limitations in the agreement between the Synertial output measurements and the goniometer measurements, the strong positive correlations indicate that the Synertial output measurements changed in the same way as the goniometer measurements changed, and there was just a systematic mismeasurement in how much they changed. Therefore, additional research could look to identify whether it is possible to quantify these variations and producing a correction to the calculations completed by the mocap system as a means to address these limitations. A way to achieve this may be through reviewing the upper and lower bounds produced from the Bland-Altman plots, because they indicate how systematic the differences between the measures were.

Although using machines or robots enable greater precision when setting the joint angles (Brennan, et al., 2011; Brodie et al., 2008; Cutti et al., 2006), the suit was assessed on a human. The reason for this was that it provides the best representation of how an individual in the suit affects its effectiveness and the movement of sensors inside the suit, or how the suit moving over the body may have contributed to the variations in the measurements between the goniometer and the mocap suit. Therefore, measurement error is likely to have been caused by human aspects, but these errors will also be measured in real use scenarios and, consequently, should be taken into account.

Further research is required to assess the reliability of the Synertial IGS-180-I mocap suit, particularly to understand how the outputs provided by the BVH data and whether other joints require the level of correction that was identified with the

knee data. The use of an inertial mocap suit is a step closer to having an objective method of measuring postural angles and a step away from current subjective methods of postural assessment for occluded conditions (Hamilton et al., 2011). Therefore, this will address the current limitations within postural assessments which are unable to provide absolute indicators of damage due to their hypothetical quantification and scoring of magnitudes of interactions (Kilbom, 1997; Li & Buckle, 1999; Roman-Liu, 2014; Winkel & Westgaard, 1992). When fully effective this information will be critical for providing greater insight into the effects that postures and movement may have on MSDs and their development, and the ability to link particular postures to specific damage (Smith, 1996). A single participant was assessed within this study, which means it is not possible to understand whether there are any differences in how the mocap suit records individuals with different physiologies. The participant was within the 96th percentile of the age range 19 to 65 anthropometric tables (Pheasant & Haslegrave, 2006), which represents the physiology of a small proportion of the population so it not representative of the average population who may wear the mocap. It is possible that accuracy and effectiveness of the suit will vary on individuals of different weights, heights, and morphologies with greater or fewer artefacts being found depending on the fit of the suit. Future work should look to investigate the effect individual differences may have on the suit, or ensure that their participant pool consists of more representative morphologies to account for this as an experimental factor.

With regard to the methodology used, traditional clinical methods were used including a goniometer and a computer-generated protractor. It provided an output that matched the measurements taken using the protractor goniometer, making them directly comparable. A limitation of these methods is that they are not the gold standard in mocap and, therefore, measurement errors are likely. The Vicon system is considered the gold standard (Dutta, 2012); consequently, the next step in the assessment of the reliability of the Synertial IGS-180-I would ideally be to compare the mocap suit with a Vicon system using static postures. This would give a clearer indication of the accuracy of the suit. Additionally, the results could be compared to those presented within this Chapter to provide

evidence of the effectiveness of the goniometer and computer-generated protractor. It was not possible to gain access to a Vicon system for this study and this direction of research was away from the psychological focus of this PhD, therefore, it was not explored further within this research.

As stated in the literature review, it is possible to integrate the output from the mocap suit into CAD programs that include DHM applications. Future work should look to investigate the accuracy of this translation before inertial mocap systems can be used with any confidence. As none of the presently available DHM CAD software packages currently have the full postural analysis tools that are required for the work completed within this thesis, this line of inquiry has not been followed. Where DHM CAD packages are able to accurately integrate and assess human movement data using mocap, the physical impact of workstations or components that operators are required to work with or in can be better understood and the negative impact can be designed against. The ability to achieve this with inertial or non-optical systems is critical for manufacturing activities where postures are occluded or completely confined. The ability to understand what these working postures look like and how they may be contributing to the development of MSDs would provide proactive ergonomics with an effective tool for understanding how designs and redesigns will be interacted with and look to design out as far as is reasonably possible aspects that cause harm to operators.

Based on the results of this case-study, it is recommended that additional forms of mocap be used alongside the IGS-180-I in future experiments. The technology is not stable and effective enough to be used independently from other forms of mocap with confidence. The use of cameras would help to contextualise and record additional posture data.

4 EXPERIMENT: PREDICTIVE COGNITIVE AND PHYSICAL INTERACTION

The experiment described in this Chapter aims to address Objective 2: to establish the combined parameters of physical strain and mental workload. The aim of this objective was to identify an equation to predict the effect of physical strain on mental workload, which could be integrated into a DHM CAD package. To do so, physical strain needed to be reliably measured, and this is why the mocap suit was reviewed in Chapter 3. The findings from Chapter 3 informed the experimental design and led to the inclusion of cameras in addition to the mocap suit for capturing physical strain.

An experiment was identified as the most suitable means of achieving these objectives, as it would enable controlled conditions, which would reduce the impact of extraneous variables on mental workload and physical strain. Furthermore, it enables the manipulation of the IV which provides a corresponding change in the DV. This was achieved by having the participants adopt postures simulating three different types of physical strain (IV) while completing two different tasks of differing complexity which affects workload capacity (DV). As both demands are drawing from the same energy resource (Kahneman, 1973; Norman & Bobrow, 1975; Van Acker et al., 2018), as energy is drawn to maintain the physical demand a corresponding effect is expected for cognitive workload capacity. Subjective and objective overall and mental workload was captured, and physical strain was recorded and analysed to produce predictive measures of mental workload. This manipulation made it possible to assess the effect of physical strain on workload and establish an equation that accounted for both high and low level of workload as defined by task complexity. The method, results, and implications are presented in full and discussed below.

4.1 Method

4.1.1 Research Ethics

This research was approved by the Cranfield University Research Ethics System (CURES) (see Appendix 6.3D.3) and conducted in accordance with the Cranfield Research Integrity Policy, the British Psychological Society's (BPS) Code of Human Research Ethics, and the General Data Protection Regulation (GDPR) 2018. Details of adherence to these policies can be found within the procedure detailed in Section 4.1.6.

4.1.2 Participants

Forty-two participants volunteered in response to internal Cranfield University advertisements uploaded on the announcement board on the University's intranet and attached to physical announcement boards around campus. Fourteen female and twenty-eight male staff and students, between 22 and 59 years (mean: 31, SD: 10.36). They ranged in height from 155.0cm to 194.7cm.

4.1.3 Design

A repeated measures experimental design was used to investigate the effect of physical strain on mental workload. This was achieved by increasing physical stress through the posture adopted by participants while they complete a low complexity and a high complexity task (IVs). The effect of this change was measured on mental workload (DV). The experiment was conducted in the Intelligent Automation (IA) laboratory at Cranfield University as this allowed controls over environmental variables and privacy.

A repeated measures experimental design was chosen because it mitigates the variation between individual's heart rate data and subjective workload scores, and allows for effective comparison between the conditions. If two consecutive conditions required the participants to work in high strain this was likely to lead to the HRV data showing an impact of the continuous physical strain rather than the strain associated with the condition being measured. Therefore, counterbalancing was employed to ensure that order effects did not impact the

results. This involved alternating the tasks from participant to participant and rotating the order in which postures were adopted so that results could not be attributed to practice or fatigue.

4.1.3.1 Independent Variables

4.1.3.1.1 Task design

To establish a comprehensive model and equation, the effect of physical strain on mental workload needed to account for high and low mental workload conditions. This was achieved by changing task complexity; to achieve this two tasks were used within this experiment, a low complexity task and a high complexity task. Campbell's (1988) definition of task complexity details low complexity tasks having an end state and a single path, and complexity being increased with interdependence between parts of the task. Therefore, both low and high complexity tasks completed within this experiment had one end state with a single path, and the high complexity task had interdependence between parts of the task, where there was none for the low complexity activity (Campbell, 1988).

The tasks were derived from current processes completed on the Airbus A320 wing flow-line, to provide ecological validity to the experiment and an understanding of the current working conditions of operators. Existing aerostructures manufacturing manual processes were considered to be ideal use cases due to a significant level of labour intensive effort and time required. Currently, little is known about the impact they have on operators' physical and mental well-being and performance and, therefore, this experiment provides insight into these factors. Furthermore, this research would contribute to ongoing research into improving output and production rates via improved working conditions (Judt et al., 2020; Lockett et al., 2014).

To identify the low complexity task, a task was chosen from a review of two manual in-tank sealing processes because of their relative simplicity: application of sealant on rib two for the helium leak test, and in-tank sealing of bolts. A single day observation visit to the Airbus Broughton facility was conducted to review both tasks and identified the in-tank sealing of bolts as the most suitable process

for replication in the experiment because it could be adapted to colouring in circles, and limited additional confounding factors. The helium leak test sealing was not suitable for translation to this experiment because it required the application of clay or putty. This would have made effective resetting of the experiment between each condition challenging. Additionally, the operators who completed the task stated that the task contributed to muscle fatigue in the forearms, wrists and fingers. As a result, there was concern that this may have a confounding effect on physical strain during the experiment, by increasing strain that was already caused by the imposed postural working conditions.

The higher complexity task was chosen from two system installation tasks marked for redesign because of a build philosophy change that involved pre-equipping major components of a single aisle wing prior to installation (Judt et al., 2016). The tasks included equipping electrical harnesses, raceways, and bleed air ducting, which were then investigated for the effectiveness of orientation and position changes to identify effects on assembly performance improvements (Judt et al., 2020). Bleed air duct installation was identified as the most suitable task for the higher complexity condition for the experiment due to: the interdependency of the task steps, and because of the variation in task steps and components. The electrical harness and raceway installation was a more repetitive task with fewer interdependent task steps and, therefore, were assessed as unsuitable as the high complexity task for this experiment.

4.1.3.1.2 Posture

To create the varying levels of physical strain required to assess their impact on workload, three levels had to be created with reference to REBA, they included:

1. Overhead (OH) standing: to put the legs, back, neck, and arms in physical strain.
2. OH sitting: to support the legs and back and to put the arms and neck under physical strain.
3. Neutral: the component was presented in the ideal orientation of 45°, all joints were supported or able to be held in a neutral position.

4.1.3.2 Dependent Variables

Mental workload was measured using both objective and subjective methods, HRV and the NASA-TLX respectively. This was to provide redundancy in the experiment in the event that the heart rate monitor malfunctioned, and overcome limitations of both techniques as mentioned in the literature review 2.2.1. Additionally, it was intended to contribute to research investigating the effectiveness of these measures as a means of assessing mental workload as detailed in Section 2.2.1 of the literature review.

4.1.4 Apparatus

A representation of the laboratory layout is presented in below in Figure 4-1. It shows the relative placement of the robot, robot controller, mock-up, mocap PC, the task parts, stadiometer, cameras, and the light curtain. It also shows the table and chair where the participants completed the surveys between conditions.

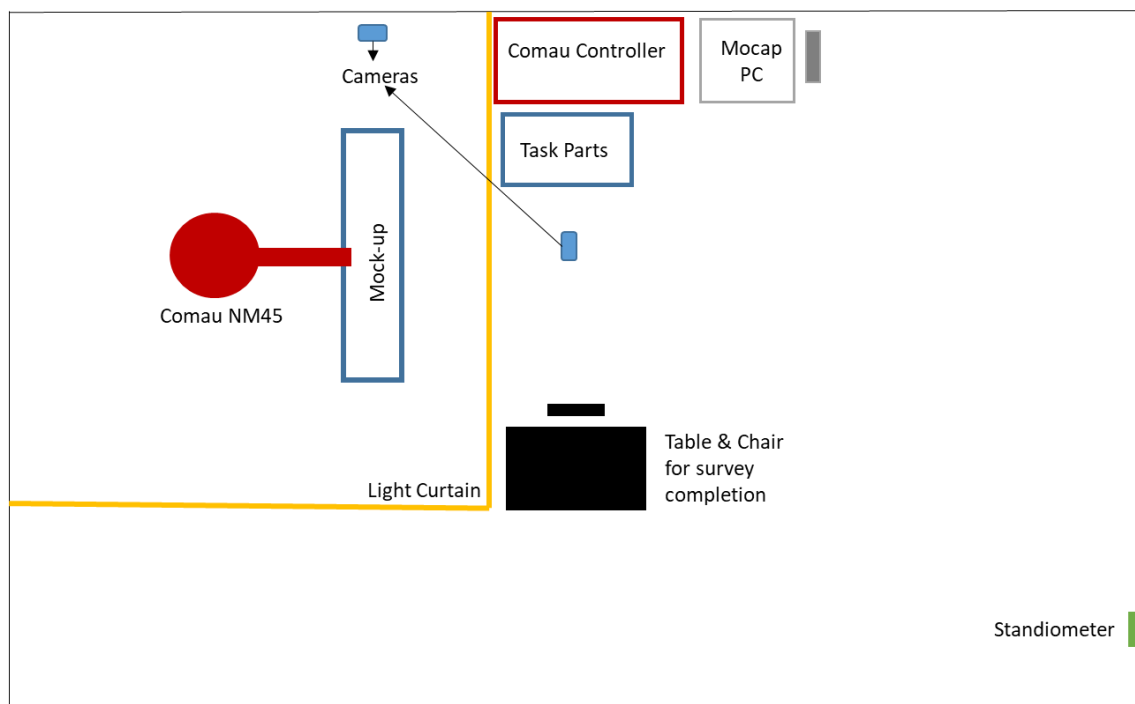


Figure 4-1: Laboratory Layout

4.1.4.1 Robot

A Comau NM45 industrial robot, with a 45kg payload, was used to hold the mock-up rather than using a fixed mock-up because the robot enabled easy and repeatable movement of the mock-up into the required heights and orientations.

4.1.4.2 Mock-up

The mock-up used within this experiment was a partial A320 Airbus wing leading edge structure and can be seen in Figure 4-2. Figure 4-3 provides a labelled diagram showing the structural components of a wing for reference. The mock-up was approximately matched, in dimensions and rib position, to an A320 wing front spar and was mounted to the robot. The mock-up consisted of a 2 m × 0.3 m machined aluminium spar fitted with six ribs. Each rib had the required system component openings and were bolted onto the front spar. The ribs were covered by an aluminium sheet skin and medium-density fibreboard (MDF) bottom skin plate. This created a bottom panel opening of the required dimensions through which both tasks were conducted, these dimensions replicated the available internal space in an A320 leading edge.

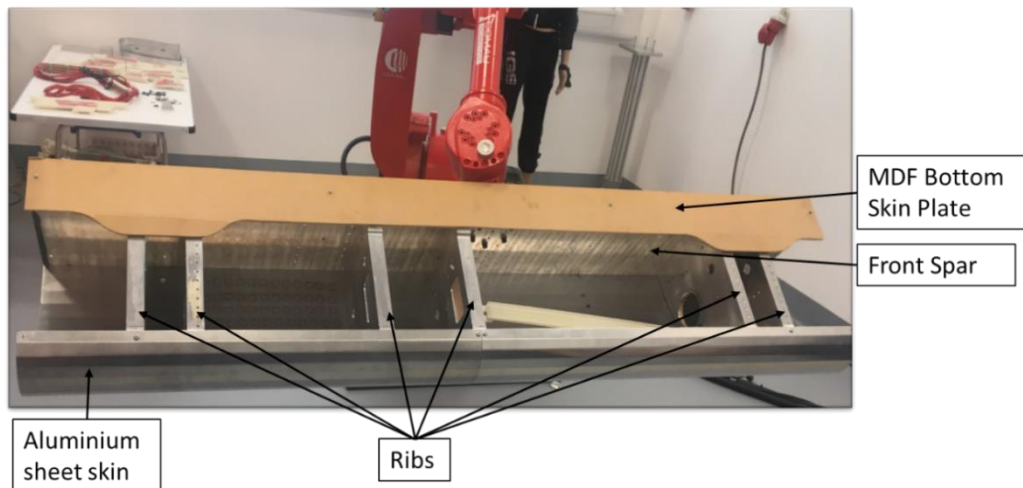


Figure 4-2: Labelled Mock-up Identifying Structural Components

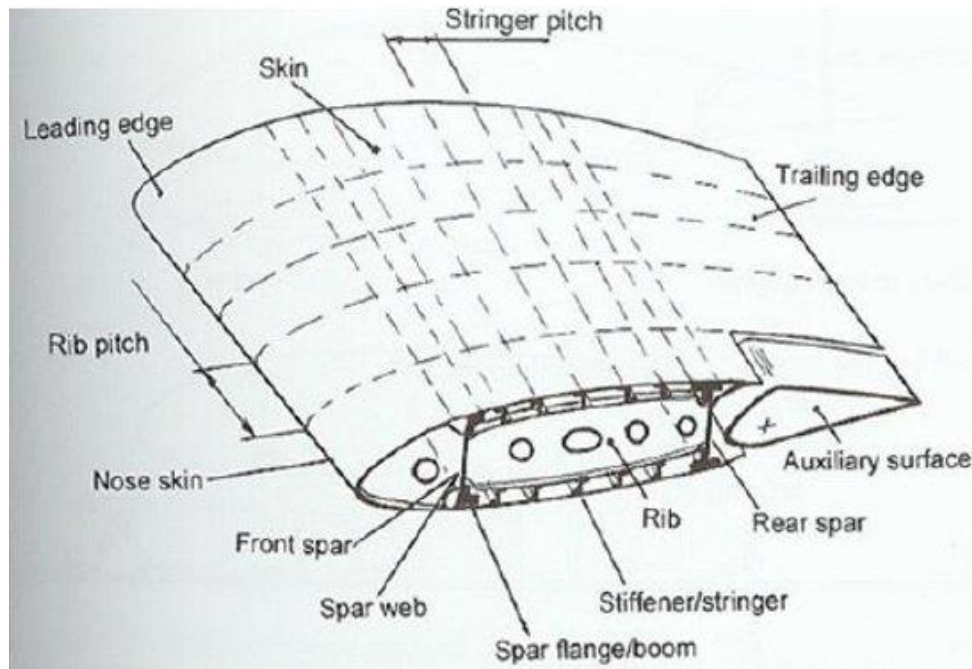


Figure 4-3: Diagram of a Labelled Cut-Section of a Wing (Howe, 2004)

The low complexity and high complexity tasks were both located within the mock-up. The low complexity task was positioned on the right hand side of the spar (when in the neutral condition as shown in Figure 4-2), and the circles seen in Figure 4-4 were drawn onto the aluminium sheet skin and covered in a plastic film to protect them from being erased between conditions. The high complexity task was completed on the left-hand side of the mock-up when presented in the neutral condition. The inclusion of both tasks within the same structure ensured that the working height was the same between each task-based condition. This safeguarded against any variation in physical strain exposure for the participants between the tasks.

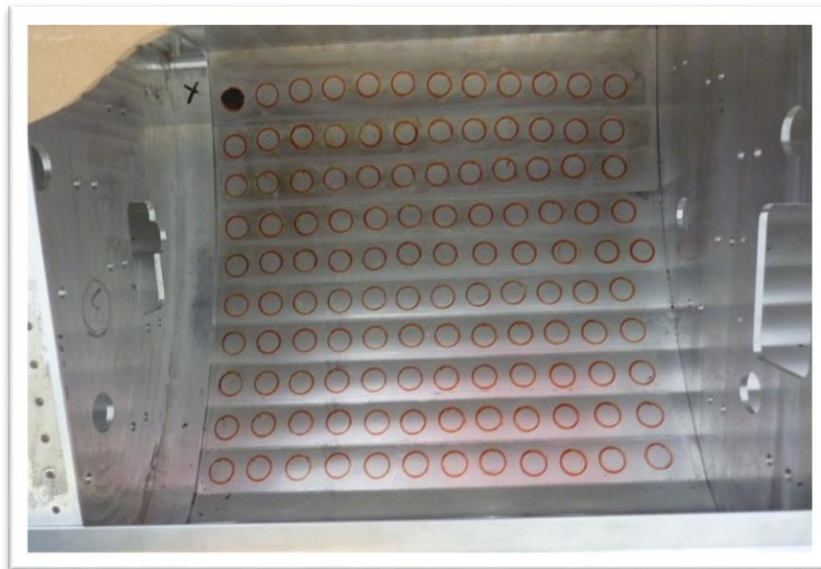


Figure 4-4: Bolt Sealant Task

4.1.4.3 Motion Capture

Two forms of mocap were employed to record the postures of the participants as they completed the tasks. The wireless full body inertial mocap suit (IGS-180-I, Synertial Ltd.) that was assessed in experiment detailed in Chapter 3 and two Panasonic Lumix cameras: a DMC-G7 and an FZ38 (Panasonic Corporation) which were placed onto tripods to ensure effective activity recording. As explained in the Literature Review in Section 2.2.1.2.1 the use of cameras to record postures is a traditional form of mocap, with two cameras used to overcome any perspective bias.

The depth of the mock-up and task requirements meant that the participants were required to complete the tasks while partially within the structure. As a result, participants head, neck, forearms, and wrists were partially obscured by the mock-up during both tasks. Therefore, the mocap suit was identified as the most suitable recording tool as it does not require line of site. The motion data from the suit was recorded in IGS-Bio. Despite clear limitations with the use of the Actor Files, where they were only found to be effective for the neck as detailed in Chapter 3 Section 3.2.3, effective limb positioning which is provided by the Actor

Files was required to assess the posture of the participants. Therefore, Actor Files for each participant were created in AutoCal and applied to the motion data in IGS-Bio to assess the mocap data.

The cameras were additionally utilised due to the limitations of the inertial mocap suit identified in Chapter 3. Although the cameras were vulnerable to occlusion restrictions, they provided additional visual contextual information (participant use of the mock-up for stabilising support, which reduces REBA scores). Furthermore, they provided an additional form of mocap which meant that data was not lost when errors occurred with the suit. The use of two cameras placed at 90° from one another helped to reduce occlusion issues inherent in video mocap (Chiasson et al., 2012; Li & Buckle, 1999) and some of the perspective bias.

4.1.4.4 Physiological Measure

As identified in the Literature Review (Section 2.2.1.1.3) cardiovascular measures were identified as the most suitable physiological measure of workload, with the HRV statistic IBI utilised due to the length of each condition, IBI requires a minimum of a minutes data collection (Shaffer & Ginsberg, 2017). Objective measures counter limitations of primacy and recency effects seen with subjective measure, because HRV is measured across the whole task and provides an objective measure of mental workload.

The objective mental workload measure of HRV, was recorded using the MindWare Technologies Ltd. Mobile Physiological system (50-2303-01) and provided an ECG. This system records the participant's HR and provides an analysis of HRV. The system required the placement of three electrodes onto the participant's torso with lead wires inserted into the MindWare mobile system. The mobile system hung from a lanyard around the participant's neck to prevent electrical signal interference with the mocap system.

4.1.5 Materials

4.1.5.1 NASA-TLX

Workload was measured using the NASA-TLX (Hart & Staveland, 1988). As explained in Section 2.2.1.1.2 of the Literature Review this measure was used because it provides a subjective understanding of the participant's perception of the task. Furthermore, it was identified as an effective measure to establish a predictive equation as its output is numerical, it has been shown to be sensitive at the dimension level (Hill et al., 1992; Nygren, 1991; Rubio et al., 2004) and has been found to support a general predictive model for experienced workload (Nygren, 1991)

The scale is a widely used subjective workload rating tool comprising of six scales to measure the six principal dimensions of workload:

- Mental Demand (Very low – Very high)
- Physical Demand (Very low – Very high)
- Temporal Demand (Very low – Very high)
- Performance (Perfect – Failure)
- Effort (Very low – Very high)
- Frustration (Very low – Very high)

The Mental Demand Subscale can be used as a subjective measure of mental workload (DiDomenico & Nussbaum, 2008). The scale can be used in either its traditional or raw form. The traditional form involves weighting the mental workload variables bulleted above before rating the scales, while in the raw version the weighting process is not completed to reduce the time required for completion. Due to the time requirements in the experiment, and because it is as valid a measure as the traditional form (Bittner et al., 1989; DiDomenico & Nassbaum, 2008; Moroney et al., 1992), the raw form was used. Additionally it can be used to assess individual dimensions of workload (DiDomenico & Nussbaum, 2008; Shaikh et al., 2012), in the case of this experiment Mental Demand was of interest.

4.1.6 Procedure

4.1.6.1 Briefing and Consent

In order to adhere to experimental ethical guidelines, upon arrival the participants were briefed regarding the nature of the study, ethical aspects and entitlements, and the key components of the procedure. They were then asked to provide their consent to take part in the experiment and their biographical information. A copy of the brief and consent form can be found in Appendix 6.3D.1.

4.1.6.1.1 Heart Rate Monitor and Mocap Set Up

The heart rate monitor was attached to the participants, which required careful placement of the ECG electrodes. Three electrodes were placed on three points on each participant's trunk: the right clavicle, the left bottom rib and the right bottom rib (a visual of the electrode placement can be found in Appendix 6.3D.1). Prior to electrode placement, the attachment site was cleaned with an alcohol wipe to ensure effective attachment. The lead wires were then all attached to the electrodes: the negative wire was attached to the clavicle electrode, the positive wire to the left rib electrode, and a ground lead wire to the right rib electrode. A service loop was created on each of the leads using micropore tape, to prevent movement noise and the electrode from being pulled off the skin in the event of the lead being pulled on. Each wire was then inserted into the corresponding ports on the MindWare mobile device.

Prior to putting the mocap suit on, the participant's height was measured using a wall mounted stadiometer located in the laboratory (Figure 4-1). The participants were then asked to put on the mocap suit, and the bottom of the suit was attached to the back origin sensor (sensor 10 as shown in Figure 3-2 in Section 3.1.5.1). The suit was then plugged into the power source and turned on. The placement of each of the sensors was assessed and, if needed, they were moved into the correct position as shown in Figure 3-2. All strapping was tightened to reduce the possible effect of artefact noise. Two calibration pictures were taken of each participant for the creation of the Actor File. This enabled the manikins to be calibrated from the standard manikin stature to the participant's stature, which includes adjustment of limb length. The participant stood inside a calibration jig

wherein front and right-side photos were taken from two cameras placed on tripods positioned at waist height and 3 meters away from the jig. Figure 4-4 shows an example of these two pictures within the calibration software AutoCal.

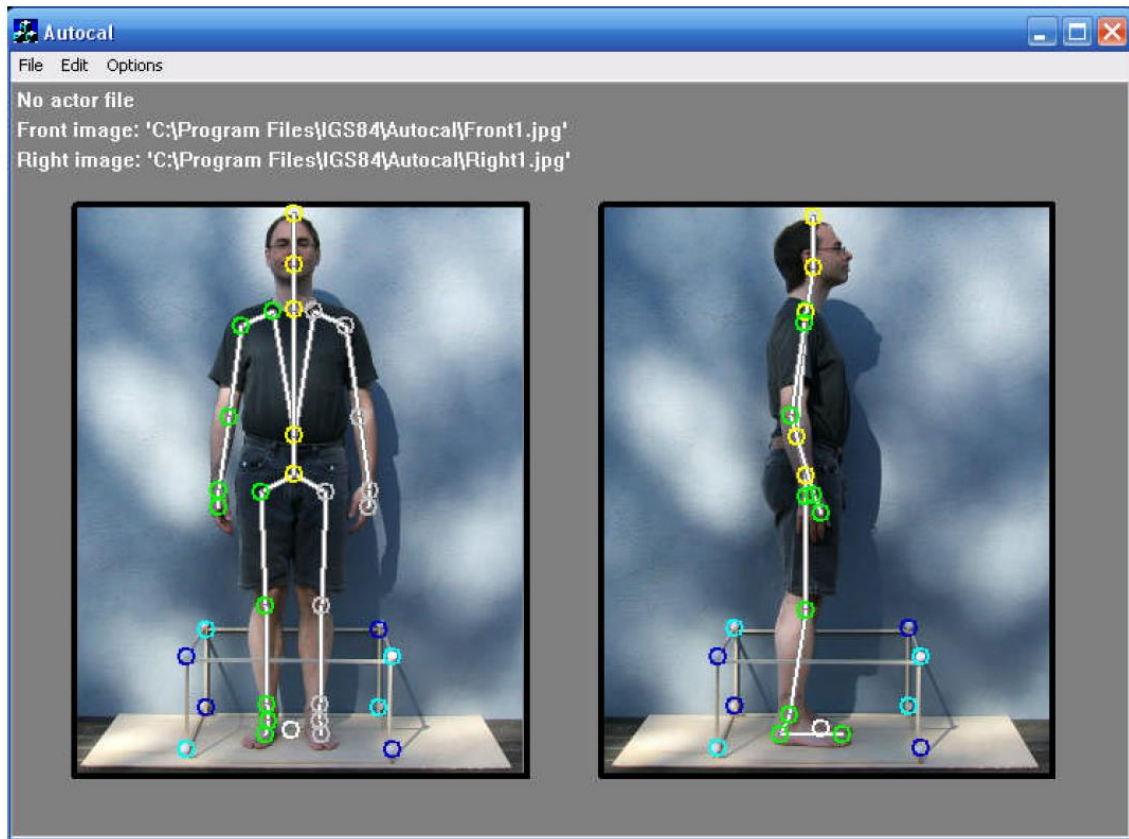


**Figure 4-5: Front and Side Calibration Pictures within AutoCal Software
(Synertial, 2009)**

The cameras were lined up with the jig central to the focal point of the camera on the horizontal plane, and with an equidistant gap between the rear and front upright poles of the jig as outlined in the Synertial user guide (Synertial, 2011). The photos were later uploaded to a PC for the Actor File to be created in AutoCal (v11.6). The Actor File for the participant was applied to the recorded data after data collection rather than prior to recording due to time constraints. The suit was then calibrated, at the start of the experiment.

4.1.6.2 Tasks

This experiment required the participants to complete two tasks in three orientations. Participants were provided with personal protective equipment (PPE) while they completed the tasks, which included a bump cap to protect their head and safety glasses to protect their eyes. Prior to starting the tasks, the videos, heart rate monitor, and mocap suit were all turned on. Recording for the mocap suit and heart rate monitor were started at the same time at the beginning of the tasks to allow alignment with the video recordings. Participants were given two minutes to make as much progress as they could on either task, in the three different postures (OH stood, OH sat, and neutral). The task was kept to two minutes due to the number of conditions completed within this experiment and to mitigate against physical fatigue from impacting performance in the later conditions. Each task was completed in all orientations before moving onto the second task, these orientations can be seen in Figure 4-6 and Figure 4-7; Figure 4-6 provides an example of a participant working in the OH sat condition while completing the Duct Installation task. The reason for completing one task in all orientations before progressing to the second task was to prevent confusion, particularly for the higher complexity task and to help ensure the participants remembered all of the task steps.

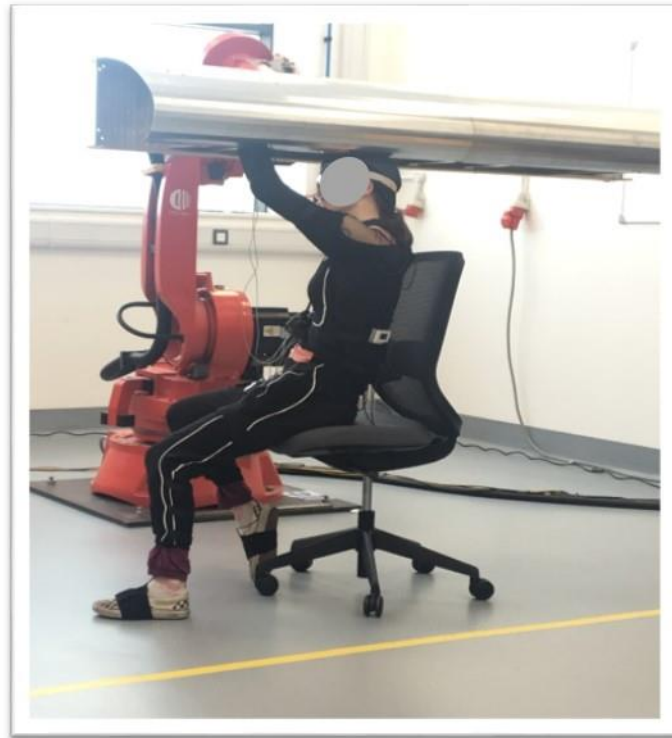


Figure 4-6: Participant Working In the OH Sat Condition While Completing the Duct Installation Task

As detailed in Section 4.1.3, the Bolt Sealant and Duct Installation tasks and physical strain were counterbalanced. For both tasks, prior to being asked to complete the activity participants were shown the task steps and provided with an explanation with the mock-up in the neutral orientation (Figure 4-7), the details of the two tasks are found in the following two subheadings 4.1.6.2.1 and 4.1.6.2.2 below. Participants were invited to ask any questions and all were answered. Once all questions had been answered, the robot was moved into the required orientation, all recordings were started, the participants were informed of the time limit and asked to begin the task.



Figure 4-6: OH Stood and Sat Working Conditions



Figure 4-7: Neutral Working Condition

4.1.6.2.1 Bolt Sealing

The Bolt Sealant task involved the participants filling in as many of the circles seen in Figure 4-8 as they could in the two minutes. Prior to starting the task the participants were shown the first circle completely filled and asked to ensure that

the circles they filled in were to the same standard. This was to ensure there was no confusion regarding what was expected of the participants. The participants were asked to work in rows horizontally, working from the 'x', and told they had two minutes to fill in as many circles as they could. Once shown the filled circle it was cleaned off, the participant was given a whiteboard marker, and asked to start the task. The task was completed in all three orientations.

4.1.6.2.2 Duct Installation

The parts installed during the Duct Installation task were laid out in the order of use, as can be seen in Figure 4-9. This was to help the participants remember the task steps. Figure 4-10 shows the task steps taken to complete the Duct Installation task. The Duct Installation involved the participants first collecting the blast tube and three bolts, they fit the blast tube into the component, ensuring that both back and front were flush with the ribs and the bolt holes aligned before they secured it in place using the three bolts and an Allen key as seen in Figure 4-10a. The participants were asked to tighten the bolts until they were flush but not overly tight, to ease the task reset process.

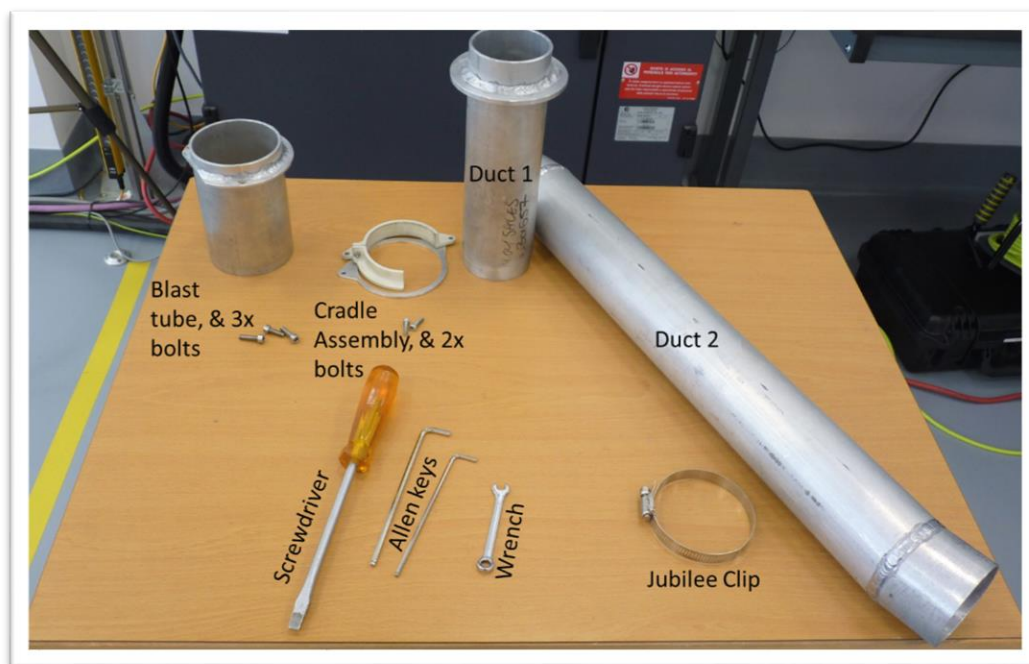


Figure 4-9: Layout of Components for Duct Installation Tasks

The participants then installed the cradle assembly by aligning the cradle (plastic component) over the backing plate (metal component) and then bolting it on to the rib using two bolts and an Allen key as seen in Figure 4-10b. Duct 1 was then inserted through the cradle and blast pipe, and the second plastic section of the cradle assembly was tightened to hold the duct in place (Figure 4-10c). The cradle assembly was then tightened using the wrench and Allen key.

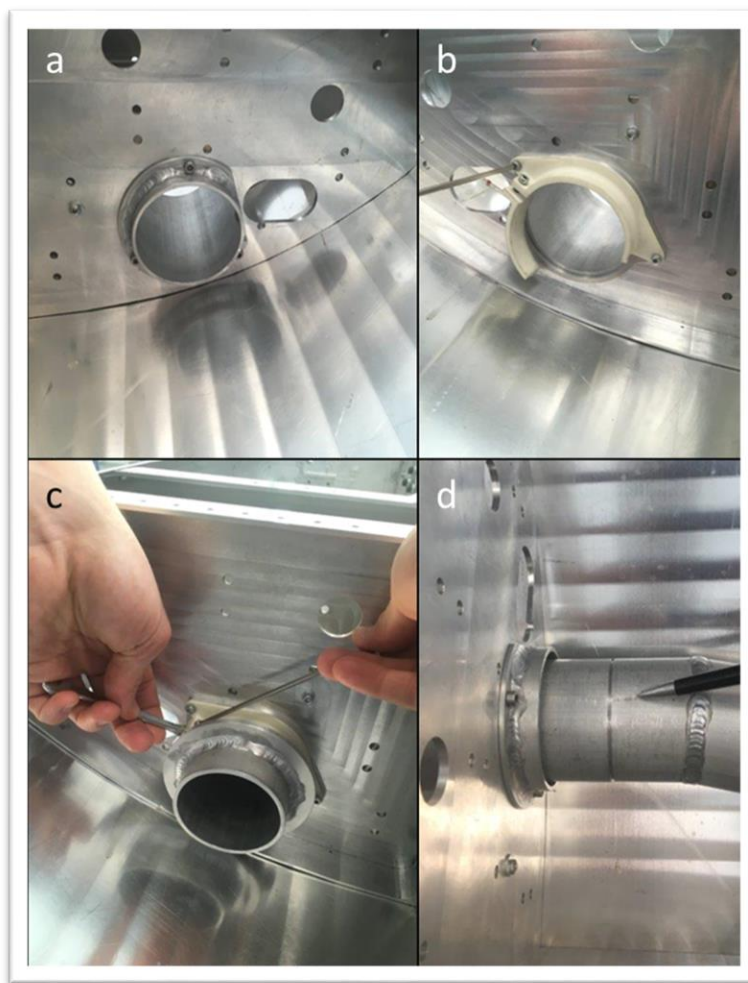


Figure 4-10: Duct Installation Task Steps: a) Blast Tube Inserted and Secured with Three Bolts, b) Installed Cradle Assembly, c) Installed Duct 1 and Cradle Assembly Tightening, d) Duct 1 and 2 Alignment

The jubilee clip was placed over Duct 2, and Duct 2 was inserted into a pipe clamp attached to the rib and the other end of Duct 2 was aligned with and inserted into Duct 1. A line on Duct 1 and Duct 2, as seen in Figure 4-10d, had to be aligned to ensure the correct fit between the two ducts. The jubilee clip was then moved into place over the join of Duct 1 and 2 and tightened using the screwdriver as shown in Figure 4-11.



Figure 4-11: Jubilee Clip Placement and Tightening

After spending two minutes on each task the participants were asked to stop, and all recordings were ended. The participants then completed the post-trial NASA-TLX (data collection sheets can be found in Appendix 6.3D.1), while the experiment was reset for the next task and the mock-up moved into its required position using the robot. Completing the scale required each participant to mark the most suitable point on each of the six workload dimension scales. Each dimension has a scale with 20 intervals, with 21 vertical tick marks dividing the scale from 0 to 100 in increments of 5. After completing the final task and NASA-

TLX the participant removed the mocap suit, was debriefed and reminded of all ethical entitlements.

4.1.7 Analysis

4.1.7.1 Postural Analysis

The postures adopted by participants in each of the conditions required analysis, to identify the level of physical strain experienced by participants in each condition. This produced a numerical value corresponding to the level of strain experienced during the task, which could be used to identify whether a statistically significant equation could be established. The whole task was observed from beginning to end, for all participants across all tasks, to identify the posture with the highest REBA score. To do this, the videos and mocap recordings were reviewed at the same time by aligning the two recordings and running them concurrently. By viewing both types of recording at the same time the video recordings provided data where the mocap suit had not worked effectively and where it had worked the video recordings provided contextual information, such as, participants leaning on the mock-up which lowers the REBA score. Mocap data was not available for all of the tasks completed by all participants due to the mocap suit malfunctioning during data collection (effective for 172 recordings out of 294), so in those cases video recordings were solely relied upon.

The highest REBA rated posture was identified with reference to the REBA scoring system and a search for postures where body segments would score highly. The reason for this is that the part of the task with the highest posture score is likely to lead to the development of MSDs, and, therefore, it is of interest for assessment and redesign. Details regarding REBA scoring are provided in the following paragraphs. Once the highest posture had been identified the timing for the two videos and a key frame from the mocap recording were identified and logged. The postures were then analysed using the systematic REBA scoring method (McAtamney & Corlett, 1993). REBA was used because in a comparison with alternative available methods it was the only method able to capture very awkward postures which include squatting, lumbar flexion greater than 90° and torso twisting (Chiasson et al., 2012). These types of postures were required with

the OH standing condition. A single frame was analysed rather than the whole task, because a method for analysing the task as a whole could not be found, and these methods are used for physical strain analysis throughout the literature (Chiasson et al., 2012; Nino et al., 2019; Pillastrini et al., 2007).

The analysis produced individual scores for six body segments – neck, trunk, legs, upper arm, lower arm and wrist, with supplementary weighting factors such as loads (weights requiring lifting), couplings (how well designed handles are), duration (time taken on the task), and dynamics (repetitive or changes in range of motion) additionally scored. A worksheet which provides visual diagrams of each of the following steps can be found in Appendix 6.3C.1. The first step was to identify the neck, trunk and leg positions. The neck position and its associated score could be between 1 if the neck was flexed between 0° and 20°, and 2 if the flexion was greater than 20° or if the neck was in extension, with any twist or bending in the neck providing an additional point to the neck score. The same was then completed for the trunk with a similar score for bend severity ranging from 1 to 4 and a +1 adjustment added for twisting or sideways bending. The legs were then scored, if weight was dispersed equally across legs, including walking or sitting, then the value given was 1. If weight was unilaterally distributed across the legs, feather weight bearing or an unstable posture then the legs were scored with a 2. The bend in the legs was then considered with a flexion in the knees between 30° and 60° leading to an additional point, and greater than 60° flexion leading to the addition of 2 points to the leg score. The scores were then inputted into their table (Table A) as found in McAtamney and Corlett, (1993) the addition of a force score was not required for the tasks within this experiment.

The second step was to identify the positions of both arms: the left and the right arms were scored separately and the arm with the overall highest score on Table B was used in the overall scoring. The first step in scoring the arms was to locate the upper arm position. The further the arm was from the midaxillary line of the trunk the higher the score ranging from 1 to 4. If the upper arm fell between 20° in extension or flexion a single score was given, and if the upper arm was in extension greater than 20° or flexion between 20° and 45° 2 points were given.

An upper arm flexion between 45° and 90° was given 3 points, and a flexion greater than 90° was scored 4 points. If the arm was supported the score was reduced by 1, if the shoulder was raised, and or the upper arm was abducted an additional point was given for each.

The lower arm position was scored either 1 if it was between 60°-100°, or 2 if it was outside these values. If the position of the wrist fell between 15° for extension or flexion it was scored 1, and a score of 2 was given if flexion or extension was greater than 15°. A +1 adjustment would be given if the wrist was bent from the midline, or twisted. These scores were then looked up on their associated table (Table B) and the participant's ability to grasp the tools, or white board marker were assessed with a coupling score. The coupling score was identified for all tools and the marker, as is defined in the REBA scoring, as "Acceptable but not ideal hand hold" which provides an additional point to the Table B score.

The Table A and B scores with additional points were then entered into a final table (Table C) where an overall score was given. Activity scores were not added to the overall score as the participants did not hold body parts in position for longer than a minute, repeat small range actions, or complete actions that caused rapid large range changes in posture (McAtamney & Corlett, 1993). Where mocap data was missing and the participants were obscured, due to working within the mock-up, a score of 1 was attributed to the body segment that could not been seen.

To ensure the reliability of the postural analysis scores identified by the researcher a second rater scored 20% of the postures. This second rater was a researcher who had been fully briefed in the use of the REBA method. 20% of the postures assessed equates to 62 postures, to identify these a random number generator was used to ensure a spread across the data set. Where duplicate postures arose a new number was generated. As a random number generator was used, this meant that in some instances the spread of identified postures was not equal across all conditions. The spread was reviewed and changes were made to ensure an equal distribution across conditions. The second rater scored all 62 postures.

The overall scores for both the first rater and the second rater were assessed using a weighted Kappa which takes the degree of discrepancy between the two raters into account. Where an unweighted Kappa identifies agreement or disagreement, a weighted Kappa takes into account the level of mismatch. The reason for using a weighted Kappa rather than an unweighted one was because some of the scores, particularly for the arms, often fell on the border of two scores. As a result, there would be a disagreement between the two raters, an unweighted Kappa identifies this as a disagreement. Whereas, a weighted Kappa takes this into account and adjusts the scores accordingly, a greater level of disagreement decreases the Kappa score to a greater degree than a smaller level of disagreement. The weighted Kappa score was 0.724, which indicated a substantial level of agreement (Landis & Koch, 1977) between rater one and two. Therefore, it was accepted that the remaining postures were reliably scored by rater one.

The mean, SD, minimum and maximum of the overall postural analysis scores across all participants were then identified. Hignett and McAtamney (2000) provide a categorisation of the scoring with the REBA method, which indicates the level of severity of the postures as identified in Table C and the associated risk of developing MSDs. The categorisation of the Table C scores range from negligible risk to very high risk depending on score. The risk levels and scores are:

- 1 negligible risk
- 2 - 3 low risk,
- 4 - 7 medium risk,
- 8 - 10 high risk,
- >11 very high risk.

These ranges provide contextual information for the severity of the postures adopted across the conditions. This enables the identification of the level of physical strain the participant is exposed to in the different conditions.

4.1.7.2 NASA-TLX

The completed NASA-TLX sheets were individually scored. For each data collection sheet each of the six dimension scales range from 0 to 100. This information is used to identify the score for each dimension of the NASA-TLX completed by the participant. Overall Workload for each participant was identified by averaging the scores of all six scales. Averages, SD, minimum and maximum values of Overall Workload were then calculated for the responses from all participants across all variables. The average, SD, minimum and maximum values were additionally provided for the scale Mental Demand as a measure of mental workload across all participants.

4.1.7.3 Heart Rate Variability

All HR data had to be processed before it could be analysed, because the software is not always able to identify HR peaks and in some cases artefacts (errors) occur within the data due to movement which causes HR peak miss labelling. To process the data HR peaks for each recording had to be visually reviewed within the HRV Analysis (version 3.2.5) software program provided by MindWare to correct any artefacts. If it was not possible to visually identify a peak and correct an artefact, a mid-beat was applied using the software. This is where the software places a peak of best fit with reference to the peak and interval data from the “good” data on either side of the artefact. The number of estimated mid-beats did not exceed the recommended 10% of the data set, as outlined by MindWare. If the recommended 10% is exceeded then the HRV data may not be reliable. Once a recording was free of artefacts the mean IBI was extracted from the software into an excel file. Mean IBI data was compiled into a single data set for analysis. Averages, SD, minimum and maximum values were calculated for the responses from all participants across all variables.

4.1.7.4 Generalised Estimating Equations

The aim for this experiment was to identify the predictive effect of physical strain on mental workload, therefore, a linear model was required. For an independent measures experiment a regression analysis would be the most suitable method for providing this equation. However, since a repeated measures experimental

design was used, generalised estimating equations (GEE) were used to identify whether linear models could be identified from the data sets. This method of analysis provides information regarding the statistical significance of the relationship between the two variables, and provides a linear equation. GEEs were applied to:

- the REBA scores and the IBI data,
- the REBA scores and the overall NASA-TLX scores,
- the REBA scores and the Mental Demand NASA-TLX scale scores

The statistically significant equations provided by the GEE analysis were used to provide predicted workload scores based on the REBA scores. These predicted and original workload scores were then compared using R-squared, which is a method of assessing how well the equations were able to predict the workload data. The analysis provides an assessment of the goodness of fit of the equation, and the percentage to which the REBA scores explained the change in the workload measures.

Multiple sets of analysis were completed, which can lead to an increase in the risk of Type I error. Type I error is the risk of the analysis providing statistically significant results, when they are not. In order to control for the increased chance of Type I error induced by completing several tests, a Bonferroni correction was applied to the results, which lowers the significance threshold from .05 to .017.

4.2 Results

4.2.1 Descriptive Statistics

Table 4-1 provides the minimum and maximum values for each condition value across all six factors in addition to the averages and SDs. Lower IBI levels are found for the more complex task than the low complexity task. Therefore, suggesting that the differing task complexities affected the HRV statistic IBI and indicates a possible effect on mental workload. Additionally, within the two tasks, IBI levels are lower when the orientation of the task is more physically demanding.

As explained in Section 4.1.7.1 the REBA method used to assess the severity of the postures adopted by participants provides a risk index which provides context to the posture results. The mean physical risk seen in the OH sat and Neutral conditions for both tasks fell in the medium risk category, whereas they fell in the high-risk category for the OH stood conditions. As can be seen from the maximum data, the highest postural scores were observed in the OH stood orientation, as a result, these scores fall within the very high risk of developing MSDs category as identified by Hignett and McAtamney (2000). The minimum score for the OH stood and Neutral Duct Installation conditions fell within the medium risk of developing MSDs category. The minimum postures for the OH sat orientation in both tasks, and the Neutral Bolt Sealant condition fell within the low risk category. The maximum postural scores for OH sat and Neutral orientations for both tasks fell within the high risk category. These results indicate that the level of physical strain was effectively manipulated between the conditions.

The NASA-TLX scores in Table 4-1 show that within task conditions Overall NASA-TLX scores increased as the component orientation increased in severity. When comparing overall NASA-TLX scores, between task conditions for the different orientations, higher scores can be seen for the more complex Duct Installation task than for the Bolt Sealant task. The Mental Demand NASA-TLX sub-category results replicated this pattern of results. These results indicate that the use of task complexity as a means of manipulating mental demand to assess the impact of physical strain on different levels of cognitive capacity appear to be successful. The GEE's clarify whether these relationships are significant, and therefore, how effective the manipulation of both physical and cognitive strain was.

Table 4-1: Minimum, Maximum, Mean and SD for Mental Workload and Postural Analysis

		Duct Installation			Bolt Sealant		
		OH stood	OH sat	Neutral	OH stood	OH sat	Neutral
Postural Analysis	Min	4	3	4	8	3	3
	Max	11	9	9	12	9	9
	Mean	8.548	6.833	6.929	10.024	7.143	5.595
	SD	1.797	1.447	1.314	0.715	1.280	1.768
NASA-TLX	Min	28.333	26.667	23.333	23.333	12.5	5
	Max	86.667	80.833	78.333	93.333	75	58.333
	Mean	61.25	53.056	47.679	59.187	46.091	30.444
	SD	15.918	14.911	15.013	15.799	15.439	12.084
Mental Demand	Min	10	10	5	5	5	5
	Max	85	85	80	95	60	55
	Mean	44.881	38.929	36.190	32.262	23.929	17.143
	SD	23.045	20.138	20.146	23.405	15.521	12.838
IBI	Min	1.797	457.517	458.995	408.437	468.180	469.879
	Max	819.834	872.059	915.006	914.877	975.174	1098.679
	Mean	593.062	661.155	682.115	596.775	728.012	756.692
	SD	96.274	101.407	116.709	112.092	125.126	148.547

4.2.2 Model Development

GEEs were used to test whether a relationship could be established between REBA scores and mental workload (NASA-TLX scale and IBI). A GEE identified that it is possible to predict NASA-TLX workload scores based on REBA scores. The statistically significant results were: $X^2(1,42) = 61.558$, $p < .001$, $r^2 = .109$. The analysis provided the following equation:

$$(\text{REBA score}) \times 3.244 + 25.251 = \text{NASA} - \text{TLX overall score} \quad (4-1)$$

After using Equation 4-1 to identify the predicted NASA-TLX overall score, the predicted and original results provided by participants were compared using R-squared. The effect size revealed that REBA scores predicted 10.9% of the variance in overall Mental Workload scores, which indicates that other factors not accounted for within this study explained the remaining 89.1% of the variation.

The NASA-TLX subscale Mental Demand which measures mental workload, was individually analysed using a GEE to establish whether a causal relationship existed. A significant effect of REBA score on Mental Demand was identified with $X^2(1,42) = 18.813$, $p = .004$. An R-squared was computed to compare the observed and predicted score values of Mental Demand indicating that REBA scores predicted 1.8% of the variance in the Mental Demand scores.

The predictive equation is therefore:

$$(\text{REBA score}) \times 1.699 + 19.459 = \text{NASA} - \text{TLX Mental Demand} \quad (4-2)$$

A significant effect of REBA scores was found for the HRV statistic IBI (Mental Workload): $X^2(1,42) = 749.391$, $p < .001$. R-squared identified that REBA scores predicted 14.8% of the variance in the IBI scores. The prediction equation is:

$$(\text{REBA score})x - 22.038 + 833.789 = \text{HRV IBI} \quad (4-3)$$

These results indicate that it is possible to identify an equation that predicted the effect of REBA scores on workload and specifically mental workload. The goodness of fit results (R-squared), however, show that there are other factors that affect these equations and additional research is required.

4.3 Discussion

The aim of this experiment was to establish a predictive model or equation of mental workload. This was achieved through identifying a predictive effect of REBA scores on mental workload, which could be integrated into a DHM CAD package. A repeated measures methodology was used to assess changes in objective and subjective workload and mental workload. Analysis of the results revealed statistically significant relationships between physical strain and overall NASA-TLX scores, the NASA-TLX subscale Mental Demand scores, and the HRV statistic IBI. The identification of these relationships support other accounts of physical strain affecting mental workload (Mehta & Agnew, 2011; Nino et al., 2019; Perry et al., 2008) and particularly on the NASA-TLX subscale Mental Demand as a measure of mental workload (Mehta & Agnew, 2011) as detailed in Chapter 2 Section 2.2.1. The reason for this as explained in the literature is due to both utilising the same underlying energy resource (Kahneman, 1973; Norman & Bobrow, 1975; Van Acker et al., 2018). More importantly though, the identification of predictive equations indicates that it is possible to identify a predictive effect of physical strain on both objective and subjective forms of mental workload, and the provision of equations by the analyses means that the relationships can more easily be integrated into CAD programs.

The identification of these equations is first step towards providing design engineers with a tool to assess a cognitive process within CAD programs prior to the development of new systems or the redesign of existing systems. This information will be a critical tool for proactive ergonomics, where designers are

designing a workstation or a structure/ component that an operator will be required to work on or within this equation provides them with the ability to identify how working posture will impact workload capacity. This then provides designers to identify whether redesign is possible, or the information can be fed through to the organisation and used to inform job design, rotation, and break scheduling. The integration of this type of information into adaptive automation, or digitally based task design and monitoring programs would greatly improve the understanding of work on operators and provide management with the tools to make informed activity and task decisions. It is expected that this would improve performance factors within the workplace, and reduce the associated costs as instances of errors reduce because operators are able to work at optimal workload levels.

The Mental Demand subscale of the NASA-TLX was assessed in addition to the Overall score because although the NASA-TLX has been shown to be sensitive to changes in mental demands, the overall scale is influenced by physical demands and the other constructs it measures (Mehta & Agnew, 2011). Therefore, as identified by DiDomenico and Nussbaum (2008) the overall score is not necessarily indicative of mental workload levels. These subscales have been shown to be sensitive subjective assessments of the various types of workload and that individuals are able to effectively distinguish and evaluate between physical and mental workload during tasks involving concurrent mental and physical demands (DiDomenico & Nussbaum, 2008). The goodness of fit for the GEE equations for the overall and Mental Demand NASA-TLX scales further reflect this, as the REBA scores explain a smaller percentage of the variation observed in Mental Demand as compared to the overall NASA-TLX scores. Therefore, researchers measuring mental workload need to present the relevant NASA-TLX subscale to participants if their intention is to assess only mental workload.

Although the statistically significant GEE for the effect of REBA scores on IBI provides an objective measure of mental workload, it is not possible to distinguish between the effect of physical exertion from mental workload in the HRV data

(Bucks et al., 1994; Wilson, 1992). Therefore, although a statistically significant predictive equation has been identified for IBI, physical strain is likely to have affected these results and, therefore, the IBI data does not necessarily represent mental workload. The repeated measures experimental design and the use of consistent physical strain was intended to reduce the effect of this confounding variable. However, this does not wholly mitigate the effect on the IBI equation and, consequently, some effect likely remains. To understand the extent of this effect on the equation additional research is required, in which the effect is identified at varying levels of physical strain and is isolated from the mental workload data.

Generalising the equations found in this experiment beyond this experiment should be done cautiously. It is possible that extraneous factors may have had an effect on mental workload due to the laboratory design. Although the tasks completed by participants reflect those currently completed within aerospace manufacturing, the laboratory environment controls for the additional factors that may affect an operator's mental workload. Such as noise levels and organisational pressures. A factor that may have influenced the effectiveness of the equations is the participant's proximity to an industrial robot. The mock-up was held and moved using an industrial medium payload robot. Although it did not move during the conditions it had to be moved between them and, due to the laboratory layout as seen in Figure 4-1, this meant that the participants were approximately three meters from the moving robot. This may have led to an unforeseen effect of emotional stress on the participants, which HRV is sensitive to (Wilson, 1992). Therefore, the IBI equation may account for stress in addition to mental workload, which would affect its predictive accuracy. To account for this effect further research should investigate the impact of stress caused by proximity to and movement of a robot. Despite these factors, the equations provide evidence that it is possible to predict the effect of physical strain on mental workload and additional research should be conducted to further investigate this interaction.

4.3.1 Study Limitations

Maintaining physical demand for the Duct Installation task was difficult due to the mock-up. The physical demand of the task reduced as the participant progressed to attaching the cradle to the rib, because participants were no longer required to bend to fit the part. Within the manufacturing environment where this task is completed this would not be the case. However, only a subsection of the component could be used due to space limitations. This was not the case for the low complexity Bolt Sealant task. Therefore, it is likely that this influenced the results and the accuracy of the equations, potentially lowering the effect of physical strain on mental workload for the high complexity task. However, variations in postures like this are representative of industrial work (Mehta & Agnew, 2011), and consequently, this may have provided some ecological validity to these results.

An additional factor that might have affected physical strain for participants is variation in their height. It may explain the low postural severity observed in the minimum postural score for the OH stood working conditions. This is likely to mean that the equations do not effectively account for low REBA scores. Although controlling for these variations would have been preferred, the experimental cohort would not have reflected the general population and, additionally, it may have hindered access to a suitable number of participants. It is advised that future research control for these variations. A possible solution could have been to vary the height of the component, however, the intention was to replicate the current manufacturing tasks as much as possible and, therefore, the component was set to the height that it would be within the Airbus A320 flow-line.

A further potential confounding variable that should be considered is increased emotional stress due to feeling self-conscious while wearing the mocap suit. The mocap suit is made of Lycra and is, therefore, skin tight. This feeling of discomfort may have been compounded by being video recorded and watched by the experimenter and laboratory technician while they completed their task. Additionally, participants may also have felt that their performance was being judged, which could have further contributed to emotional stress. Both factors

may have led to an increase in emotional discomfort and consequently, may have affected the HRV data and possibly contributed to the adoption of unnatural postures which would have potentially affected the REBA scores. Unfortunately, for this experiment both forms of mocap were required due to the limitations of the mocap suit.

Due to technical difficulties and possibly the age of the mocap suit, it was only effective for 58.5% of the data collection sessions, which meant that postural analysis was conducted using the video recording where it had not been possible to record mocap data. A consequence of this is that some body segments were obscured and, therefore, a lower REBA score was calculated, because only a single point could be attributed to a body segment that could not be seen. It is possible that this has implications for the equations identified within this experiment. These difficulties further reinforce the findings from Chapter 3, that the technology is not stable and effective enough to be used independently from other forms of mocap with confidence. Future experiments should look to either use a more reliable mocap suit or just use the video cameras. However, if video cameras are used the occlusion problem remains. The HRV monitor was much more effective than the mocap suit and there were only five instances of missing data resulting from problems with sensor attachment or operator error, difficulties with sensor attachment are well known issues with HR monitors (Wilson, 1992).

4.3.2 Conclusion

Despite limitations of generalisability with the equations identified from this experiment, it is a first step towards providing design engineers with a tool to assess a cognitive process within CAD programs prior to the development of new systems. The equations should enable an assessment of the impact that physical components of a task may have on mental workload, and provide the ability to control for the factors that may excessively increase mental workload required for tasks. As a result, this should help reduce negative performance effects resulting from high levels of mental workload and reduce the need for redesign of tasks or systems. Further research should be conducted to validate and investigate the reliability of these findings, and should look to develop and fine-tune other

equations like this for integration into DHM capabilities. Due to the complexity of mental workload and the factors that may affect it, it was not possible to integrate these equations into a DHM CAD program for validation purposes. A more easily controlled equation required identification, and this was achieved in the following Chapter.

5 STUDY: DEVELOPMENT AND INTEGRATION OF A HUMAN ROBOT COLLABORATION EQUATION INTO DHM

As detailed in the Literature Review (Chapter 2), one of the ways that manufacturing companies are looking to reduce the physical impact of tasks on their operators, where it is not possible to redesign the task or environment, is to introduce HRC. This involves operators working in close quarters and collaboratively with a robot. The psychological impact of introducing this type of close work requires consideration prior to implementation (Grahm et al., 2018; Koppenborg et al., 2017; Parasuraman & Riley, 1997; Parasuraman et al., 2008; Vidulich & Tsang, 2012), and arose as an area of interest in the experiment detailed in the previous Chapter (4) due to the potential impact it had on operator's workload. This Chapter aimed to assess whether the three key psychological constructs expected to affect HRC as identified in the literature review (Section 2.2) (mental workload, trust, and mood) could be used to establish equations for integration into DHM CAD programs. Only the trust data could be used to establish an equation and, therefore, this data was then used to address the secondary aim of this Chapter, which was to assess the reliability of the equation once integrated into CAD. By achieving this, Objective 3: 'to establish human robot interaction equations/s for industrial HRC that can be integrated into DHM CAD programs', and 4: 'to assess the predictive accuracy of the HRC equation/s after integration into a DHM CAD program', were met. This Chapter is comprised of two parts, the first half details the identification and establishment of the equation/s, and the second half details the reliability process.

The work detailed within this Chapter was part of a Connected Everything funded project and was completed in collaboration with Airbus, and the Advanced Manufacturing Research Centre (AMRC). Airbus provided the use case selection for the reliability experiment, and the AMRC provided their programming expertise to create the equation and to integrate it into a CAD program. The author of this thesis conducted the experiments to established and to assess the reliability of the equation.

5.1 Development of a Human Robot Collaboration Equation

5.1.1 Method

5.1.1.1 Research Ethics

This research was approved by CURES (see Appendix 6.3E.3) and conducted in accordance with the Cranfield Research Integrity Policy, the BPS's Code of Human Research Ethics, and the GDPR 2018. Details of adherence to these policies can be found within the procedure detailed in Section 5.1.1.6.

5.1.1.2 Participants

Twenty-two participants volunteered in response to internal Cranfield University advertisements uploaded on the announcement board on the University's intranet and attached to physical announcement boards around campus. There were ten female and twelve male staff and student participants, between 22 and 62 years (mean: 32, SD: 13.24).

5.1.1.3 Design

A repeated measures experimental design was used to enable measurement of the DVs (Trust, Mood and Workload) across different conditions involving changes to the IV (robot speed). The experiment was conducted in the IA laboratory at Cranfield University as this allowed control over environmental variables and privacy.

5.1.1.3.1 Variable Selection

Robot speed was chosen as the IV, because it has been shown to have an effect on individual's perceived trust during HRC tasks (Arai et al., 2010; Charalambous et al., 2016; Koppenborg et al., 2017). Furthermore, precise and repeatable changes can be achieved with speed. It was intended that the ability to control the IV in this way would reduce the possible impact of extraneous variables and

would overcome some of the complexities observed in the developed equations in Chapter 4.

Trust, workload and mood were the DVs as interacting with a collaborative robot was expected to have an effect on these psychological responses as detailed in the Literature Review in Section 2.2.2. Additionally, their assessment methods produce numerical outputs which can be used to create equations that are suitable for integration into CAD programs. Mood was only measured with a proportion of the participants (n=6), because a subject advisor recommended its inclusion after data collection had already started.

5.1.1.4 Experimental Task

To ensure that effects of the IVs were captured without any confounding interference due to effects of task complexity, a simple task was required. A simple pipe-fitting task, which was used in the development of the Trust in Industrial HRC scale was used. This involved the following steps:

1. the robot (initiated by technician) picked up a pipe from its left side and brought it to the participant and then stopped, presenting one end for assembly
2. the participant then attached four fittings onto the end of the pipe that had been presented to them
3. the robot (initiated by technician) then moved away again with the assembled pipe and placed it in a box before returning to its start point

The robot movement was initiated by the technician rather than having the robot run automatically, because the robot is not a fully collaborative robot and could harm the participant if it moved while they were interacting with it.



Figure 5-1: Attachment of the Fittings to the End of a Pipe

5.1.1.5 Materials and Equipment

5.1.1.5.1 Robot Cell

The same Comau NM45 industrial robot used in Chapter 4 was used for this experiment because of its availability in the IA laboratory and, its suitability for a repeated measures study of this kind. Additionally, it was able to safely run at a range of speeds that would enable effective assessment of the effect of speed on the DVs. Figure 5-2 shows the experimental layout, indicating the pipe's start and end positions at either side of the robot, the 'participant marker' where participants were asked to stand when the robot was in 'active mode' and moving, and the 'fittings' which the participants fit onto the end of the pipe.

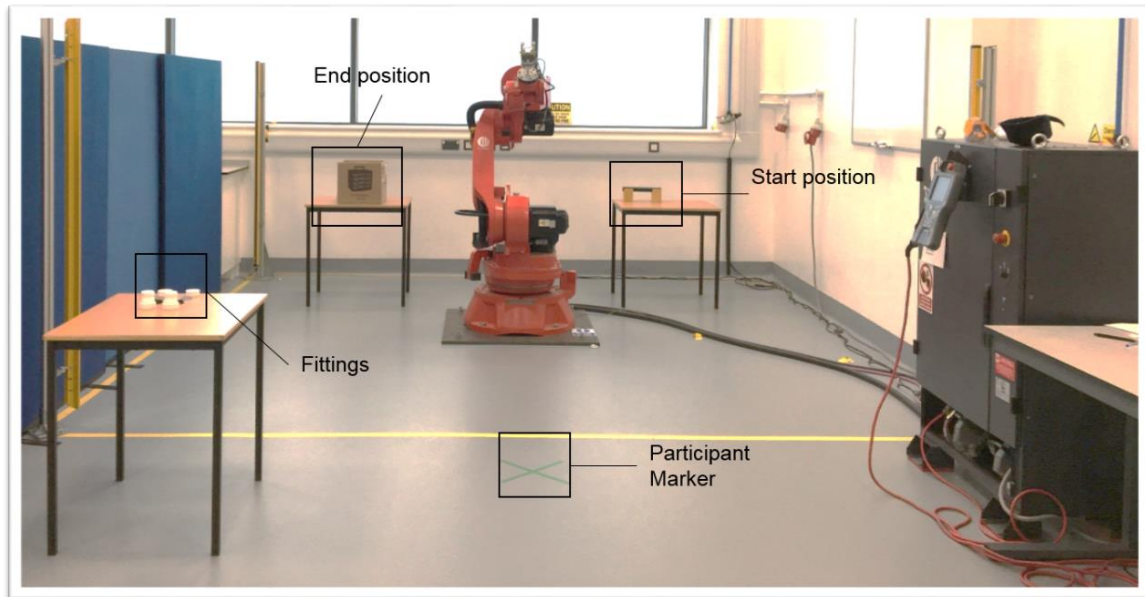


Figure 5-3: Experiment Robot and Layout

As the Comau NM45 industrial robot has no integral collaborative safety functions and the participants would be in proximity to the robot while it was moving, a safety light curtain was installed to ensure protection of the participant during the robot's 'active modes'. The light curtain is the yellow vertical apparatus that can be seen on the left of the image in Figure 5-2, and its perimeter is indicated by the yellow line across the floor. When the robot is in motion and the device is operational it would initiate an immediate robot stop should there be any object or human incursion across the perimeter. The participant marker was used to prevent participants from changing proximity to the robot which could affect their perception of the robot speed.

5.1.1.5.2 Independent Variable Measurement

As robot speed was the IV, the different speeds that required testing needed to fit within a predefined range, which was determined by both regulatory and experimental criteria. At the upper threshold of this range, it was necessary to

comply with current safety standards for HRC (ISO TS 15066) (ISO, 2016) and this dictated that the highest speed at which the robot could be run safely in collaborative mode was at 50% of its overall potential speed (=1250mm/s). At the lower threshold, it was considered that the experiment would be confounded if the robot's speed was so slow that it increased task time and prolonged task performance. Therefore, it was decided that the lowest speed at which the robot could run should be at just 5% of its overall potential speed (=125mm/s).

Having defined the maximum and minimum thresholds of the robot's speed, it was decided that five speeds within this range would be sufficient for comparison: any fewer may not have offered enough discrimination but any more was likely to be unnecessary and may have induced fatigue and monotony in the participants. Table 5-1 below sets out both the exact speed values and what percentage of the robot's overall potential speed this represents across all five IV conditions within the specified range.

Table 5-1: Independent Variable Conditions: Robot Speeds

Condition	Actual Speed mm/s	Potential Speed %
Program 1	125mm/s	5% Speed
Program 2	400mm/s	16% Speed
Program 3	550mm/s	22% Speed
Program 4	850mm/s	34% Speed
Program 5	1250mm/s	50% Speed

5.1.1.5.3 Dependent Variable Measurement

As detailed in Section 2.2.2 of the Literature Review these measures were chosen because of their likelihood of being affected by collaborative robotics, and their ability to establish numeric outputs from their measures of assessment. A

numerical assessment was critical to ensure that an equation could be established for integration into DHM CAD programs.

5.1.1.5.3.1 Trust in Industrial Human-Robot Collaboration

To evaluate the impact of the different robot speeds on Trust, participants completed a paper-based psychometric questionnaire after each trial. The Trust in Industrial HRC Scale (Charalambous et al., 2016) is a questionnaire, which requires participants to subjectively report their perceived levels of trust in a robot system across three robotic dimensions: motion and pick-up speed, safe co-operation, and robot and gripper reliability. This scale was chosen because it measures HRC within a manufacturing context and its items and dimensions provide a numerical output that can be used to create an equation for integration into a DHM CAD program, and they have a diagnostic potential in predicted trust in HRC.

The scale uses a five-point Likert scale ranging from strongly disagree to strongly agree in response to the ten items measuring those dimensions (Table 5-2).

Table 5-2: Trust in Industrial Human Robot Collaboration Scale

Dimensions	Item	Identifying Letter
Robot motion & pick-up speed	The way the robot moved made me uncomfortable	A
	The speed at which the gripper picked up and released the components made me uneasy	C
Safe co-operation	I trusted that the robot was safe to cooperate with	D
	I was comfortable the robot would not hurt me	F
	The size of the robot did not intimidate me	H
	I felt safe interacting with the robot	I
Robot and gripper reliability	I knew the gripper would not drop the components	B
	The robot gripper did not look reliable	E
	The gripper seemed like it could be trusted	G
	I felt I could rely on the robot to do what it was supposed to do	J

5.1.1.5.3.2 Workload

The NASA-TLX (Hart & Staveland, 1988) was used in this study in the same raw form as was described in Section 4.1.5.1, to measure workload. The scale was used because it provides a numerical output that can be utilised to create an equation that can be integrated into a DHM CAD program. Although the literature recommended using multiple measures to assess workload, it was not possible in this study, as physiological measurement tools were unavailable for use at the time of experimentation. Therefore, the NASA-TLX was used on its own.

The NASA-TLX scale is comprised of six scales used to measure the six principal dimensions of workload, detailed below:

- Mental Demand (Very low – Very high)
- Physical Demand (Very low – Very high)
- Temporal Demand (Very low – Very high)
- Performance (Perfect – Failure)
- Effort (Very low – Very high)
- Frustration (Very low – Very high)

The survey uses scales ranging from 0 to 100 in increments of 5. Overall Workload was identified by averaging the scores from the six workload variables.

5.1.1.5.3.3 Mood

Mood was measured using the UMACL questionnaire (Matthews et al., 1990). As detailed in Section 2.2.2.3 of the Literature Review it was chosen because it is a three factor measure of mood that provides data on energetic and tense arousal, and hedonic tone (capacity to experience pleasure). By using a three factor rather than a two factor model the measure is able to provide a more diagnostic investigation of how HRC impacts mood. Additionally, it provides a factor of frustration, which gives insight into a further negative mood that may be associated with interacting with a collaborative robot.

The UMACL consists of 28 mood-related adjectives across the four scales shown below, the full questionnaire and items can be found in Appendix 6.3E.1:

- Energetic arousal: 8 items,
- Tense arousal: 8 items,
- Hedonic tone: 8 items,
- Anger/frustration: 4 items

The survey uses four-point Likert scales ranging from ‘definitely’, ‘slightly’, ‘slightly not’ to ‘definitely not’.

5.1.1.6 Procedure

5.1.1.6.1 Briefing and Consent

Participants took part in the experiment individually. Upon arrival, the participants were given a standardised briefing regarding the nature of the study, ethical aspects and entitlements, and the key components of the procedure. They were then asked to provide their written consent to take part in the experiment and report their age and gender details on a separate form (which were stored separately from the surveys to ensure anonymity). The six final participants that completed the mood survey completed it prior to the experimental tasks to establish a benchmark to compare with the mood responses collected during the experiment.

5.1.1.6.2 Experimental Task

Participants were first provided with PPE (bump cap and safety glasses). Then they were given a demonstration of the task, so that they were prepared for the task sequence and which task steps the robot would perform versus those that they would be required to perform. The demonstration involved a full run-through of the experimental task procedure, as described in Section 5.1.1.4 with the robot running at Program 2 to ensure enough time for the experimenter to explain everything but also to not overwhelm or influence participants with the faster speeds at their first introduction to the robot. Participants were reminded that after each condition, they would need to complete the Trust in Industrial HRC Scale, and the NASA-TLX surveys. The final six participants were also reminded of the UMACL survey.

During the task briefing, three key points were highlighted to the participants:

1. The robot's stopping point in front of the participant for the assembly task was explicitly shown to participants to minimise any confusion during the trials about when and where the fittings should be attached to the pipe.
2. During the development of the Trust in Industrial HRC Scale, the sound of a pneumatic gripper had been found to negatively affect trust in a robot (Charalambous, 2014). Therefore, the normal pneumatic sound of the gripper was pointed out to the participants and explained so that they were prepared and not influenced by the sound when completing the survey.
3. Participants were asked to stand on the participant marker and not cross the yellow line while the robot was in motion, but advised that they could do so when the robot stopped for the task. It was explicitly explained to the participants that crossing the yellow line while the robot was in motion would activate the safety light curtain and the robot would be stopped, which would require a delaying re-start.

After the demonstration and task briefing, the five robot speed trials were conducted, each lasting two minutes. The task time was kept to two minutes to ensure that the experiment did not become too long and to fatigue effects from confounding the results. The order in which the participants were exposed to the robot speeds were counterbalanced to mitigate order effects. After each condition, the participants completed the trust and workload surveys, with the six final participants also completing the mood survey. Then, at the end of the experiment, the participants were fully debriefed. The surveys administered to the participants can be found in Appendix 6.3E.1.

5.1.2 Analysis

5.1.2.1 Trust in Industrial Human-Robot Collaboration

The Trust in Industrial HRC Scale was used to measure participants' trust at the five different programmed speeds. The raw data gathered from the scale was analysed using the analysis guidelines provided by Charalambous et al. (2016).

The statements were identified based on the letter shown in Table 5-2 and were scored based on the scoring scheme detailed in Table 5-3. The score for each dimension was derived by averaging each participant's item scores under their dimension. A Trust score was identified by adding the dimension scores together. The final computed scores were averaged across all participants and are found in Table 5-5.

Table 5-3: Scoring Scheme for Each Trust Statement

Scale points	Scoring scheme for items: B, D, E, F, H, I, J.	Scoring scheme for items: A, C, G.
Strongly agree	5	1
Agree	4	2
Neutral	3	3
Disagree	2	4
Strongly disagree	1	5

5.1.2.2 Workload

Each of the six workload scales were scored 0 to 100, with 20 intervals and 21 vertical tick marks in increments of 5. Overall Workload for each participant was identified by averaging the score for each scale. Averages and SDs of Overall Workload were then calculated for the responses from all participants across all conditions.

5.1.2.3 Mood

The UMACL has both positive and negative items within the first three scales and then only positive items for the Anger/Frustration scale. Each item was scored using the scoring details can be found in Table 5-4 below.

Table 5-4: Scoring Scheme for Mood

Scale points	Scoring scheme: negative items	Scoring scheme: positive items
Definitely	1	4
Slightly	2	3
Slightly Not	3	2
Definitely Not	4	1

The item numbers associated with the negative and positive items for each scale are detailed below:

- Energetic Arousal:
 - Positive: 3, 5, 16, 22
 - Negative: 7, 11, 19, 24
- Tense Arousal:
 - Positive: 6, 9, 10, 17
 - Negative: 4, 13, 15, 21
- Hedonic Tone:
 - Positive: 1, 8, 18, 23
 - Negative: 2, 12, 14, 20
- Anger/Frustration:
 - Positive: 25, 26, 27, 28, 29

The items within a scale were added together to produce an overall score. Averages and SDs for each of the scales were then calculated for the responses from all participants across all conditions.

5.1.3 Results

5.1.3.1 Trust in Industrial Human-Robot Collaboration

Table 5-5 presents the means and SDs for all participants' responses to each dimension of the Trust in Industrial HRC Scale (D1, 2, and 3). A score greater than 25 indicates a 'good' level of Trust in a robot, which should signify good operational effectiveness and efficiency (although very close to 50 might signify dangers of overreliance and complacency) (Charalambous et al., 2016). Therefore, the Trust scores across the five speeds were, overall, in the optimal range. Although the results indicate that the participants trusted the robot at all speeds, they trusted the robot less at the higher speeds and had the greatest overall Trust for Program 2.

Statistical analysis was applied to the data to identify whether scores between the different speeds were significantly different. A Shapiro-Wilk test was first used to identify that the data was normally distributed; this meant a repeated measures Analysis of Variance (ANOVA) could be performed. The ANOVA, with a Greenhouse-Geisser correction, confirmed that the overall mean scores for Trust were significantly different between at least two speeds ($F(2.864, 60.135) = 13.356, p < 0.05$). Therefore, this data was used to create an equation that would be integrated into a CAD program. Post hoc tests using the Bonferroni correction procedure showed mean Trust scores for Program 5 to be significantly different from the mean Trust scores for Programs 1 to 4 ($p < 0.05$). There was no significant change in the mean Trust scores between Programs 1 to 4 ($p > 0.05$). These results are shown in Table 5-5.

Table 5-5: Means and SDs for Trust in Human-Robot Collaboration across Dimensions

	D1: Robots motion and pick-up speed		D2: Safe co-operation					D3: Robot and gripper reliability						
Program	The way the robot moved made me uncomfortable	The speed at which the gripper picked up and released the components made me uneasy	D1	I felt safe interacting with the robot	The size of the robot did not intimidate me	I was comfortable the robot would not hurt me	I trusted that the robot was safe to cooperate with	D2	I felt I could rely on the robot to do what it was supposed to do	I know the gripper would not drop the components	The robot gripper did not look reliable	The gripper seemed like it could be trusted	D3	Trust Score
1	3.7 (1.3)	3.6 (1.3)	7.3 (2.3)	4.5 (0.7)	3.9 (1.2)	4.4 (0.8)	4.5 (0.7)	17.2 (2.8)	4.4 (0.5)	4.5 (0.6)	4.3 (1.0)	4.4 (0.6)	17.6 (1.8)	42.1 (4.1)
2	4.3 (1.1)	4.5 (0.8)	8.8 (1.5)	4.4 (0.5)	3.8 (1.2)	4.2 (0.7)	4.4 (0.5)	16.7 (2.1)	4.4 (0.5)	4.3 (0.6)	4.3 (0.8)	4.3 (0.5)	17.3 (1.7)	42.8 (3.4)
3	3.7 (1.2)	4.2 (0.8)	7.9 (1.6)	4.1 (0.8)	4.2 (0.9)	4.4 (0.7)	4.3 (0.7)	17.0 (2.6)	4.4 (0.6)	4.4 (0.6)	4.2 (1.2)	4.3 (1.0)	17.2 (2.2)	42.1 (5.3)
4	3.5 (1.1)	4.0 (0.9)	7.5 (1.8)	4.0 (0.9)	3.6 (1.2)	4.0 (0.9)	4.1 (0.8)	15.7 (3.4)	4.0 (0.6)	4.4 (0.9)	4.5 (0.5)	4.4 (0.6)	17.2 (1.7)	40.4 (5.7)
5	2.6 (1.1)	3.2 (1.2)	5.8 (2.1)	3.4 (0.7)	3.5 (1.1)	3.5 (0.8)	3.7 (0.7)	14.1 (2.8)	3.5 (0.8)	4.3 (0.7)	4.0 (0.8)	3.9 (0.8)	15.7 (2.3)	35.5 (5.9)

As no significant difference was found between the overall mean Trust scores across Programs 1 to 4, the individual scores required a closer analysis to identify the most suitable speed for use in the reliability test. To gain a clearer understanding of the trend of the data, a graph of the overall trust scores, as seen in Figure 5-3, was reviewed. The graph shows that with Program 4, a downward trend in Trust levels begins, while in Program 3 the horizontal trend of the previous speed continues, with a value higher than Program 1's value. Whilst the primary goal of this project is to examine the feasibility of integrating psychological data into a DHM CAD program, a secondary intention for Airbus was to assess the feasibility of introducing a robot to the shop floor. For this reason, production needs were also considered so, for improved time efficiency, the fastest speed with a high overall trust scores – Program 3 – was chosen as most suitable for the trust speed equation to be taken forward for the design of the demonstrator and reliability testing.

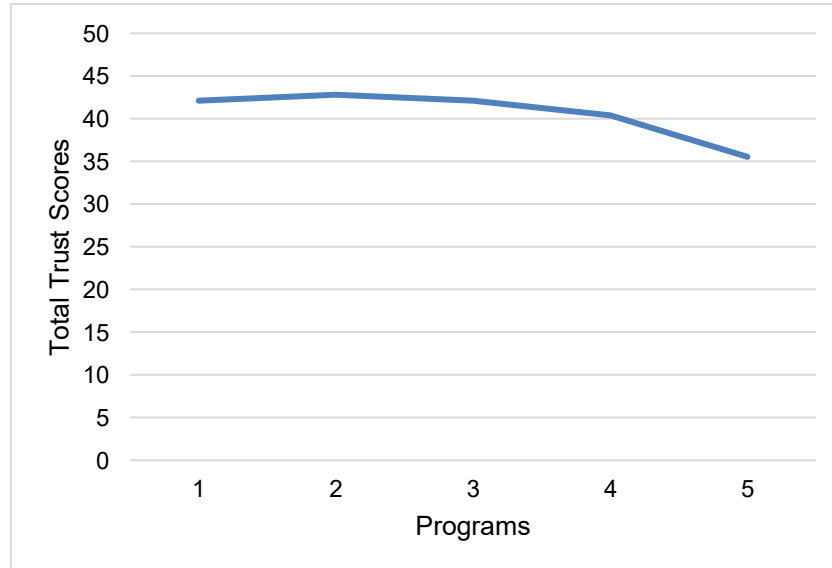


Figure 5-4: Total Trust Scores Across Robot Program Speeds

The repeated measures study led to the identification of a psychological equation between Trust and Speed, whereby the 'ideal' Trust scores (above 25) were achieved when the robot was running at Program 3, which is 550mm/s.

5.1.3.2 Workload

Table 5-6 presents the means and SDs for all participants' responses for the NASA-TLX Overall workload score. The scales range from 0 to 100, and as most scores fell between 20 and just over 30 for each of the Programs they indicate that the task required a low level of workload from the participants. A Shapiro-Wilk test and a review of skew and kurtosis values identified that the data was not normally distributed; this meant a Friedman test was performed on the data. The Friedman confirmed that the overall workload scores were significantly different across the five speeds ($\chi^2(4) = 27.753$, $p < 0.01$). Therefore, these results could be taken forward to develop a robot speed workload equation. However, this was not possible due to the time and resource constraints of the project this work was conducted in.

Table 5-6: NASA-TLX Overall Workload Scores and SDs across Robot Program Speeds

	Workload
Program 1	23.651 (5.636)
Program 2	23.016 (5.698)
Program 3	23.929 (7.327)
Program 4	26.071 (9.506)
Program 5	30.476 (10.909)

Post hoc tests using the Bonferroni correction procedure showed no significant differences between any of the programs ($p > 0.01$).

5.1.3.3 Mood

The UMACL does not produce an overall measure of mood but rather four scales that indicate variations in a participant's mood. Table 5-7 provides the means and SDs for all participants across the pre-task assessment and mood dimensions for each of the programmed speeds. The statistical analysis conducted on each scale and results are described separately below.

Table 5-7: UMACL Means and SDs across Robot Program Speeds

	Energetic Arousal	Tense Arousal	Hedonic Tone	Anger/ Frustration
Pre-task	25 (4.561)	15 (3.899)	27.5 (2.429)	6.333 (1.506)
Program 1	22 (3.347)	15 (3.795)	26.333 (3.266)	7.667 (3.077)
Program 2	23.5 (4.764)	16 (2.191)	27.333 (2.733)	6.333 (2.338)
Program 3	23.667 (3.386)	18.833 (1.941)	25.5 (4.806)	6.833 (2.858)
Program 4	24.333 (2.944)	18.5 (1.975)	26.667 (2.658)	7.167 (2.787)
Program 5	22.667 (1.966)	17.5 (2.258)	25.5 (3.619)	7.5 (2.95)

Prior to identifying whether the Energetic Arousal scores between the different program speeds were significantly different, a Shapiro-Wilk test and a review of skew and kurtosis values revealed that the data was normally distributed which meant that an ANOVA could be applied to the data. The ANOVA showed that the Energetic Arousal scores were not significantly different across the pre-task assessment and five programmed speeds ($F(5, 25) = 1.415, p = 0.253$).

A Shapiro-Wilk test and a review of skew and kurtosis values identified that the Tense Arousal Scores between the different robot speeds were not all normally

distributed. Therefore, a Friedman test was performed and found that the Tense Arousal scores were not significantly different across the five speeds and pre-task assessment ($\chi^2(5) = 9.949, p = 0.08$).

The Hedonic Tone scores were found to be normally distributed using a Shapiro-Wilk test and through reviewing the skew and kurtosis values. Therefore, an ANOVA was applied to the data. The ANOVA, with a Greenhouse-Geisser correction, found that the Hedonic Tone scores were not significantly different across the five speeds and pre-task assessment ($F(1.395, 6.977) = 1.504, p = 0.274$).

The Anger/Frustration scale scores were not all normally distributed as identified through a Shapiro-Wilk test and from reviewing the skew and kurtosis values. Consequently, a Friedman test was applied to the scores and found that the Anger/Frustration score was not significantly different across the five speeds and pre-task assessment ($\chi^2(5) = 6.373, p = 0.272$).

5.1.4 Conclusion

As can be seen from the results, significant findings were found for the Trust and workload responses. Due to the time and resource constraints within the project, it was only possible to take one psychological concept forward into the second phase of the study to develop an equation that could be integrated into CAD and its predictive accuracy tested. Trust was chosen due to the significant differences found between trust at different speeds and enabled an identification of an optimal speed to be used in the reliability testing completed in the following Section 5.2. Future work will look to develop an equation for the workload robot speed responses and test the reliability of the results.

5.2 Assessment of the Accuracy of the CAD Integrated Human Robot Collaboration Trust Speed Equation

This second part of the Chapter provides the reliability assessment of the trust speed equation established in the first part of the Chapter (5.1), and by doing so meets Objective 4 of this thesis: 'to assess the predictive accuracy of the HRC equation after integration into a DHM CAD program'. This part of the study was designed to measure how accurately the CAD model developed by the AMRC with the new integrated trust speed equation, established in Section 5.1, predicts human responses. This assessment of the reliability of the equation was conducted using a demonstrator developed to represent an Airbus installation process down-selected for automation redesign. The reason for choosing an Airbus manual process as the experimental task, is because it was considered an ideal use case due to the potential advantages of introducing large-scale industrial robots to aerospace assembly. As detailed in the Literature Review, by introducing HRC into this industry it is expected that it will be possible to improve both worker health and task performance in tasks that involve large / heavy components.

5.2.1 Method

5.2.1.1 Research Ethics

This research was approved by CURES (see Appendix 6.3E.3) and conducted in accordance with the Cranfield Research Integrity Policy, the BPS's Code of Human Research Ethics, and the GDPR 2018. Details of adherence to these policies can be found within the procedure detailed in Section 5.2.1.6.

5.2.1.2 Participants

Thirty-seven participants volunteered in response to an internal Cranfield University advertisement uploaded on the announcement board on the University's intranet and attached to physical announcement boards around

campus. There were fifteen female and twenty-two male participants aged between 21 and 59 years old (mean: 30, SD: 8.17).

5.2.1.3 Design and Key Variables

As the preceding work had led to specification of the most suitable robot speed (Section 5.1) and predicted an overall trust score per person at a single speed of 550mm/s (Section 5.1.3.1), it was not deemed necessary to replicate the repeated measures design of the previous participant studies. The reason for this was that the aim of this section was to assess the reliability of the optimal score identified in the previous Section (5.1.3.1), and the predictive score identified by the AMRC. Furthermore, the assessment of the reliability of all speeds was not possible due to the safety restrictions of the Fanuc CR-35iA which was used in the present study, and prevents the robot from reaching the two faster speeds from Section 5.1 (850mm/s and 1250mm/s). Therefore, this study was designed to measure participant trust in response to the one identified 'most suitable' speed. Where the ideal speed had been identified as 550mm/s, the functional limitations of the Fanuc CR-35ia required the speed to be increased by 2% to 600mm/s.

5.2.1.4 Experimental Task

5.2.1.4.1 Task Identification

While the experimental task used in the first participant trials (Section 5.1) was a simple abstract assembly task, for validity purposes, the task completed within this experiment needed to represent an Airbus installation task. This provides ecological validity for the results and provide evidence for effective deployment of a collaborative robot on to the shop floor within Airbus. Therefore, a demonstrator was planned to assess how well the CAD model with the integrated trust speed equation would predict human trust in the robot at the specified speed. To identify the most appropriate task for this project and experiment a task

was down-selected based on the following criteria for suitability of process for transfer to HRC:

1. Access and spatial layout – the current workstation layout / space would allow implementation of a robot of similar size and dimensions to the Fanuc CR-35iA.
2. Human physical performance – transfer of task steps to a robot would reduce manual handling demands, decrease task repetition and frequency, and improve body postures necessary to complete the task.
3. Logistics – the implementation of a robot would not impede logistical requirements, e.g. transportation and positioning of large components.
4. Payload restrictions – the process would not require the robot to handle a maximum payload exceeding 35 kg (the weight limit for the Fanuc CR-35iA).

The Airbus researcher collaborating on this project identified potential tasks and the project members applied the criteria detailed above to identify the Airbus single aisle aircraft slat installation process as most suitable for the demonstrator. Technically, the angle of placement of the slat component was considered most appropriate for the FANUC CR-35iA and the slats were thought to be within the robot's payload limits. For Airbus, transferring this manual process to HRC was expected to increase productivity, cycle time and ergonomic and health and safety improvements.

The Airbus researcher completed a technical breakdown of the task from beginning to end. From the breakdown the pick-up and placement of the slat onto the wing was identified as the most suitable aspect of the slat installation process for the Fanuc CR-35iA robot to complete as a demonstrator. The role of operators would change from moving the slat into place and securing it to the wing, to only bolting the slat into place. The benefit to this would include improved efficiency

and a reduced risk of musculoskeletal damage to operators, as they would no longer be required to manually lift the slats and hold them in place while they are secured onto the wing.

5.2.1.4.2 Demonstrator Task

The demonstrator task involved the robot performed a manual handling role of picking up and placing a representative slat into a frame and the participant secured the slat into place:

1. Participant sat under the frame as the operators would be expected to if this task was being completed on the shop floor.
2. Robot picked up the slat and moved it into place
3. Participant secured the slat with two bolts using an Allen key through the inside of the frame
4. Participant moved out from under frame and completed the Trust scale

5.2.1.5 Materials and Equipment

5.2.1.5.1 Robot Cell

A Fanuc CR-35iA industrial robot with a pneumatic gripper was used for this experiment because it is designed specifically for HRC with heavier payloads (35kg). As such, the robot has integral safety features which allowed participants to work in close proximity to it without the need for additional measures, such as safety light curtains or physical guarding. The robot and experimental set-up is shown below in Figure 5-4.

A representative slat section was custom made for this study. Although the Fanuc CR-35iA robot can lift items of up to 35kg close to its base, when the arm and component is further away from its base less weight can be moved. Therefore, the slat section model was scaled down from its original size to 17kg to accommodate the distance needed for the robot to place the slat onto the frame.

The frame representing the main wing component was designed to precisely fit the slat but also to protect the participant if the slat was dropped.



Figure 5-5: Experiment Robot and Layout

A chair with an adjustable back was provided for the participants so that they could lean backwards to fit under the frame and move into place. Although the height of the chair could be adjusted participants were asked not to make any changes as it was set to the same height as the benches used by the operators on the shop floor in Airbus.

A table was set up in front of the demonstrator cell, as shown in Figure 5-4, to house the PPE, bolts, and Allen key. The proximity of the tool and bolts was to ensure that time during the task was not taken up with moving between the tools and the mock-up. This is also where participants completed the Trust in Industrial HRC Scale after task completion.

5.2.1.5.2 Trust in Industrial Human-Robot Collaboration

Trust data was again collected using a paper-based questionnaire of the Trust in Industrial HRC Scale as previously described in Section 5.1.1.5.3.1.

5.2.1.6 Procedure

5.2.1.6.1 Experiment

5.2.1.6.1.1 Briefing and Consent

Participants took part in this study individually. Upon arrival, they received a verbal briefing, and were asked to read the written Brief (an example of which can be found in Appendix 6.3E.2). To ensure informed consent, it covered: the nature of the study and procedure, right to withdraw from the study, confidentiality and privacy, data management (including limitations on withdrawal of submitted data due to anonymisation procedures), dissemination of results, and researcher contact information for any subsequent enquiries. Participants were then asked to sign and confirm informed consent, and to provide some biographical data.

5.2.1.6.1.2 Experimental Task

A demonstration of the entire experimental task was given to participants, to familiarise them with the procedure as well as the robot's movement, speed, and the sound of the gripper. This was to ensure that the participants responses to the Trust scale were not based on a first impression of the robot, which may have negatively influenced the results. To begin the demonstration, participants were shown the PPE they needed to wear (a bump cap and safety glasses). Next, the researcher both demonstrated and described the full task procedure that participants would need to follow whereby they were sat under the frame, the robot picked up the slat and moved it into place, and the participant secured it with two bolts using an Allen key, with their arms through the inside of the frame.

It was emphasised to participants that they should sit under the frame while the robot moved the slat into place, and fit the bolts overhead, as this resembled the

shop floor slat installation process in Airbus. The start and finish positions of the experiment are shown in Figure 5-5 below.



Figure 5-5: Experiment Layout with Representative Slat Start (a) and Finish Positions (b)

After the demonstration, the participant's trial began. The technician started the robot program, which involved the robot picking up the representative slat and moving it into place onto the frame. The participant tightened the bolts and the task was complete. The participants were then asked to fill in the Trust in Industrial HRC Scale. Then the participants were then fully debriefed regarding the nature of the project, and reminded of their ethical entitlements. The researcher answered any questions the participant had regarding the study before thanking them for their participation.



Figure 5-7: Participant Completing the Bolt Tightening Task under the Frame

5.2.1.6.2 Virtual Environment and Model Development

While the experiment was being conducted, the AMRC engineers were transferring the demonstrator to a virtual environment and establishing a simulation based on the trust speed equation established in Section 5.1. This simulation was used to predict the expected Trust score based on the speed of the Fanuc CR-35iA.

The AMRC selected a Visual Components® package due to the following key features, which made it the most suitable:

1. Physical layout (CAD) and simulation in one package
2. A fully functioning model of the Fanuc robot (that could output Fanuc code)
3. Physical object simulation in real-time (for part jig interaction)
4. Customisable person and behaviour modelling

5. Creation of scripts and macros to run with the simulation (ideal for implementing the behaviour equation/s)

The demonstrator was transferred to a virtual environment and programmed to complete the task detailed in Section 5.2.1.4.2, the results of which can be seen in Figure 5-7. The AMRC engineers then proceeded to develop a Trust model based on the results from Section 5.1, with the polynomial equation shown below identified as the most suitable to model the trust relationship.

$$y = -9E-07x^2 + 0.0006x + 4.1526 \quad (5-1)$$

Equation 5-1 presents y as the Trust score from the Trust in Industrial HRC Scale and x as the robot speed expressed in millimetres per second. They then proceeded to add the equation into a Python script in the Visual Components model that looked at the movement speeds of the robot during the simulation run. They were then able to output, on the screen, the calculated Trust score for that of the robot speed in the virtual environment, as can be seen in Figure 5-7.

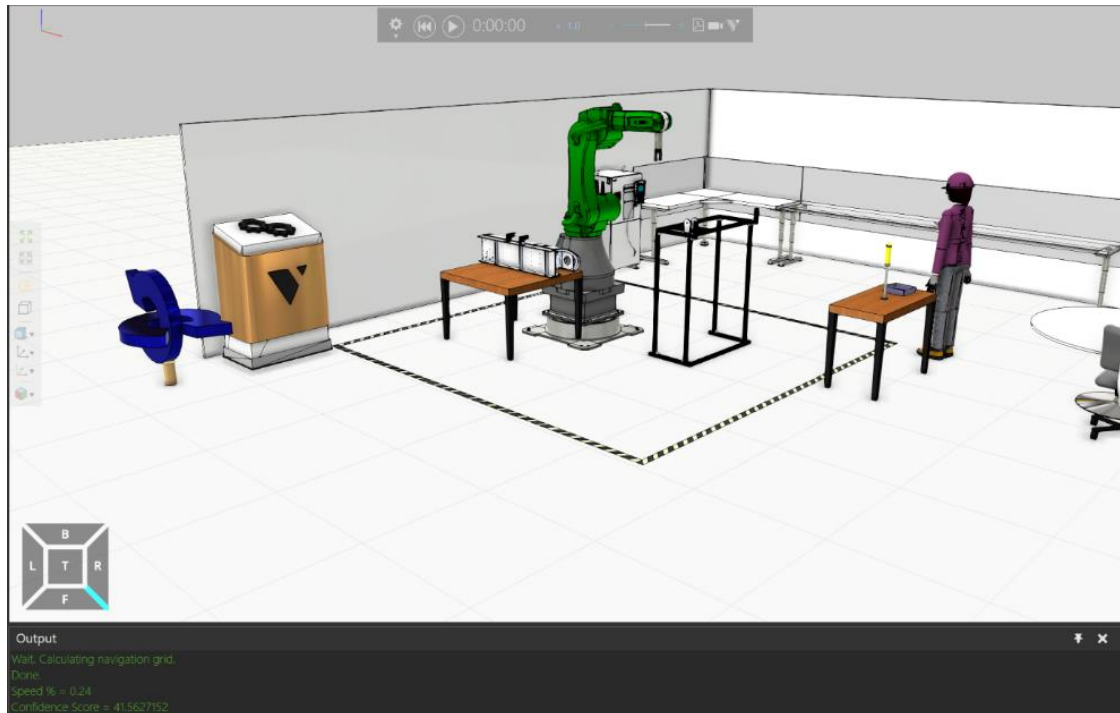


Figure 5-8: End Output from Simulation Run

The model provided a predicted Trust score of 41.5672, based on robot speed of 600mm/s, for the reliability experiment. The Results Section 5.2.2 identifies the accuracy of this prediction when compared to the experimental results.

5.2.1.7 Analysis

The Trust in Industrial HRC Scale was used to measure participants' trust within this experiment. Their responses were analysed in the same way as is detailed in Section 5.1.2.1, using the analysis guidelines provided by Charalambous et al. (2016). The final computed scores were averaged across all participants and are found along with their SDs in Table 5-8.

The results detailed in Section 5.1.3.1 revealed that the Trust score of interest was when the robot ran at a speed of 550mm/s. The experiment detailed in this

Section (5.2) looked to replicate this speed with an industrial scenario and a fully collaborative robot to assess the reliability of this finding and test the accuracy of the CAD HRC model. Therefore, a comparison of the overall trust scores obtained in Section 5.1, with the overall trust scores acquired from the reliability experiment, was required to show that the trust speed equation was valid. As ANOVAs do not provide evidence of an absence of effect within a population, alternative methods were required.

Quertemont (2011) recommends using a confidence interval test to provide evidence of an absence of an effect in the population. The confidence interval provides details about the highest difference likely to exist at the population level. Therefore, if the upper and lower bounds of the confidence interval of the trust scores obtained from the Reliability experiment conducted within this subheading fall within those from the experiment used to establish the trust speed equation (Section 5.1), it is possible to conclude that any variation between the two data sets is a consequence of natural population variation rather than response to the variables. Therefore, the two data sets from both participant studies were analysed using descriptive analysis in SPSS to identify the 95% confidence interval for the means of the two data sets (trust scores for the speed 550mm/s from the equation experiment (Section 5.1) and the trust scores gained from the Reliability study conducted within this subsection).

To further support these findings the minimum mean difference was identified for two samples of the same size and SDs of the observed data, to identify what significantly different data should look like. The minimum mean difference was identified as 2.927. Although the observed difference between the two values was below this threshold, it was necessary to ensure that it was significantly smaller and not likely to be generated by natural variance occurring within the population. Thus, to test for this, a Monte Carlo simulation was conducted.

The Monte Carlo simulation involved two datasets being simulated, both having the same sample size and SD to that of the real datasets but with their real mean difference being set to 2.927. In effect, 37 values were randomly generated from a normal distribution (mean = 100; SD = 5.485) for the first dataset, while 22 values were randomly generated from another normal distribution (mean = 102.927; SD = 5.294), thus replicating the sample sizes of the original datasets. The mean difference between the two simulated samples was recorded. After 1,000 iterations of this simulation, a normal distribution of expected mean differences was obtained, under the condition that the two populations had the minimum significant difference of 2.927 identified in the first step.

To compare the simulated and observed data, the simulated average and SD were compared with the observed difference to ensure that, if the two populations were different, the observed difference was significantly smaller than the smallest possible significant difference. A one-sample t-test then compared the smallest simulated mean difference and SD with the observed mean difference.

5.2.2 Results

Table 5-8 provides the descriptive statistics and confidence intervals for the Program 3 Trust score from Section 5.1.3.1, and the Reliability Trust score from the experiment conducted within this Section (5.2).

The confidence interval results presented in Table 5-8 below shows how closely the Trust level data derived from the experiment to establish the trust speed equation (Section 5.1.3.1) corresponds with the data collected in the Reliability experiment with the demonstrator. The one-sample t-test of the Monte Carlo simulated and the observed data revealed significance ($t(999) = 56.83$, $p < 0.05$) which suggests that the observed difference between the two conditions (trust responses to the robot moving at speed 3 in Section 5.1.3.1, and trust responses

from the reliability experiment detailed in Section 5.2) was due to random variation and not due to actual difference between the two populations.

Table 5-8: Descriptive Statistics and Confidence Interval Data

		Trust Speed Experiment Program 3 Trust Score	Reliability Experiment Demonstrator Trust Score
Mean (SD)		42.14 (5.294)	42.43 (5.485)
95% Confidence Interval for Mean	Lower Bound	39.789	39.22
	Upper Bound	44.484	43.689

The trust speed equation generated by the AMRC expected that operator trust scores would be around 41.56 in response to robot speed in the Reliability experiment. As shown in the Confidence interval results presented in Table 5-8, the average Trust score measured during the Reliability experiment was 42.43 (SD: 5.485). The trust speed equation's predicted value and the observed value from the reliability experiment are, therefore, very similar, with the expected value falling within a SD from the observed mean. Thus, the new HRC model suggests that the AMRC predicted the Trust score with good accuracy.

5.3 Discussion

This Chapter showed the effect of robot speed on participant trust, workload, and mood and the development of a psychological response equation for trust, this equation was developed into a virtual model and its accuracy assessed using a HRC demonstrator. Therefore, the final two Objectives (3 and 4) of this thesis as detailed in Section 2.2.2.5 have been met. This Chapter successfully

demonstrated the ability to integrate human psychological data into CAD, in a way that is accessible to design engineers or systems integrators looking to include a level of psychological analysis in the assessment of their designs. Where this level of integration had not been possible in Chapter 4 due to the complexity of workload, trust and speed provided a simplified construct for equation development and successful integration into CAD. This Chapter provides support for using this method as a means to improve CAD DHM tools by integrating further psychological constructs.

As detailed in the Literature Review (Chapter 2), one of the ways that manufacturing companies are looking to reduce the physical impact of work on their operators, where it is not possible to redesign the task or environment, is to introduce HRC (Judt et al., 2016). This involves operators working in close quarters and collaboratively with a robot to complete their tasks. The results of the current study provide a means of understanding this effect prior to the implementation of a robot into a manufacturing environment, and inform design decisions regarding robot speed. The speeds that could be achieved in the equation experiment, detailed in the first half of this Chapter, meant that the trust scores fell close to one another, which makes it difficult to generalise the model outside of these speeds. Ideally, more varied and extreme speeds would be tested to give a wider range of trust responses, which may affect the model. However, this is limited by what can be safely tested with people. Additionally, the assessment of all five speeds used to establish the equation was not possible due to the safety features of the collaborative robot (Fanuc CR-35iA), had it been possible, the reliability assessment would have been more robust. As safety technology improves, it may be possible to assess trust at faster speeds and to assess the accuracy of the trust response at all of these speeds, therefore, further informing and adjusting the model. As can be seen in the items that comprise the Trust in Industrial HRC Scale there are further factors that affect trust and

additional research that could be conducted to extend the trust equation further for implementation in CAD.

In the first section of this Chapter (5.1), the effect of robot speed on two additional constructs: mood and workload were assessed. The range of speeds that the participants were exposed to was found to have an overall significant effect on workload but not on mood. The time and resource constraints of the project meant that a single construct could be taken forward to develop and integrate an equation into CAD. As the trust speed results from Section 5.1.3.1 had a significant difference between the programs it meant that a discernible difference could be found and used to justify the speed chosen for the reliability assessment detailed in Section 5.2. Although the ANOVA for workload showed significance, the application of the Bonferroni correction changed the significance levels for the post hoc tests and meant that a significant difference was not found between the scores for each speed. A reason for the lack of an effect on workload may have been that the task was very simple, and that the robot did not continue to move while the participant was completing the task. Had the robot continued its movement throughout the task, there might have been an effect found in the post hoc tests on workload. The overall significance can be used to establish an equation and future research will look to assess the reliability of the developed equation. With regard to the mood results, although all of the statistical analyses showed no significant effect, the statistical analysis of the tense arousal scores were approaching significance, with a greater number of participants some significant results may have been found. It is possible that significant results may be found for mood and workload with different tasks, different robots and variation in other robot factors. Wider more extensive investigations into the various robot factors should be investigated, this includes robot size and type, gripper types, and robot movement.

A limitation of the mood survey was found during the experiment detailed in the first section of this Chapter (Section 5.1), none native English speakers struggled to understand some of the adjectives within the survey. The terms that participants requested definitions of included: Jittery, Sluggish, Vigorous and Unenterprising. Three of the participants asked for these terms to be defined, and, consequently, there is concern that some of the participants did not ask for the definitions of these terms and may have guessed at their meaning. A consequence of this is that the participants provided responses that did not reflect their true mood. It is advised that the UMACL be revised for non-native English speakers or an alternative method of assessing mood is used in experiments with none-native English speakers.

Care should be taken when generalising this model as the experimental population was derived mostly from students from Cranfield University which has a focus on science and technology and, therefore, may have led to a bias in the trust and mood results. Further research should look to assess trust responses to robot speed from a greater population pool, and include those who work in manufacturing environments and may find themselves working with collaborative robots. These studies would provide further evidence for the reliability of the equation and an understanding of variation in the equation that results from different robots, environments, and population samples. Additionally, larger populations should be conscripted in future experiments to enable modelling of the error bands of the equation. This would offer a result with a confidence range attached, allowing for a more accurate and useful output, and to determine a degree of confidence in the simulation's results.

5.3.1 Conclusion

This Chapter provides evidence that it is possible to develop a psychological equation that can be integrated into DHM CAD tools, for improving design and

robot integration decisions. Although it was not possible to establish an equation for all three psychological responses, an equation was established for trust. In establishing this equation and assessing its reliability the final two objectives of this thesis: 'to establish human-robot interaction equations for industrial HRC' (Objective 3) and 'to assess the predictive accuracy of the HRC equation after integration into a DHM CAD program' (Objective 4) were met.

6 DISCUSSION AND CONCLUSION

6.1 Discussion

The aim of this thesis was to identify and establish psychological parameters that can be used to improve the capability of DHM for manufacturing system design. The aim was achieved through meeting the four objectives: Objective 1 looked to assess the effectiveness of an inertial mocap suit to identify whether it could be used with confidence to measure postures in the experiment to meet Objective 2. Objective 2 proceeded to identify whether it would be possible to establish a cognitive parameter for mental workload through establishing the combined parameters of physical strain and mental workload. By identifying these parameters, it was possible to identify a predictive equation between physical strain and mental workload. The 3rd Objective involved establishing a human robot interaction equation for industrial HRC that could be integrated into DHM CAD programs and Objective 4 looked at assessing the predictive accuracy of the equation after integration into a DHM CAD program. The following paragraphs provide a summary of each of the objectives, how they were met in the studies detailed within this thesis, and how they sit within the context of the literature. Each section will then proceed to detail the implications of any limitations and areas for future research.

6.1.1 Objective 1: Investigating the Effectiveness of Motion Capture Technology

The assessment of the effectiveness of the Synertial IGS-180i mocap suit identified variations in agreement between the different software outputs and the goniometer measurements for the whole body and individual joints. A result of which is that in the process of looking to identify whether it would be effective for capturing a person's postures during shop floor work, the suit was identified as not wholly reliable; with a lack of agreement found between the goniometer and

various outputs for the right shoulder, both elbows, both wrists and both ankles. Therefore, in the experiment that followed to meet Objective 2, additional forms of mocap were utilised. These findings reinforced findings in the literature from Cloete and Scheffer (2008) and Lebel et al. (2017) regarding the limitations in accuracy of inertial mocap suits, particularly for the ankles.

This study looked specifically at angles of flexion and extension within the joints, and the measurements were conducted while the participant was stationary. Furthermore, the whole suit was tested rather than groups of two to three sensors. The reason for this was to rule out movement artefacts as a cause of error in the data, and to begin to develop an understanding of how the suit acts as a whole. This expanded understanding of the effectiveness of inertial mocap systems. Within the literature, their accuracy and effectiveness has been assessed during movement (Kim & Nussbaum, 2013; Lebel et al., 2017; Robert-Lachaine et al., 2017), and when limited to between two and three sensors (Brennan, et al., 2011; Brodie et al., 2008; Cutti et al., 2006; Roetenberg et al., 2007; Schall et al., 2016). Although these studies provide an understanding of inertial sensors in a short chain, or during movement, they do not provide an understanding of how accuracy may change across the suit as a whole, and with dynamic assessments it is challenging to identify where variation occurs with regard to specific body segments. The results presented in this thesis show that when the person is stationary, there are limitations in the effectiveness of the suit, that require addressing prior to being able to use the suit with confidence. It is likely that they will be addressed as the technology matures and becomes more stable.

There were limitations with this study and, consequently, areas for further investigation in future research. A single participant was assessed within this study, which means it is not possible to understand whether there are any differences in how the mocap suit records individuals with different physiologies.

Future work should look to investigate the effect individual physiological differences may have on the accuracy and effectiveness of inertial mocap suits, or ensure that the participant sample consists of more representative morphologies to account for this as an experimental factor.

A further limitation of this study was the comparison to the goniometer, although clinically this is a recommended method, it would have been preferred to compare the Synertial suit to the Vicon system. The reason for this is that the Vicon optical mocap system is considered the gold standard (Dutta, 2012) and, therefore, most accurate form of mocap. Although clinically the goniometer is acceptable, it is likely that it introduced measurement errors. This may mean that the suit is more accurate than the results suggested. Furthermore, the assessment was limited to measuring only flexion and extension, with no rotational movement accounted for. It is possible that the measurements were more accurate in these other orientations. A Vicon system would have provided information regarding these factors, however, it was not possible to gain access to a Vicon system for this thesis. Further research should look to compare the two devices, and establish a clearer understanding of the suits accuracy and effectiveness across all movements.

The investigation into the effectiveness of the mocap suit meant that its limitations were understood prior to attempting the experiment to meet Objective 2. As a result, cameras were sourced and integrated into the experimental design prior to beginning experimentation. It is possible that the limitations with inertial mocap are a result of the youth of the technology and since this study was first conducted, the technology has continued to develop. Therefore, it is possible that more effective results may be obtained with a newer inertial suit. As inertial technology continues to develop, its accuracy and effectiveness should be assessed, and if found to be suitably effective could be used with confidence within manufacturing to assess working postures. This technology would provide

valuable insight into the working postures of operators who work in confined and occluded environments. Furthermore, because it can give objective angular measurements when fully effective this information will be critical for providing greater insight into the effects that postures and movement have on MSDs and their development. It will provide an ability to link particular postures to specific damage (Smith, 1996), which is not currently possible due to the inability to provide absolute indicators of damage (Kilbom, 1997; Li & Buckle, 1999; Roman-Liu, 2014; Winkel & Westgaard, 1992). If found to be fully effective one of the additional benefits to this technology is that the recorded data can be uploaded into DHM CAD programs and used to control the movement of manikins. The advantage of this is that it would enable the application and assessment of activity and postural analysis tools across a task faster than it takes a person to carry out the assessment. This would significantly reduce the time required for postural analysis and would provide valuable insight into the physical demands of whole tasks (Johnson & Fletcher, 2015).

6.1.2 Objective 2: Establishing the Combined Parameters of Physical Strain and Mental Workload

Objective 2 contributed to the aim of identifying and establishing psychological parameters that will improve the capability of digital human modelling for manufacturing system design by providing a means of measuring mental workload within a DHM CAD program. This was achieved through “establishing the combined parameters of physical strain and mental workload”. A repeated measures experiment was used to establish these parameters and produced a statistically significant predictive equation for the NASA-TLX Mental Demand dimension. The change in mental workload was predicted from physical strain as measured using the REBA postural assessment method. The identification of this equation is a step closer to virtual environments described by Sundin and Örtengren (2006) which provides design engineers with a more holistic means of

assessing their designs, by accounting for impacts on both physical and psychological processes. Statistically significant predictive equations were also identified for the Overall NASA-TLX scores and the HRV IBI statistic, however, there is difficulty in isolating mental demand from physical demand within both of these results (Backs et al., 1994; DiDomenico & Nussbaum, 2011).

The establishment of the predictive model between REBA and the NASA-TLX Mental Demand dimension by means of an experiment brings empirical evidence supporting the interaction between physical and mental workload (DiDomenico & Nussbaum, 2008; Marras & Hancock, 2014; Mehta & Agnew, 2011; Shaikh et al., 2012a, 2012b). The results support that an interaction effect occurs between the two factors, and that it does so in a measurable way. This provides an understanding that when operators are expected to work in environments that concurrently require physical and cognitive expenditure, there is likely to be a cognitive cost. If this cost is in the form of a high load on mental workload, the acute effects of errors and poor performance are likely for manufacturing companies. As suggested by Marras and Hancock (2014), it is likely that this effect on cognitive capacity results from the need to manage discomfort and later fatigue. Based on this, it is possible that the more time individuals spend in postures that have high risks of MSDs, the greater the degradative effect on cognition. Therefore, when establishing a Mental Demand equation or any other cognitive equation, research should look to assess the effect of time on the equation's reliability. It is possible that demographic factors such as age may additionally have an effect on this equation and should be investigated. Furthermore, by establishing this understanding of the cost of physical strain on mental workload capacity and the time components of it, it should be possible to provide companies with an understanding of how best to manage breaks, task rotation and recovery times when integrated into digital prediction and monitoring programs.

The effectiveness of the mocap suit being established prior to this experiment led to the inclusion of two cameras for recording postures during the experiments. Where the suit functioned effectively, it provided invaluable information regarding the neck, wrists, and forearm when they were occluded by the mock-up. Unfortunately, the suit was effective for just over 50% of the conditions, where it was not effective and the joints could not be seen, the lower score given to the joint may have biased the data by attributing greater mental workload to lower physical demands, which would have affected the reliability of the predictive equation. It is likely that newer mocap suits are more reliable, however, future work looking to assess postures should ensure the accuracy of the mocap technology. Where the accuracy of the suit cannot be guaranteed it is recommended that tasks be redesigned in such a way that the whole posture can be observed and assessed.

The predictive equation for NASA-TLX Mental Demand could be introduced into a CAD program to be used to predict mental workload based on the postural demands identified during an assessment of a virtual design. However, this would require the introduction of REBA to DHM CAD programs. Although no DHM CAD programs currently have REBA within their postural assessment tools, they do have RULA. RULA was assessed as unsuitable for use within the experiment because it would not have effectively assessed the awkward postures that would be adopted during the tasks in the experiment (Chiasson et al., 2012). REBA, on the other hand, more effectively assesses these types of postures. As REBA is an extension of RULA, it should be possible to include REBA in DHM CAD programs too. If it were to be integrated, it would provide design engineers with a more effective means of assessing the whole body and specifically awkward postures.

The DV in the experiment conducted to meet Objective 2 was based on changes in an IV that could be measured within a CAD program, if REBA were integrated.

The reason for this decision was to overcome the issues found with current computer models of cognition, which tend not to be in a form readily available and accessible to design engineers (Duffy, 2012). Postural assessment tools are available within DHM CAD programs and this assessment type is already used by design engineers (Chiu et al., 2012; Cimino et al., 2009; Estefani, 2011; Gomez, 2014; Judt et al., 2016; Lockett et al., 2014). Therefore, by tying the cognitive assessment to physical assessments, it was felt that engineers would be better able to easily understand and interpret the results.

A limitation of the mental workload equation is the lack of an upper and lower bound of what is acceptable for workload. The lack of a “red line” in mental workload, where someone moves from effective workload levels to ineffective levels (Hart & Wickens, 2008 as cited in Grier et al., 2008), makes identification of over or underload and, therefore, identifying the effectiveness of a task redesign difficult. Grier’s (2015) work has looked to establish these bounds, however, the upper and lower bounds are task dependent and further research is required to establish more generalised boundaries. It is likely that this will be a common limitation of all psychological models developed for integration into DHM CAD programs. Identifying “red lines” becomes difficult where variation in capacity occurs across individuals and other factors or conditions contribute to variation.

Generalising the equations found in this experiment beyond this experiment should be done cautiously. Further work is required to validate the equation established by the experiment conducted to meet this objective, with a larger participant pool and a range of different tasks. The validation process should look to investigate and account for variations between types of tasks, levels and types of physical strain, and types of mental workload or processing capacities as defined by multiple resource theory (Wickens, 2002). Experiments could be conducted to identify capacities in verbal or spatial information coding, visual and

auditory perception, aspects of memory and decision-making, and both manual and verbal response strategies. This expanded validation process should be considered when looking to establish other psychological equations for integration into DHM CAD programs.

6.1.3 Objectives 3 & 4: Establish Human-Robot Interaction Equation/s for Industrial Human Robot Collaboration and Assessment of the Predictive Accuracy of the HRC Equation/s after Integration into DHM

Objective 3 and 4 looked to not only establish equations that could be integrated into a DHM CAD program, but to also integrate them into a CAD program and assess their predictive accuracy. The first experiment conducted to identify equations that could be integrated into a DHM CAD program was only able to establish two statistically significant equations for Trust, and workload. However, one could not be established for Mood. The time and resource constraints of the study meant that only a single construct could be taken forward to establish an equation, model it in a CAD program and test its reliability. The trust speed interaction was chosen due to the significant differences found in the post hoc tests, as this meant a speed could be identified for the reliability assessment purposes. With the skill and expertise of engineers from the AMRC, a predictive equation was established and the trust speed equation was integrated into a CAD program. The predictive accuracy for the trust speed equation was then successfully assessed using a mock-up of a redesigned pre-existing Airbus process and used a fully collaborative robot. Thereby, establishing psychological parameters that could improve the capability of digital human modelling for manufacturing system design and fulfilling the aim of this thesis.

As with the results from the experiment conducted to meet Objective 2, the DV in the experiments conducted to meet Objectives 3 and 4 were based on changes in an IV that could be measured within a CAD program. In this case, the IV was

robot speed. Robots and their programmable variables can be imported into CAD programs, and are readily used by design engineers. Therefore, by connecting Trust with robot speed it is hoped that this will provide design engineers and systems integrators with an easily accessible and understandable tool for use. This inclusion of perceived trust that relates to robot speed would benefit design engineers and systems integrators looking to understand how a redesign of a process that includes implementation of collaborative robots may affect operators. Further, this speed trust equation could be integrated into digital based task design and monitoring programs and enable management to plan workspaces that account for the psychological requirements of their workforce.

There is a limitation with this equation and that is the large range of what is considered acceptable for trust, from a Trust score of 25 to 50 on the Trust in Industrial HRC scale (Charalambous et al., 2016). It is likely that there are nuanced levels of trust at different increments within this range and further work should be conducted to understand these differences. Changes in the different trust dimensions may influence this, either due to the item weighting or due to the factors that negatively affect trust. A recommendation is that thresholds for each of the three dimensions may provide clear indicators for good trust levels. This could help to mitigate any possible bias in the different item loadings across the three dimensions, and provide a better diagnostic tool for measuring trust.

A limitation of the Trust scale is that there is a biased weighting between the dimensions. There are four items for both the “Safe co-operation” and “Robot and gripper reliability” dimensions, whereas the dimension “Robot motion & pick-up speed” only has two items. To achieve the overall trust score, the scores are summed together. This produces a weighting towards the two dimensions with more items. Further work is required to understand the effect of this weighting and establish whether the scale requires further refinement. Limitations were additionally found in the UMACL mood scale, with the language of some of the

items found to be unsuitable to non-native English speakers. In future research adjective type surveys should be reviewed for understanding within a pilot study to ensure that all items are comprehensible.

Although the workload results were found to be significant, the lack of significance found in the post hoc tests meant that this psychological construct was not taken forward within this research because it did not enable an effective identification of a suitable speed for the reliability assessment. However, future research will look to establish an equation using a nonparametric regression and test its reliability.

The mood results were not found to significantly vary as robot speed changed in the context of this experiment. However, these factors are known to be affected when interacting with robots and automation (Parasuraman et al., 2008, Stokes et al., 2010, Vidulich & Tsang, 2012). Therefore, it should be assessed in the context of other factors that vary within HRC such as robot size, movement, and level of interaction between the person and the robot. Within the literature, it is clear that mood and trust can shift from negative to positive as individuals spend more time with a robot (Stokes et al., 2010). Therefore, future research should look to develop an understanding of the role of time and exposure on these variables, particularly when looking to establish equations for integration into DHM CAD programs.

6.2 Contribution to Knowledge

The research detailed within this thesis has contributed not only to intellectual knowledge but it has also had an impact within the manufacturing industry. The literature review has provided insight into the current capabilities and limitations of DHM CAD programs. These were not clear within the literature, and the information presented on the various company websites were not comprehensive. This review has identified clear gaps and limitations with the

available software, particularly in relation to manufacturing, and highlighted areas for improvement for both physical and psychological assessments. This information makes it very clear to DHM CAD developers regarding what could be improved to aid in manufacturing workplace design.

The motion capture suit experiment provides identification of the limitations in the effectiveness of motion capture technology, as well as their effectiveness for recording work completed in occluded and confined working conditions. The results indicate specific limbs where the suit does not effectively capture motion, providing the developers of the system with a targeted means of improving it. As already stated within this thesis, the age of the suit likely contributed to these findings, however, similar accuracy and effectiveness assessments can only provide insight into the suitability of this technology for assessing occluded working conditions. Although a comparison was not made with a Vicon system, the protractor goniometer provides a low technology means of conducting an effectiveness assessment.

The establishment of the workload and REBA equations contribute to the current limited understanding of the interaction between the two as identified by Marras and Hancock (2014). The equations provide a starting point for further work, for validation and further refinement of this interaction. Furthermore, by establishing this relationship between the two conditions a more easily accessible means of assessing cognition has been provided, compared to current cognitive modelling tools. These findings have wider implications for understanding cognition and suggest that the implication of other factors should be considered during assessment.

The low complexity task used to identify the predictive cognitive and physical interaction equations was identified from the Airbus A320 factory in Broughton. The visit and observation of tasks led to the production of a report for the Health

and Safety Controller in Airbus Broughton. The report covered working postures for Rib 2 sealing activities, in-tank bolt sealing, assessment of environmental conditions and psychological factors with in-tank working. This provided insight into the working conditions of operators, and provided recommended remedial changes to improve current working conditions.

The development of the trust speed equation and integration into a CAD program provides supporting evidence for the importance of speed in the HRC trust relationship. This knowledge contributes to our current understanding of how the introduction of this relatively new technology may affect those who may be required to work with it. The reliability of the integrated trust speed equation in a CAD program provides evidence for the ability to identify psychological equations and include them in DHM CAD programs. This work was part of a feasibility study funded by the EPSRC Connected Everything Network Plus and was called ICHORD (Integrating the Cognitions of Human Operators in digital Robot Design). Despite requiring further refinement of the equation, it shows that it is possible to establish equations to predict psychological reactions that can be used to aid the development of HRC systems, with the intended impact of supporting operators' acceptance of robots on the shop floor. This work has led to an in progress funding application, to further develop understanding of HRC psychological reactions with the intention of building a library of predictive equations that can be used to inform design decisions when introducing HRC.

As a whole, this thesis has shown that it is possible to predict psychological responses and provide access to these predictions in easily accessible forms for design engineers. This ability to predict psychological responses begins to make it possible to also predict possible behaviours, and the ability to predict these behaviours means that workplaces will be better able to adapt or take into account operator variability, working pace, and preferences. This understanding and predictive capability provides Industry 4 with a step closer to being able to

account for human variation and to find a way to enable people to work at their most productive.

6.3 Conclusion

This thesis aimed to identify whether it was possible to measure psychological responses to stimuli in such a way that they could be integrated into DHM CAD programs. The intention was to be able to provide design engineers with a method of considering the psychological responses of those who will interact with their designs, in a similar way to how they account for physical interactions. The establishment of predictive equations between mental workload and REBA scores; and the identification and reliability assessment of the trust speed equation has shown that the provision of psychological reactions in DHM CAD programs is possible. Although, more research is required to validate and investigate further factors that should be considered within DHM CAD programs. This thesis provides a step towards achieving DHM that is capable of assessing both psychological and physical processes for working conditions.

REFERENCES

- Albin, T. J. (2013). Evaluating the efficiency of checklists used to assess risk of musculoskeletal disorder. *Proceedings of the Human Factors and Ergonomics Society 57th Annual Meeting*, 57(1), 985-988.
- Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, 24(2), 225-235.
- Anderson, J. R., & Lebiere, C. (1998). *The Atomic Components of Thought*. Lawrence Erlbaum.
- Arai, T., Kato, R., & Fujita, M. (2010). Assessment of operator stress induced by robot collaboration in assembly. *CIRP Annals*, 59(1), 5-8.
- Archer, S.G., & Adkins, R. (1999) *IMPRINT User's Guide*. Prepared for U.S. Army Research Laboratory, Human Research and Engineering Directorate.
- Backs, R. W. (2001). An autonomic space approach to the psychophysiological assessment of mental workload. In P. A. Hancock, & P. A. Desmond, (Eds.), *Human Factors in Transportation. Stress, Workload, and Fatigue* (pp. 279-289). CRC Press.
- Backs, R. W., Ryan, A. M., & Wilson, G. F. (1994). Psychophysiological measures of workload during continuous manual performance. *Human Factors*, 36(3), 514- 531.
- Bao, S., Howard, N., Spielholz, P., Silverstein, B., & Polissar, N. (2009). Interrater reliability of posture observations. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 51(3), 292-309.

Battini, D., Faccio, M., Persona, A., & Sgarbossa, F. (2011). New methodological framework to improve productivity and ergonomics in assembly system design. *International Journal of Industrial Ergonomics*, 41(1), 30-42.

Battini, D., Persona, A., & Sgarbossa, F. (2014). Innovative real-time system to integrate ergonomic evaluations into warehouse design and management. *Computers & Industrial Engineering*, 77, 1-10.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, 78(5), B231-B244.

Bittner, A. V., Byers, J. C., Hill, S. G., Zaklad, A. L., & Christ, R. E. (1989). Generic workload ratings of a mobile air defense system (LOS-FH). In *Proceedings of the 33rd Annual Meeting of the Human Factors and Ergonomics Society*, 1476-1480. HFES.

Bonnechere, B., Jansen, B., Salvia, P., Bouzahouene, H., Omelina, L., Moiseev, F., Sholukha, V., Cornelis, J., Rooze, M., & Van Sint Jan, S. (2014). Validity and reliability of the Kinect within functional assessment activities: comparison with standard stereophotogrammetry. *Gait & Posture*, 39(1), 593-598.

Brennan, A., Zhang, J., Deluzio, K., & Li, Q. (2011). Quantification of inertial sensor-based 3D joint angle measurement accuracy using an instrumented gimbal. *Gait & Posture*, 34(3), 320-323.

Broberg, O. (1997). Integrating ergonomics into the product development process. *International Journal of Industrial Ergonomics*, 19(4), 317-327.

Brodie, M. A., Walmsley, A., & Page, W. (2008). Dynamic accuracy of inertial measurement units during simple pendulum motion. *Computer Methods in Biomechanics and Biomedical Engineering*, 11(3), 235-242.

- Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of Management Review*, 13(1), 40-52.
- Cappozzo, A., Catani, F., Leardini, A., Benedetti, M. G., & Della Croce, U. (1996). Position and orientation in space of bones during movement: experimental artefacts. *Clinical Biomechanics*, 11(2), 90-100.
- Castellucci, H. I., Viviani, C. A., Molenbroek, J. F. M., Arezes, P. M., Martínez, M., Aparici, V., & Bragança, S. (2019). Anthropometric characteristics of Chilean workers for ergonomic and design purposes. *Ergonomics*, 62(3), 459-474.
- Causse, M., Chua, Z., Peysakhovich, V., Del Campo, N., & Matton, N. (2017). Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. *Scientific Reports*, 7(1), 1-15.
- Chang, S., & Wang, M. (2007). Digital human modelling and workplace evaluation: using an automobile assembly task as an example. *Human Factors and Ergonomics in Manufacturing*, 17(5), 445-544.
- Chang, Y. J., Chen, S. F., & Huang, J. D. (2011). A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in Developmental Disabilities*, 32(6), 2566-2570.
- Charalambous, G. (2014). *The Development of a Human Factors Tool for The Successful Implementation of Industrial Human-Robot Collaboration* [Unpublished doctoral dissertation, Cranfield University]. Cranfield University.
- Charalambous, G., Fletcher, S., & Webb, P. (2016). The development of a scale to evaluate Trust in industrial human-robot collaboration. *International Journal of Social Robotics*, 8(2), 193-209.
- Chertman, C., Santos, H. M. C. D., Pires, L., Wajchenberg, M., Martins, D. E., & Puertas, E. B. (2010). A comparative study of lumbar range of movement in

healthy athletes and non-athletes. *Revista Brasileira De Ortopedia*, 45(4), 389-394.

Chiasson, M., Imbeau, D., Aubry, K., & Delisle, A. (2012). Comparing the results of eight methods used to evaluate risk factors associated with musculoskeletal disorders. *International Journal of Industrial Ergonomic*, 42(5), 478-488.

Chiu, Y. C., Chen, S., Wu, G. J., & Li Y. H. (2012). Three-dimensional computer-aided human factors engineering analysis of a grafting robot. *Journal of Agricultural Safety and Health*, 18(3), 181-194.

Cimino, A., Longo F. & Mirabelli, G. (2009). A multimeasure-based methodology for the ergonomic effective design of manufacturing system workstations. *International Journal of Industrial Ergonomics*, 39(2), 447-455.

Clarkson, H. M. (2000). *Musculoskeletal assessment (2nd ed.)*. Lippincott, Williams & Wilkins.

Cloete, T., & Scheffer, C. (2008). Benchmarking of a full-body inertial motion capture system for clinical gait analysis. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 4579-4582). IEEE.

Colle, H. A., & Reid, G. B. (1998). Context effects in subjective mental workload ratings. *Human Factors*, 40(4), 591-600.

Colombini, D. & Occhipinti, E. (2006). Preventing upper limb work-related musculoskeletal disorders (UL-WMSDs): New approaches in job (re)design and current trends in standardization. *Applied Ergonomics*, 37(4), 441-450.

Cook, J. R., Baker, N. A., Cham, R., Hale, E., & Redfern, M. S. (2007). Measurements of Wrist and Finger Postures: A Comparison of Goniometric and Motion Capture Techniques. *Journal of Applied Biomechanics*, 23(1), 70-78.

Cuff, D.J. & Pupello, D.R. (2012). Prospective randomized study of arthroscopic rotator cuff repair using an early versus delayed postoperative physical therapy protocol. *Journal of Shoulder and Elbow Surgery*, 21(11), 1450-1455.

Cutti, A. G., Giovanardi, A., Rocchi, L., & Davalli, A. (2006). A simple test to assess the static and dynamic accuracy of an inertial sensors system for human movement analysis. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 5912-5915). IEEE.

Cutti, A. G., Paolini, G., Troncossi, M., Cappello, A., & Davalli, A. (2005). Soft tissue artefact assessment in humeral axial rotation. *Gait & Posture*, 21(3), 341- 349.

Damgrave R.G.J., & Lutters, D. (2009). The Drift of the Xsens Moven Motion Capturing Suit during Common Movements in a Working Environment *Proceedings of the 19th CIRP Design Conference – Competitive Design*, (pp. 338- 342). Cranfield University Press.

Daniell, N., Fraysse, F., & Paul, G. (2012). How does the size and shape of local populations in China compare to general anthropometric surveys currently used for product design?. *Work*, 41(1), 4088-4090.

Das, B., Shikdar, A.A., & Winters, T. (2007). Workstation Redesign for a Repetitive Drill Press Operation: A Combined Work Design and Ergonomics Approach. *Human Factors and Ergonomics in Manufacturing*, 17(4), 395–410.

DiDomenico, A., & Nussbaum, M. A. (2008). Interactive effects of physical and mental workload on subjective workload assessment. *International Journal of Industrial Ergonomics*, 38(11-12), 977-983.

DiDomenico, A., & Nussbaum, M. A. (2011). Effects of different physical workload parameters on mental workload and performance. *International Journal of Industrial Ergonomics*, 41(3), 255-260.

Dockrell, S., O'Grady, E., Bennett, K., Mullarkey, C., Mc Connell, R., Ruddy, R., Twomey, S., & Flannery, C. (2012). An investigation of the reliability of Rapid Upper Limb Assessment (RULA) as a method of assessment of children's computing posture. *Applied Ergonomics*, 43, 632-636.

Drury, C.G. (2000). Global quality: linking ergonomics and production. *International Journal of Production Research*, 38(17), 4007–4018.

Duffy, V. G. (2012). Human digital modeling in design. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics* (4th ed., pp. 1016 - 1030). John Wiley & Sons, Inc.

Dunning, K.K., Davis, K.G., Cook, D., Kotowski, S.E., Hamrick, C., Jewell, G. & Lockey, J. (2010). Costs by industry and diagnosis among musculoskeletal claims in a state workers compensation system: 1999-2004. *American Journal of Industrial Medicine*, 53(3), 276-284.

Dutta, T. (2012). Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. *Applied Ergonomics*, 43(4), 645-649.

Dyer, S., Martin, J., & Zulauf, J. (1995). *Motion Capture White Paper*. Accessed 02/18/2014. ftp://ftp.sgi.com/sgi/A%7CW/jam/mocap/MoCapWP_v2.0.html

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference* (pp. 789-795). IEEE.

Endsley, M. R. (2012) Situation awareness. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics* (4th ed., pp. 553 - 568). John Wiley & Sons, Inc.

Estefani, J. (2011). *A Maintainability Predication Method Integrated with CAD*. [Unpublished Master's dissertation]. Cranfield University.

Evans, J. D. (1996). *Straightforward Statistics for the Behavioral Sciences*. Thomson Brooks/Cole Publishing Co.

Eysenck, H. J. (1983). *Individual Differences and Psychopathology*. Academic Press.

Ferguson, S.A., Marras, W.S., Allread, W.G., Knapik, G.G., Vandlen, K.A., Splittstoesser, R.E. & Yang, G. (2011). Musculoskeletal disorder risk as a function of vehicle rotation angle during assembly tasks, *Applied Ergonomics*, 42(5), 699-709.

Fletcher, S. R., Johnson, T. L., & Larreina, J. (2019). Putting people and robots together in manufacturing: are we ready?. In *Robotics and Well-Being: Vol.95* (pp. 135-147). Springer.

Fletcher, S. R., Johnson, T. L., & Thrower, J. (2018). A study to trial the use of inertial non-optical motion capture for ergonomic analysis of manufacturing work. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 232(1), 90-98.

Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics*, 73, 90-99.

Gomez, J. (2014). *Tooling Conceptual Design for Ergonomics Improvement on Fuel System Installation of an A320 Aircraft*. [Unpublished Master's dissertation]. Cranfield University.

Gopinath, V., Ore, F., Grahn, S., & Johansen, K. (2018). Safety-focussed design of collaborative assembly station with large industrial robots. *Procedia Manufacturing*, 25, 503-510.

Grahn, S., Gopinath, V., Wang, X. V., & Johansen, K. (2018). Exploring a Model for Production System Design to Utilize Large Robots in Human-Robot Collaborative Assembly Cells. *Procedia Manufacturing*, 25, 612-619.

Grier, R. A. (2015). How high is high? A meta-analysis of NASA-TLX global workload scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59(1), pp. 1727-1731. SAGE Publications.

Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J. G., & St. John, M. (2008, September). The red-line of workload: Theory, research, and design. In *Proceedings of the human factors and ergonomics society annual meeting*, 52(18) pp. 1204-1208. Sage Publications.

Hamilton, N., Weimar, W., & Luttgens K., (2011). *Kinesiology: Scientific Basis of Human Motion* (12th ed.). McGraw-Hill Higher Education

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 139-183.

Hart, S., & Wickens, C. D. (2008). Mental Workload. *NASA Human Integration Design Handbook*.

Helander, M. G. & Khalid, H. M. (2012). Affective Engineering and Design. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics* (4th ed., pp. 569-596). John Wiley & Sons, Inc.

Hendy, K. C., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: when is one scale better than many?. *Human Factors*, 35(4), 579-601.

Henelius, A., Hirvonen, K., Holm, A., Korpela, J., & Muller, K. (2009). Mental workload classification using heart rate metrics. *2009 Annual International*

Conference of the IEEE Engineering in Medicine and Biology Society (pp. 1836-1839). IEEE.

Hignett, S., & McAtamney, L. (2000). Rapid Entire Body Assessment (REBA). *Applied Ergonomics*, 31, 201-205.

Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner Jr, A. C., Zaklode, A. L., & Christ, R. E. (1992). Comparison of four subjective workload rating scales. *Human Factors*, 34(4), 429-439.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, 45(1-3), 73-93.

Howe, D. (2004). *Aircraft Loading and Structural Layout*. AIAA

HSE (2019). *Work related musculoskeletal disorder statistics (WRMSDs) in Great Britain, 2019*. Accessed 19/9/2019:

<http://www.hse.gov.uk/statistics/causdis/msd.pdf>

Hsiao, H., Whitestone, J., Kau, T. Y., Whisler, R., Routley, J. G., & Wilbur, M. (2014). Sizing firefighters: method and implications. *Human Factors*, 56(5), 873-910.

International Federation of Robotics. (2021). *IFR Press Release Top 5 Robot Trends 2021: Robots learn new tricks and help to reduce carbon footprint*. Accessed: 23/02/2021: <https://ifr.org/ifr-press-releases/news/top-5-robot-trends-2021>

International Organization for Standardization. (2016). Robots and robotic devices — Collaborative robots (ISO/TS 15066:2016). Accessed 14/08/2020: <https://www.iso.org/standard/62996.html>

International Organization for Standardization. (2018). Mechanical Vibration – Measurement and Evaluation of Machine Vibration (ISO 20816-8:2018). Accessed 9/06/2020: <https://www.iso.org/obp/ui/#iso:std:iso:20816:-8:ed-1:v1:en>

Janowitz, I. L., Gillen, M., Ryan, G., Rempel, D., Trupin, L., Swig, L., Mullen, K., Rugulies, R., & Blanc, P. D. (2006). Measuring the physical demands of work in hospital settings: Design and implementation of an ergonomics assessments. *Applied Ergonomics*, 37, 641–658.

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International journal of cognitive ergonomics*, 4(1), 53-71.

Johnson, C., & Goodman, N. W. (2006). Time to stop sniffing the air: snapshot survey. *The British Medical Journal*, 333(7582), 1295-1296.

Johnson, T. L., & Fletcher, S. R. (2014). A Computer Software Method for Ergonomic Analysis Utilising Non-Optical Motion Capture. In S. Sharples, & S. Shorrock, (Eds.), *Contemporary Ergonomics and Human Factors*. (pp. 93-100). CRC Press.

Johnson, T. L., & Fletcher, S. R. (2015). The limitations of using only CAD and DHM in design relating to high value manufacturing. In *Contemporary Ergonomics and Human Factors 2015: Proceedings of the International Conference on Ergonomics & Human Factors* (pp. 329-332). CRC Press.

Judt, D., Forster, K., Lockett, H. L., Lawson, C. P., & Webb, P. (2016). Aircraft wing build philosophy change through system pre-equipping of major components. *SAE International Journal of Aerospace*, 9(1), 190-197

Judt, D., Lawson, C., & Lockett, H. (2020). Experimental investigation into aircraft system manual assembly performance under varying structural component

orientations. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 234(4), 840-855.

Juul-Kristensen, B., Fallentin, N., & Ekdahl, C. (1997). Criteria for classification of posture in repetitive work by observation methods: A review. *International Journal of Industrial Ergonomics*, 19, 397-411.

Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall.

Karhu, O., Kansi, P., & Kuorinka, I. (1977). Correcting working postures in industry: A practical method for analysis. *Applied Ergonomics*, 8(4), 199-201.

Karwowski, W. (2012). The discipline of human factors and ergonomics. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics* (4th ed., pp. 1053–1078). John Wiley & Sons, Inc.

Keyserling, W. M., Wiggermann, N., Werner, R. A., & Gell, N. (2010). Inter-Worker Variability in Lower Body Postures During Assembly Line Work: Implications for Exposure Assessment. *Journal of Occupational and Environmental Hygiene*, 7(5), 261-271.

Kilbom, A. (1997). Repetitive work of the upper extremity: guidelines for the practitioner. In P. Seppalla, T. Luopajarvi, C. H. Nygrrd & M. Mattila (Eds.), *Experience to Innovation, Proceedings of the 13th Triennial Congress of the International Ergonomics Association*, 4, (pp. 66- 68).

Kim, S., & Nussbaum, M. A. (2013). Performance evaluation of a wearable inertial motion capture system for capturing physical exposures during manual material handling tasks. *Ergonomics*, 56(2), 314-326.

Klippert, J., Gudehus, T., & Zick, J. (2012). A Software-Based Method for Ergonomic Posture Assessment in Automotive Preproduction Planning: Concordance and Difference in Using Software and Personal Observation for

Assessment. *Human Factors and Ergonomics in Manufacturing and Service Industries*, 22(2), 156-175.

Koppenborg, M., Nickel, P., Naber, B., Lungfiel, A., & Huelke, M. (2017). Effects of movement speed and predictability in human–robot collaboration. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 27(4), 197-209.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.

Laughery, K. R., Plott, B., Matessa, M., Archer, S., & Lebiere, C. (2012). Modeling human performance in complex systems. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics* (4th ed., pp. 931 - 961). John Wiley & Sons, Inc.

Lebel, K., Boissy, P., Nguyen, H., & Duval, C. (2017). Inertial measurement systems for segments and joints kinematics assessment: towards an understanding of the variations in sensors accuracy. *Biomedical Engineering Online*, 16(1), 56.

Lee, J. D. (2006). Affect, attention, and automation. In A. F. Kramer, D. A. Wiegmann, & A. Kirlik (Eds.), *Series in human-technology interaction. Attention: From Theory to Practice* (p. 73–89). Oxford University Press.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243-1270.

Leiden, K., Byrne, M. D., Corker, K. M., Deutsch, S. E., Lebiere, C., & Wickens C.D. (2007). Overview of human performance modeling tools. In D. C. Foyle, & B. L. Hooey (Eds), *Human Performance Modeling in Aviation* (pp. 67-76). CRC Press.

Li, G., & Buckle, P. (1999). Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. *Ergonomics*, 42(5), 674-695.

Lin, Y. C., Wang, M. J. J., & Wang, E. M. (2004). The comparisons of anthropometric characteristics among four peoples in East Asia. *Applied Ergonomics*, 35(2), 173-178.

Lockett, H., Fletcher, S., & Luquet, N. (2014). Applying Design for Assembly Principles in Computer Aided Design to Make Small Changes that Improve the Efficiency of Manual Aircraft Systems Installations. *SAE International Journal of Aerospace*, 7(2), 284-291.

Luximon, A., & Goonetilleke, R. S. (2001). Simplified subjective workload assessment technique. *Ergonomics*, 44(3), 229-243.

Mackay, C., Cox, T., Burrows, G., & Lazzerini, T. (1978). An inventory for the measurement of self-reported stress and arousal. *British Journal of Social & Clinical Psychology*, 17(3), 283–284.

Marras, W. S., & Hancock, P. A. (2014). Putting mind and body back together: a human-systems approach to the integration of the physical and cognitive dimensions of task design and operations. *Applied Ergonomics*, 45(1), 55-60.

Matthews, G., Jones, D. M., & Chamberlain, A. G. (1990). Refining the measurement of mood: The UWIST mood adjective checklist. *British Journal of Psychology*, 81(1), 17-42.

McAtamney, L., & Corlett, E. N. (1993). RULA: a survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, 24(2), 91-99.

- McGinley, J. L., Baker, R., Wolfe, R., & Morris, M. E. (2009). The reliability of three-dimensional kinematic gait measurements: a systematic review. *Gait & Posture*, 29(3), 360-369.
- Mehrabian, A. (1977). Individual differences in stimulus screening and arousability. *Journal of Personality*, 45(2), 237-250.
- Mehta, R. K., & Agnew, M. J. (2011). Effects of concurrent physical and mental demands for a short duration static task. *International Journal of Industrial Ergonomics*, 41(5), 488-493.
- Miyake, S. (2001). Multivariate workload evaluation combining physiological and subjective measures. *International Journal of Psychophysiology*, 40(3), 233-238.
- Miyake, S., Yamada, S., Shoji, T., Takae, Y., Kuge, N., & Yamamura, T. (2009). Physiological responses to workload change. A test/retest examination. *Applied Ergonomics*, 40(6), 987-996.
- Molteni, E., Butti, M., Bianchi, A. M., & Reni, G. (2008, August). Activation of the prefrontal cortex during a visual n-back working memory task with varying memory load: a near infrared spectroscopy study. *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 4024-4027). IEEE.
- Moroney, W. F., Biers, D. W., Eggemeier, F. T., & Mitchell, J. A. (1992). A comparison of two scoring procedures with the NASA task load index in a simulated flight task. *Proceedings of the IEEE 1992 National Aerospace and Electronics Conference* (pp. 734-740). IEEE.
- Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11), 1905-1922.

Mulder, L. J. M., De Waard, D. D., & Brookhuis, K. A. (2004). Estimating mental effort using heart rate and heart rate variability. In N. Stanton, A. Hedge, H.W. Hendrick, K.A. Brookhuis, E. Salas (Eds.), *Handbook of Ergonomics and Human Factors Methods* (pp. 201-208). Taylor & Francis.

Nino, V., Marchak, F., & Claudio, D. (2019). Association between perceived workload and adverse body posture. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care: Vol. 8(1)*, (pp. 182-186). SAGE Publications.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive psychology*, 7(1), 44-64.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, 33(1), 17-33.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 30(3), 286-297.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, 2(2), 140-160.

Parkinson, M. B., & Reed, M. P. (2010). Creating virtual user populations by analysis of anthropometric data. *International Journal of Industrial Ergonomics*, 40(1), 106-111.

- Paul, G., & Wischniewski, S. (2012). Standardisation of digital human models. *Ergonomics*, 55(9), 1115-1118.
- Perry, C. M., Sheik-Nainar, M. A., Segall, N., Ma, R., & Kaber, D. B. (2008). Effects of physical workload on cognitive task performance and situation awareness. *Theoretical Issues in Ergonomics Science*, 9(2), 95-113.
- Peruzzini, M., Grandi, F., & Pellicciari, M. (2020). Exploring the potential of Operator 4.0 interface and monitoring. *Computers & Industrial Engineering*, 139, 105600.
- Pheasant, S. & Haslegrave, C.M. (2006) *Bodyspace: Anthropometry, Ergonomics, and the Design of Work*. CRC Press.
- Pillastrini, P., Mugnai, R., Farneti, C., Bertozzi, L., Bonfiglioli, R., Curti, S., Mattioli, S., & Violante, F. S. (2007). Evaluation of two preventive interventions for reducing musculoskeletal complaints in operators of video display terminals. *Physical Therapy*, 87(5), 536-544.
- Punnett, L. & Wegman, D.H. (2004). Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. *Journal of Electromyography and Kinesiology*, 14(1), 13–23.
- Quertemont, E. (2011). How to statistically show the absence of an effect. *Psychologica Belgica*, 51(2), 109-127.
- Ranganathan, P., Pramesh, C. S., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Measures of agreement. *Perspectives in Clinical Research*, 8(4), 187.
- Ravenell, R. A., Camasta, C. A., & Powell, D. R. (2011). The Unreliability of the Intermetatarsal Angle in Choosing a Hallux Abducto Valgus Surgical Procedure. *The Journal of Foot & Ankle Surgery*, 50(3), 287 – 292.

Recarte, M. A., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied*, 6(1), 31.

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied*, 9(2), 119.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Advances in Psychology*. 52, 185-218.

Reyterou, E. (2009). *Development of an Accessibility Prediction Methodology*. [Unpublished Master's dissertation]. Cranfield University.

Robert-Lachaine, X., Mecheri, H., Larue, C., & Plamondon, A. (2017). Validation of inertial measurement units with an optoelectronic system for whole-body motion analysis. *Medical & Biological Engineering & Computing*, 55(4), 609-619.

Rodrick, D., Karwowski, W. & Marras, W. S. (2012). Work-Related Upper Extremity Musculoskeletal Disorders. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics* (4th ed., pp. 826–867). John Wiley & Sons, Inc.

Roetenberg, D., Baten, C. T., & Veltink, P. H. (2007). Estimating body segment orientation by applying inertial and magnetic sensing near ferromagnetic materials. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15(3), 469-471.

Roman-Liu, D. (2014). Comparison of concepts in easy-to-use methods for MSD risk assessment. *Applied Ergonomics*, 45, 420 – 427.

Romero, D., Stahre, J., Wuest, T., Noran, O., Bernus, P., Fast-Berglund, Å., & Gorecky, D. (2016, October). Towards an operator 4.0 typology: a human-centric

perspective on the fourth industrial revolution technologies. In *Proceedings of The International Conference On Computers And Industrial Engineering (CIE46), Tianjin, China* (pp. 29-31).

Rubio, S., Díaz, E., Martín, J., & Puente J. M. (2004). Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology: An International Review*, 53(1), 61–86.

Salmon, P., Stanton, N., Walker, G., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics*, 37(2), 225-238.

Schall Jr, M. C., Fethke, N. B., Chen, H., Oyama, S., & Douphrate, D. I. (2016). Accuracy and repeatability of an inertial measurement unit system for field-based occupational studies. *Ergonomics*, 59(4), 591-602.

Schmitz, A., Ye, M., Shapiro, R., Yang, R., & Noehren, B. (2014). Accuracy and repeatability of joint angles measured using a single camera markerless motion capture system. *Journal of Biomechanics*, 47(2), 587-591.

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5(258), 1-17.

Shaikh, S., Cobb, S. V. G., Golightly, D., Segal, J. I., & Haslegrave, C. M. (2012a). Investigating the effects of physical and cognitive demands on the quality of performance under different pacing levels. *Work*, 41(1), 1625-1631.

Shaikh, S., Cobb, S. V. G., Golightly, D., Segal, J. I., & Haslegrave, C. M. (2012b). Effects of TAKT time on physical and cognitive demands in a mixed model assembly line and a single model assembly line. In *Contemporary Ergonomics and Human Factors 2012: Proceedings of the international conference on Ergonomics & Human Factors* (p. 137-144). CRC Press.

Smith, A. (1996). Upper-limb disorders – time to relax?. *Physiotherapy*, 82 (1), 31 – 38.

Stokes, C. K., Lyons, J. B., Littlejohn, K., Natarian, J., Case, E., & Speranza, N. (2010). Accounting for the human in cyberspace: Effects of mood on trust in automation. In *2010 International Symposium on Collaborative Technologies and Systems* (pp. 180-187). IEEE.

Sundin, A., & Örtengren, R. (2006). Digital Human Modeling for CAE Applications. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics* (3rd ed., pp. 1053–1078). John Wiley & Sons, Inc.

Svebak, S. (1993). The development of the Tension and Effort Stress Inventory (TESI). In J. H. Kerr, S. J. Murgatroyd, & M. J. Apter (Eds.), *Advances In Reversal Theory* (p. 189–204). Swets & Zeitlinger Publishers.

Synertial. (2011). *IGS-180-I Basic User's Guide v1.0*

Synertial. (2012). *IGS Inertial Motion Capture System for Biomechanical Applications*.

Tattersall, A. J. (2000). Practical guidelines for workload assessment. In J.A Wise & V.D.Hopkin (Eds.), *Human Factors in Certification* (pp. 223- 235). CRC Press

Tattersall, A.J., & Hockey, G.R., (1995). Level of operator control and changes in heart rate variability during simulated flight maintenance. *Human Factors*, 37(4), 682-698.

Taylor, R.M. (1990). Situational awareness rating technique (SART): the development of a tool for aircrew systems design. *Situational Awareness in Aerospace Operations (AGARD-CP-478)*. NATO-AGARD. pp. 3/1 –3/17.

Tenney, Y. Y., Diller, D. E., Deutsch, S., & Godfrey, K. (2006). The AMBR experiments: methodology and human benchmark results. In Gluck, K. A., & Pew,

R. W. (Eds.). *Modeling Human Behavior with Integrated Cognitive Architectures: Comparison, Evaluation, and Validation*. Lawrence Erlbaum.

Thies, S. B., Tresadern, P., Kenney, L., Howard, D., Goulermas, J. Y., Smith, C., & Rigby, J. (2007). Comparison of linear accelerations from three measurement systems during “reach & grasp”. *Medical Engineering and Physics*, 29(9), 967- 972.

Tomlinson, E. C., & Meyer, R. C. (2009). The role of causal attribution dimensions in trust repair. *Academy of Management Review*, 34(1), 85-104.

Tsang, P. S., & Vidulich, M. A. (1994). The roles of immediacy and redundancy relative subjective workload assessment. *Human Factors*, 36(3), 503-513.

Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics* (3rd ed., pp. 243–268). John Wiley & Sons, Inc.

Turk, M., Resman, M., & Herakovič, N. (2021). The impact of smart technologies: A case study on the efficiency of the manual assembly process. *Procedia CIRP*, 97, 412-417.

Van Acker, B. B., Parmentier, D. D., Vlerick, P., & Saldien, J. (2018). Understanding mental workload: from a clarifying concept analysis toward an implementable framework. *Cognition, Technology & Work*, 20(3), 351-365.

Van Nieuwenhuyse, A., Somville, P. R., Crombez, G., Burdorf, A., Verbeke, G., Johannik, K., Van den Bergh, O., Masschelein, R., Mairiaux, P. & Moens, G. F. (2006). The role of physical workload and pain related fear in the development of low back pain in young workers: evidence from the BelCoBack Study; results after one year of follow up. *Occupational and Environmental Medicine*, 63(1), 45- 52.

- Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, 41(5), 656-669.
- Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied*, 2(3), 270-285.
- Vidulich, M. A., & Tsang, P. S. (2012) Mental workload and situation awareness. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics* (4th ed., pp. 243–268). John Wiley & Sons, Inc.
- Wang, L., & Lau, H.Y.K. (2013). Digital Human Modeling for Physiological Factors Evaluation in Work System Design. In V.G. Duffy (Eds.), *Digital Human Modeling and Applications in Health, Safety, Ergonomics, and Risk Management* (pp. 134-142). Springer.
- Waters, T. R. (2004). National efforts to identify research issues related to prevention of work-related musculoskeletal disorders. *Journal of Electromyography and Kinesiology*, 14(1), 7-12.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070.
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 159-177.
- Wilson, G. F. (1992). Applied use of cardiac and respiration measures: Practical considerations and precautions. *Biological Psychology*, 34(2-3), 163-178.
- Wilson, G. F., Fullenkamp, P., & Davis, I. (1994). Evoked potential, cardiac, blink, and respiration measures of pilot workload in air-to-ground missions. *Aviation, Space, and Environmental Medicine*, 65, 100-105.

- Winkel, J., & Westgaard R. (1992). Occupational and individual risk factors for shoulder-neck complaints: Part II. The scientific basis (literature review) for the guide. *International Journal of Industrial Ergonomics*, 10(1-2), 85–104.
- Wogalter, M. S., Laughery, K. R., & Mayhorn, C. B. (2012). Warnings and Hazard Communications. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics* (4th ed., pp. 1053–1078). John Wiley & Sons, Inc.
- Wong, C., Zhang, Z. Q., Lo, B., & Yang, G. Z. (2015). Wearable sensing for solid biomechanics: A review. *IEEE Sensors Journal*, 15(5), 2747-2760.
- Xu, X., & McGorry, R. W. (2015). The validity of the first and second generation Microsoft Kinect™ for identifying joint centre locations during static postures. *Applied Ergonomics*, 49, 47-54.
- Yang, J., Abdel-Malek, K., Farrell, K. (2004). The IOWA Interactive Digital Human Modeling Virtual Environment. *Proceedings of IMECE 2004, ASME International Engineering Congress, The 3rd Symposium on Virtual Manufacturing and Application*, 1059–1067.
- Yang, Y., McDonald, M., & Zheng, P. (2012). Can drivers' eye movements be used to monitor their performance? A case study. *IET Intelligent Transport Systems*, 6(4), 444-452.
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. In E. E. Boe & R. M. Church (Eds), *Punishment: Issues and Experiments*, (pp. 27-41). Appleton-Century-Crofts.
- Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics*, 58(1), 1-17.

APPENDICES

Appendix A BVH Reference Table

International Society of Bio-mechanics JCS							Axis Perpendicular to Saggital			IMU (Gyro) Number	Axis Perpendicular to Frontal Plane		Axis Perpendicular to Transverse Plane (Longitudinal)			
Animazoo Kinematic Model JCS							(-) Z	(+) Z	(-) X		(+) X	(-) Y	(+) Y			
Animazoo Modified BVH Output JCS							(+) X	(-) X	(+) Y		(-) Y	(+) Z	(-) Z			
							(-) Z	(+) Z	(-) Y		(+) Y	(-) Z	(-) Y			
Segment Bearing IMU Sensor	Joint(s) Measured	Proximal Parent Segment	Joint origin	Animazoo Kinematic Model (AKM) Joint Measured	(AKM) Proximal Parent Segment	(AKM) Proximal Parent Joint	BVH Output Column				BVH Output Column		BVH Output Column			
Head	Head	Neck C1-C7	Top Center ofC1 (Atlas)	Head	Neck	Neck	Flexion	41	Extension	11	Lateral tilt LT	40	Lateral tilt RT	Axial Rot. Left	42	Axial Rot. Right
Lower Neck	Neck C3-C5	Spine T1-T8	Cervical C7	Neck	Spine	Spine1	Flexion	38	Extension	17	Lateral tilt LT	37	Lateral tilt RT	Axial Rot. Left	39	Axial Rot. Right
Left Clavicle* Right Clavicle*	Clavicle + Scapula **	Spine T1-T8	SternoClavicu lar Joint	L Shoulder R Shoulder	Spine Spine	Spine1 Spine1	Axial Rot. Fore Axial Rot. Fore	44 56	Axial Rot. Back Axial Rot. Back	6 12	Depression Elevation	43 55	Elevation Depression	Retraction Protraction	45 57	Protraction Retraction
Dorsal Middle Upper Arms	Glenomumerus (GH) joint	From SC to AA to AC ***	Glenohumeral rot. Center	L Arm R Arm	L Clavicle R Clavicle	L Shoulder R Shoulder	Extension Extension	47 59	Flexion Flexion	7 13	Adduction Abduction	46 58	Abduction Adduction	Outward Medial Rotation Inward Medial Rotation	48 60	Inward Medial Rotation Outward Medial Rotation
Dorsal, Distal 1/3rd of Lower Arms	Elbow	Humerus	Hinge Axes of EL & EM *****	L ForeArm R ForeArm	L Upper Arm R Upper Arm	L Arm R Arm	Extension Extension	50 62	Flexion Flexion	8 14	Not Used Anatomically		Supination (External Rotation) Pronation (Internal Rotation)		51 63	Pronation (Internal Rotation) Supination (External Rotation)
Dorsal Capitate of Hands	Wrist	Forearm	US ****	L Hand R Hand	L ForeArm R ForeArm	L Forearm R ForeArm	Ulnar Ulnar	53 65	Radial Radial	9 15	Flexion Extension	52 64	Extension Flexion	supination pronation	48 60	pronation supination
Thorax T1-T8	Spine T1-L5	Sacral	Sacral S1	Spine1	Lower Spine	Spine	Flexion	35	Extension	16	Lateral Flexion LT	34	Lateral Flexion RT	Rotation Left	36	Rotation Right
Lumbar L1-L5	IMU to L5*****	Sacral	Sacral S1	Spine	Pelvis	Hips	Flexion	32	Extension	18	Lateral Flexion LT	31	Lateral Flexion RT	Rotation Left	33	Rotation Right
Pelvis	Pelvis/Sacral	None	None	Hips (Root)	Global 0,0,0	Global 0,0,0	Flexion	5	Extension	10	Lateral Flexion LT	4	Lateral Flexion RT	Rotation Left	6	Rotation Right
Dorsal Middle Upper Legs/Thighs	L Hip R Hip	Pelvis	L&R Hip Joint Center of Rot	L UpperLeg R UpperLeg	L Pelvis R Pelvis	Hips	Extension Extension	8 20	Flexion Flexion	0 3	Adduction Abduction	7 19	Abduction Adduction	External Rotation Internal Rotation	9 21	Internal Rotation External Rotation
Below Tibial Tuberosity Ventral	Knee	Femur	Inter-condylar point	L Leg R Leg	L Upper Leg R Upper Leg	L UpperLeg R UpperLeg	Flexion Flexion	11 23	Extension Extension	1 4	Not Used Anatomically		External Rotation Internal Rotation		12 24	Internal Rotation External Rotation
Left Foot Right Foot	Ankle	Tibia / fibula	inter-malleolar	L Foot R Foot	L LowerLeg R LowerLeg	L Leg R Leg	Plantar Flexion Plantar Flexion	14 26	Dorsi Flexion Dorsi Flexion	2 5	Inversion Eversion	13 25	Eversion Inversion	External Rotation Internal Rotation	15 27	External Rotation Internal Rotation

Appendix B REBA Assessment Worksheet

REBA Employee Assessment Worksheet

Based on Technical note: Rapid Entire Body Assessment (REBA); Hignett, McAtumney; Applied Ergonomics 31 (2000) 201-205

A. Neck, Trunk and Leg Analysis

Step 1: Locate Neck Position

 Step 1a: Adjust...
 If neck is twisted: +1
 If neck is side bending: +1
 Neck Score:

Step 2: Locate Trunk Position

 Step 2a: Adjust...
 If trunk is twisted: +1
 If trunk is side bending: +1
 Trunk Score:

Step 3: Legs

 Adjust: 30-60° Add +1, 60-90° Add +2
 Leg Score:

Step 4: Look-up Posture Score in Table A
 Using values from steps 1-3 above, locate score in Table A.
 Table A Score:

Step 5: Add Force/Load Score
 If load < 11 lbs: +0
 If load 11 to 22 lbs: +1
 If load > 22 lbs: +2
 Adjust: If shock or rapid build up of force: add +1
 Force/Load Score:

Step 6: Score A. Find Row in Table C
 Add values from steps 4 & 5 to obtain Score A.
 Find Row in Table C.
 Score A:

Scoring:
 1 = negligible risk
 2 or 3 = low risk, change may be needed
 4 to 7 = medium risk, further investigation, change soon
 8 to 10 = high risk, investigate and implement change
 11+ = very high risk, implement change

B. Arm and Wrist Analysis

Step 7: Locate Upper Arm Position:

 Step 7a: Adjust...
 If shoulder is raised: +1
 If upper arm is abducted: +1
 If arm is supported or person is leaning: -1
 Upper Arm Score:

Step 8: Locate Lower Arm Position:

 Lower Arm Score:

Step 9: Locate Wrist Position:

 Step 9a: Adjust...
 If wrist is bent from midline or twisted: Add +1
 Wrist Score:

Step 10: Look-up Posture Score in Table B
 Using values from steps 7-9 above, locate score in Table B.
 Posture Score B:

Step 11: Add Coupling Score
 Well fitting Handle and mid range power grip: good: +0
 Acceptable but not ideal hand hold or coupling acceptable with another body part: fair: +1
 Hand hold not acceptable but possible: poor: +2
 No handles, awkward, unsafe with any body part: unacceptable: +3
 Coupling Score:

Step 12: Score B. Find Column in Table C
 Add values from steps 10 & 11 to obtain Score B.
 Find column in Table C and match with Score A in row from step 6 to obtain Table C Score.
 Score B:

Step 13: Activity Score
 +1 1 or more body parts are held for longer than 1 minute (static)
 +1 Repeated small range actions (more than 4x per minute)
 +1 Action causes rapid large range changes in postures or unstable base
 Activity Score:

Table C Score	+	Activity Score
Final REBA Score		

Task name: _____ Reviewer: _____ Date: ____/____/____
 This tool is provided without warranty. The author has provided this tool as a simple means for applying the concepts provided in REBA. © 2004 Waters Consulting, Inc. provided by Practical Ergonomics
 r.barker@engosmart.com (816) 444-1667

Appendix C Investigating the Effectiveness of Motion Capture Technology

C.1 Cranfield University Research Ethics System Approval Email

Johnson, Teegan

From: SEREC
Sent: 14 November 2013 09:22
To: Johnson, Teegan
Cc: Fletcher, Sarah
Subject: Ethics Proposal 108-2013

Dear Teegan

Your proposed research activity "A computer software method for Ergonomic analysis utilising Non-optical Motion capture" has been reviewed by SEREC and confirmed as posing a low risk in terms of research ethics.

You can now proceed with the research activities you have sought approval for and we wish you a successful project.

Best wishes

Sue

Sue Garrod
Secretary
Centre for Advanced Systems
Aerospace Engineering Division
School of Engineering
01234 754165

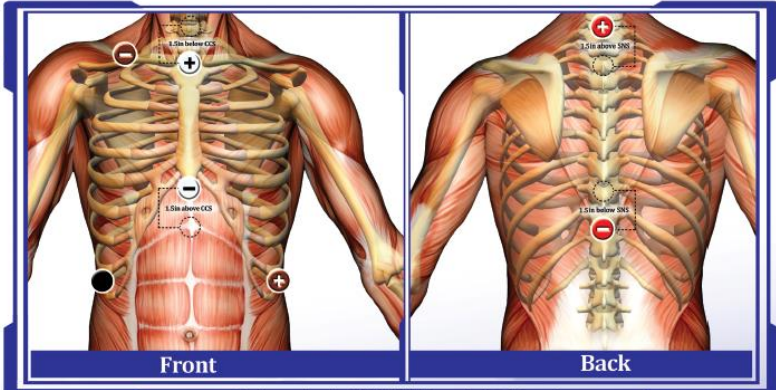
This e-mail and any attachments to it may be confidential and are intended only for the named addressee. If you are not the named addressee, please accept our apology, notify the sender immediately and then delete the e-mail. We request that you do not disclose, use, copy or distribute any information within it.

Any opinions expressed are not necessarily the corporate view of Cranfield University. This e-mail is not intended to be contractually binding unless specifically stated and the sender is an authorised University signatory.

Whilst we have taken steps to ensure that this e-mail and all attachments are free from any virus, we advise that, in keeping with good computing practice, the recipient should ensure they are actually virus free.

Appendix D Predictive Cognitive and Physical Interaction

D.1 Electrode Placement MindWare Poster



Front

Back

Electrode Placement
ECG, Impedance Cardiography, GSR

Recommended Practices

Before placing electrodes, clean areas (except GSR) with rubbing alcohol. Use water in place of alcohol for GSR site cleaning. For comfort, attach snap leads to electrode before placing on the subject.

- Place on the bottom left rib near the side
- Place on the right collar bone (clavicle)
- Place on the bottom right rib on the subjects side (Is not required when used with Impedance Cardiography)
- Place on the spine, 1.5 inches (3.81cm) above the SNS electrode
- Place on the spine, 1.5 inches (3.81cm) below the SNS electrode
- Similar to the Qu configuration, place on the jugular notch just above where the collar bones meet. If the subject is speaking during a study, offset the electrode to the left or right to prevent artifact
- Similar to the Qu configuration, place just below the sternum (on xiphoid process). Make sure any straps or bands do not interfere with the electrodes
- Place high on the subject's palms (thenar and hypothenar eminence) or on middle and pointer digits (distal & medial phalanges). Polarity is not a factor. Do not prep area with alcohol

To order more electrodes call: (888) 765-9735 or email: sales@mindwaretech.com

Secure lead wire using medical tape to prevent noise caused by movement. Significant slack should be left between electrode and tape

Configuration 1

Configuration 2

ECG	+	-
Ground	+	-
CCS	+	-
SNS	+	-
GSC	+	-

MindWare
TECHNOLOGIES LTD.
WWW.MINDWARETECH.COM

D.2 Brief, Consent and Data Collection Sheets

Brief

This experiment is part of a PhD thesis being completed at Cranfield University within the Industrial Psychology and Human Factors (IPHF) group. The aim of the experiment is to investigate the combined parameters of cognitive workload and physical strain during assembly task performance.

The procedure for this experiment will involve you wearing a heart rate monitor and a motion capture suit. Your heart rate and the postures you adopt during the experiment will be recorded during the experiment. The heart rate monitor requires 5 ECG sensors to be placed on your skin, one on your collarbone and two on your lowest ribs on each side of your body. Two further sensors will be placed on the palm of your non-dominant hand. The motion capture suit will be worn over your clothing. Two cameras will additionally record the postures you adopt while completing the tasks.

You will be asked to complete two tasks, between three and four times. One task will involve you colouring in circles using a white board marker; the second task will involve you installing piping. Each task takes 2 minutes at the end of each task you will complete a survey that will gather information about your experience.

This study is conducted in accordance with the General Data Protection Regulation 2018 and the data you will provide will be stored and maintained by the IPHF group with full confidentiality. The IPHF group will be the only personnel with access to the data you provide in this study, which will be stored securely.

The consent form requires your name only for initial identification until your data is processed. This means that you can withdraw your data from the study up to the point at which you complete the study, without giving us any reason, but after this time it will not be possible as the data will be fully anonymised *and only identified by code number*. Individual data will be reported in aggregate form so that no single participant is identifiable and if any individual quotes are used they will be made completely anonymous and unidentifiable to any individual. You will be identifiable in the video recordings and it will be possible to withdraw this data up to 7 days after you have completed the experiment. Clips may be taken from these recordings, if they are your face will be obscured and thereby anonymising the clip.

If you have any questions please do not hesitate to contact me:
t.l.johnson@cranfield.ac.uk

Thank you for your participation.

Consent Form

I understand what has been asked of me and I agree to take part in the study. I understand that I can withdraw and ask for my data to be removed up to the point of submission, after which time it will be completely anonymised. I agree to the publication of my responses in anonymised form so that they cannot be traced back to

Printed

Name:

Signature:

Date:

Biographical information

Participant number: ____

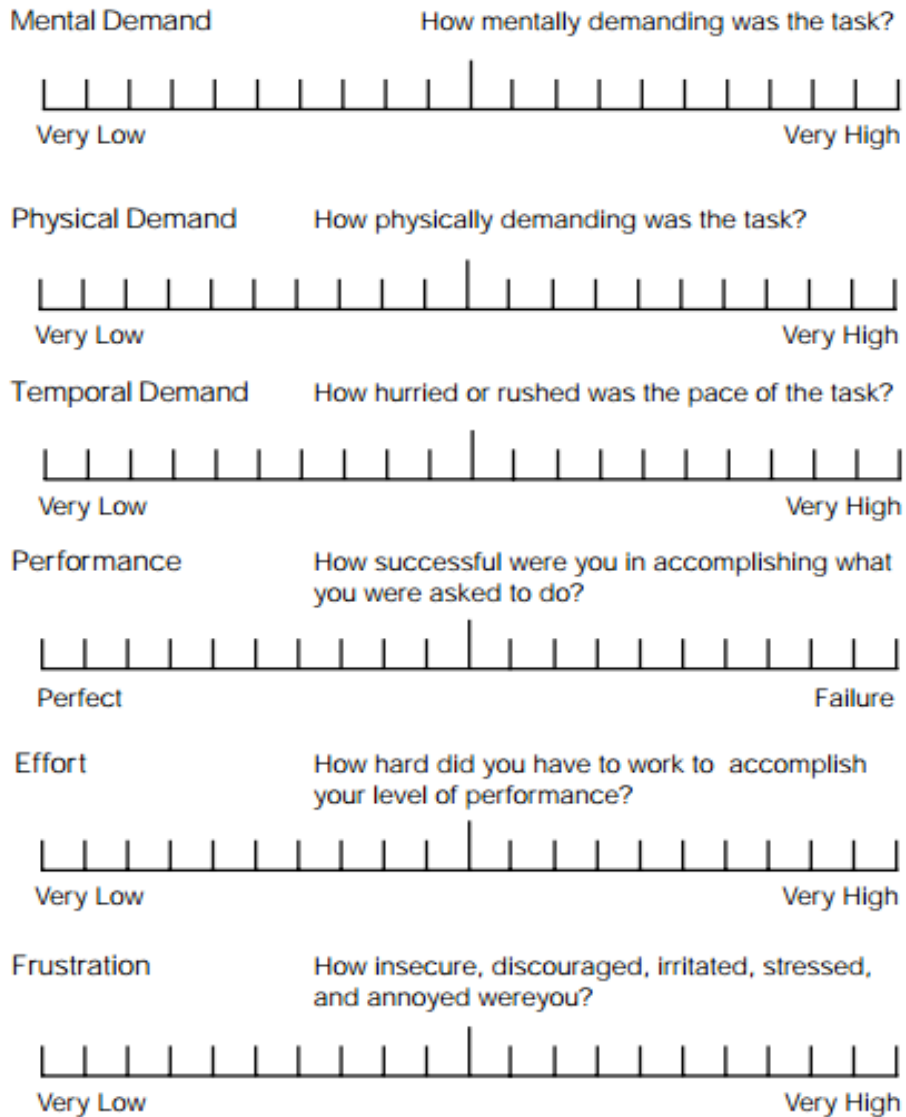
Age: _____

Gender: _____

Height: _____

Circles: OH stood

Thinking about the task you have just completed please consider each question carefully and answer by placing a cross on the scales.



Circles: OH sat

Thinking about the task you have just completed please consider each question carefully and answer by placing a cross on the scales.

Mental Demand	How mentally demanding was the task?
	
Very Low	Very High
Physical Demand	How physically demanding was the task?
	
Very Low	Very High
Temporal Demand	How hurried or rushed was the pace of the task?
	
Very Low	Very High
Performance	How successful were you in accomplishing what you were asked to do?
	
Perfect	Failure
Effort	How hard did you have to work to accomplish your level of performance?
	
Very Low	Very High
Frustration	How insecure, discouraged, irritated, stressed, and annoyed were you?
	
Very Low	Very High

Circles: 45°

Thinking about the task you have just completed please consider each question carefully and answer by placing a cross on the scales.

Mental Demand	How mentally demanding was the task?
	
Very Low	Very High
Physical Demand	How physically demanding was the task?
	
Very Low	Very High
Temporal Demand	How hurried or rushed was the pace of the task?
	
Very Low	Very High
Performance	How successful were you in accomplishing what you were asked to do?
	
Perfect	Failure
Effort	How hard did you have to work to accomplish your level of performance?
	
Very Low	Very High
Frustration	How insecure, discouraged, irritated, stressed, and annoyed were you?
	
Very Low	Very High

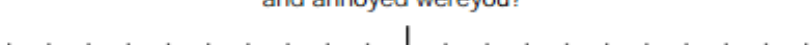
Ducts: OH stood

Thinking about the task you have just completed please consider each question carefully and answer by placing a cross on the scales.

Mental Demand	How mentally demanding was the task?
	
Very Low	Very High
Physical Demand	How physically demanding was the task?
	
Very Low	Very High
Temporal Demand	How hurried or rushed was the pace of the task?
	
Very Low	Very High
Performance	How successful were you in accomplishing what you were asked to do?
	
Perfect	Failure
Effort	How hard did you have to work to accomplish your level of performance?
	
Very Low	Very High
Frustration	How insecure, discouraged, irritated, stressed, and annoyed were you?
	
Very Low	Very High

Ducts: OH sat

Thinking about the task you have just completed please consider each question carefully and answer by placing a cross on the scales.

Mental Demand	How mentally demanding was the task?
	
Very Low	Very High
Physical Demand	How physically demanding was the task?
	
Very Low	Very High
Temporal Demand	How hurried or rushed was the pace of the task?
	
Very Low	Very High
Performance	How successful were you in accomplishing what you were asked to do?
	
Perfect	Failure
Effort	How hard did you have to work to accomplish your level of performance?
	
Very Low	Very High
Frustration	How insecure, discouraged, irritated, stressed, and annoyed were you?
	
Very Low	Very High

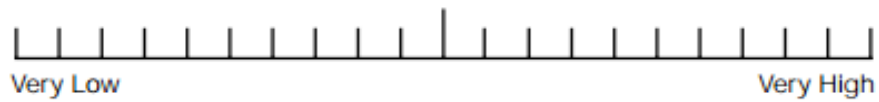
Ducts: 45°

Thinking about the task you have just completed please consider each question carefully and answer by placing a cross on the scales.

How mentally demanding was the task?



How physically demanding was the task?



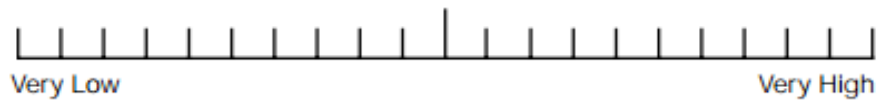
How hurried or rushed was the pace of the task?



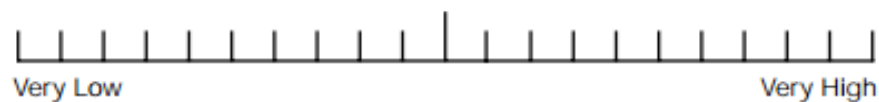
How successful were you in accomplishing what you were asked to do?



How hard did you have to work to accomplish your level of performance?

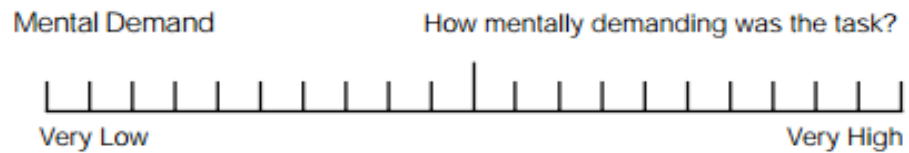
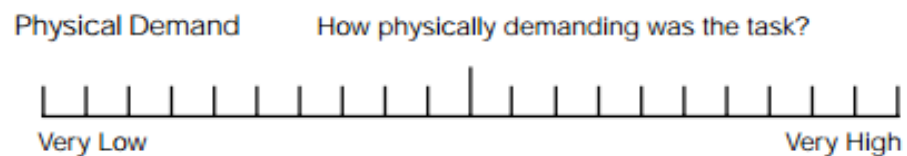
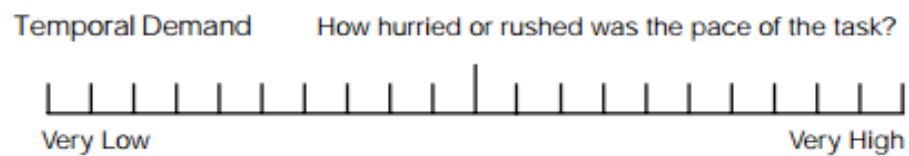
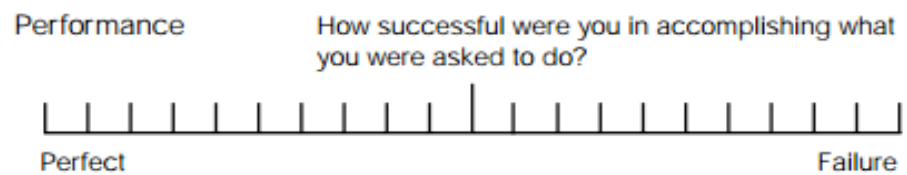
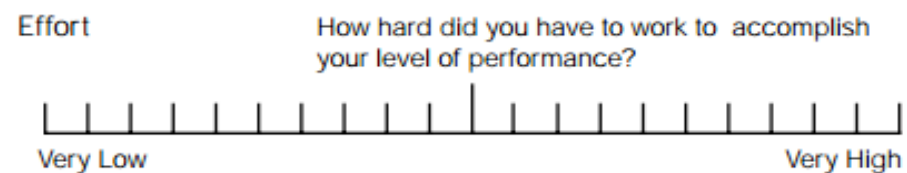
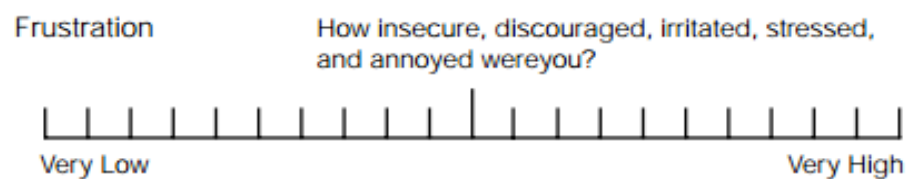


How insecure, discouraged, irritated, stressed,
and annoyed were you?



Ducts: Vertical

Thinking about the task you have just completed please consider each question carefully and answer by placing a cross on the scales.



D.3 Cranfield University Research Ethics System Approval Email

Johnson, Teegan

From: donotreply@infonetica.net
Sent: 02 May 2018 14:16
To: Johnson, Teegan
Cc: Fletcher, Sarah
Subject: CURES Submission: Approved

Dear Teegan

Reference: CURES/4999/2018

Title: Cognitive Workload and Physical Strain In Assembly Task Performance

Thank you for your application to the Cranfield University Research Ethics System (CURES).

Your proposed research activity has been confirmed as Level 2b risk in terms of research ethics. You may now proceed with the research activities you have sought approval for.

Please remember that CURES occasionally conducts audits of projects. We may therefore contact you during or following execution of your fieldwork. Guidance on good practice is available on the [research ethics intranet pages](#).

If you have any queries, please contact cures-support@cranfield.ac.uk

We wish you every success with your project.

Regards

CURES Team

May we remind you of the importance of addressing health and safety issues in your research. Templates and further guidance are available [here](#).

Appendix E Development and Integration of a Human Robot Collaboration Trust Speed Equation into DHM

E.1 Development of a Human Robot Collaboration Trust Speed Equation: Brief, Consent and Data Collection sheets

Brief

This experiment is part of an EPSRC funded project being completed by the Industrial Psychology and Human Factors (IPHF) group. The aim of the experiment is to investigate how people respond to working directly with an industrial robot.

The procedure for this experiment will involve you completing a task that requires you to attach a fitting to the end of a pipe when it is presented to you by the robot. After each task you will be asked to complete two short surveys. You will complete the same fitting activity five times with the robot presenting the pipe to you at different speeds. Prior to completing the task you will be given a full explanation and walk through of what will be expected of you. After the study you will be debriefed regarding the full nature of the experiment.

This study was approved by the Cranfield University Research Ethics Committee and is conducted in accordance with the Data Protection Act 2018. The data you will provide will be stored and maintained by the IPHF group securely and with full confidentiality. The IPHF group will be the only personnel with access to the data you provide in this study.

You can withdraw from the study at any point without giving us any reason, but you will be unable to withdraw after submission as your data will be fully anonymised. Individual data and results will only be reported in aggregate or anonymised form so that no single participant is identifiable.

If you have any questions at a later date please do not hesitate to contact me: t.l.johnson@cranfield.ac.uk

Thank you for your participation.

Consent Form

I understand what has been asked of me and I agree to take part in the study. I understand that I can withdraw and ask for my data to be removed up to the point of submission, after which time it will be completely anonymised. I agree to the publication of my responses in anonymised form so that they cannot be traced back to

Printed Name:

Signature:

Date:

Biographical information

Participant number: ____

Age: _____

Gender: _____

Pre-Test

There is a list of words which describe people's moods or feelings. Please indicate how well each word describes how you feel **AT THE MOMENT**. For each word, circle the answer from 1 to 4 which best describes your mood.

	Definitely	Slightly	Slightly Not	Definitely Not
1. Happy	1	2	3	4
2. Dissatisfied	1	2	3	4
3. Energetic	1	2	3	4
4. Relaxed	1	2	3	4
5. Alert	1	2	3	4
6. Nervous	1	2	3	4
7. Passive	1	2	3	4
8. Cheerful	1	2	3	4
9. Tense	1	2	3	4
10. Jittery	1	2	3	4
11. Sluggish	1	2	3	4
12. Sorry	1	2	3	4
13. Composed	1	2	3	4
14. Depressed	1	2	3	4
15. Restful	1	2	3	4
16. Vigorous	1	2	3	4
17. Anxious	1	2	3	4
18. Satisfied	1	2	3	4
19. Unenterprising	1	2	3	4
20. Sad	1	2	3	4
21. Calm	1	2	3	4
22. Active	1	2	3	4
23. Contented	1	2	3	4
24. Tired	1	2	3	4
25. Impatient	1	2	3	4
26. Annoyed	1	2	3	4
27. Angry	1	2	3	4
28. Irritated	1	2	3	4
29. Grouchy	1	2	3	4

Task 1 - 5

Please read each statement carefully and then check (x) the preferred box:

The way the robot moved made me uncomfortable				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

I felt I could rely on the robot to do what it was supposed to do				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

The speed at which the gripper picked up and released the components made me uneasy				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

I felt safe interacting with the robot				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

I knew the gripper would not drop the components				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

The size of the robot did not intimidate me				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

The robot gripper did not look reliable				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

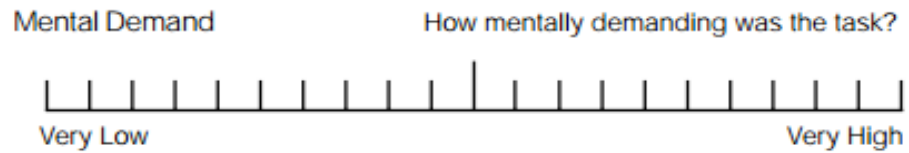
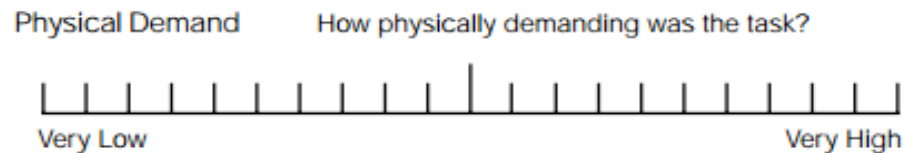
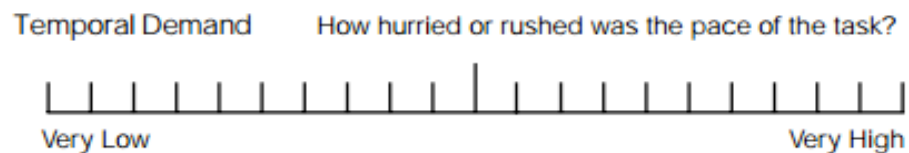
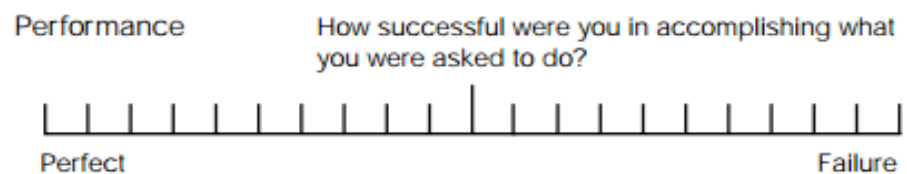
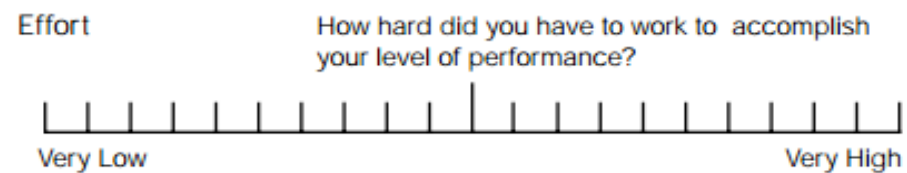
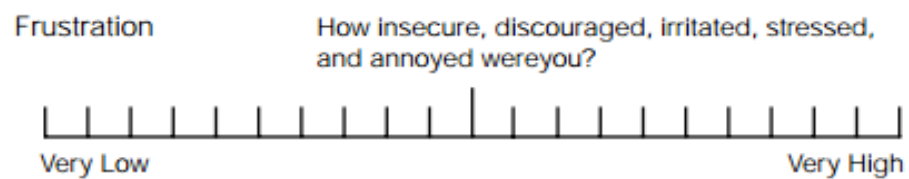
I was comfortable the robot would not hurt me				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

I trusted that the robot was safe to cooperate with				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

The gripper seemed like it could be trusted				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

The speed at which the robot moved made me uneasy				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Thinking about the task you have just complete please consider each question carefully and answer by placing a cross on the scales.



There is a list of words which describe people's moods or feelings. Please indicate how well each word describes how you feel **AT THE MOMENT**. For each word, circle the answer from 1 to 4 which best describes your mood.

	Definitely	Slightly	Slightly Not	Definitely Not
1. Happy	1	2	3	4
2. Dissatisfied	1	2	3	4
3. Energetic	1	2	3	4
4. Relaxed	1	2	3	4
5. Alert	1	2	3	4
6. Nervous	1	2	3	4
7. Passive	1	2	3	4
8. Cheerful	1	2	3	4
9. Tense	1	2	3	4
10. Jittery	1	2	3	4
11. Sluggish	1	2	3	4
12. Sorry	1	2	3	4
13. Composed	1	2	3	4
14. Depressed	1	2	3	4
15. Restful	1	2	3	4
16. Vigorous	1	2	3	4
17. Anxious	1	2	3	4
18. Satisfied	1	2	3	4
19. Unenterprising	1	2	3	4
20. Sad	1	2	3	4
21. Calm	1	2	3	4
22. Active	1	2	3	4
23. Contented	1	2	3	4
24. Tired	1	2	3	4
25. Impatient	1	2	3	4
26. Annoyed	1	2	3	4
27. Angry	1	2	3	4
28. Irritated	1	2	3	4

E.2 Assessment of the CAD Integrated Human Robot Collaboration Trust Speed Equation: Brief, Consent and Data Collection sheets

Brief

This experiment is part of an EPSRC funded project being completed by the Industrial Psychology and Human Factors (IPHF) group. The aim of the experiment is to investigate how people respond to working directly with an industrial robot.

The procedure for this experiment will involve you completing a task that requires you to attach a component, which has been moved into place and held by a robot, to a frame. After you have completed the task you will be asked to complete three short surveys. Prior to completing the task you will be given a full explanation and walk through of what will be expected of you. After the study you will be debriefed regarding the full nature of the experiment.

This study was approved by the Cranfield University Research Ethics Committee and is conducted in accordance with the Data Protection Act 2018. The data you will provide will be stored and maintained by the IPHF group securely and with full confidentiality. The IPHF group will be the only personnel with access to the data you provide in this study.

You can withdraw from the study at any point without giving us any reason, but you will be unable to withdraw after submission as your data will be fully anonymised. Individual data and results will only be reported in aggregate or anonymised form so that no single participant is identifiable.

If you have any questions at a later date please do not hesitate to contact me:
t.l.johnson@cranfield.ac.uk

Thank you for your participation.

Consent Form

I understand what has been asked of me and I agree to take part in the study. I understand that I can withdraw and ask for my data to be removed up to the point of submission, after which time it will be completely anonymised. I agree to the publication of my responses in anonymised form so that they cannot be traced back to

Printed Name:

Signature:

Date:

Biographical information

Participant number: ____

Age: _____

Gender: _____

Please read each statement carefully and then check (x) the preferred box:

The way the robot moved made me uncomfortable				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

I felt I could rely on the robot to do what it was supposed to do				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

The speed at which the gripper picked up and released the components made me uneasy				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

I felt safe interacting with the robot				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

I knew the gripper would not drop the components				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

The size of the robot did not intimidate me				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

The robot gripper did not look reliable				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

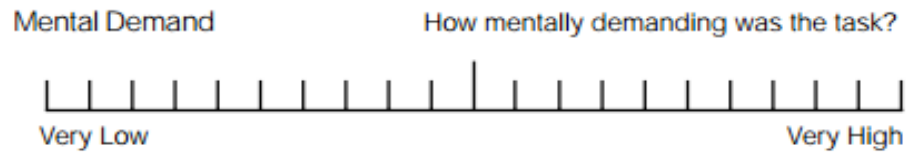
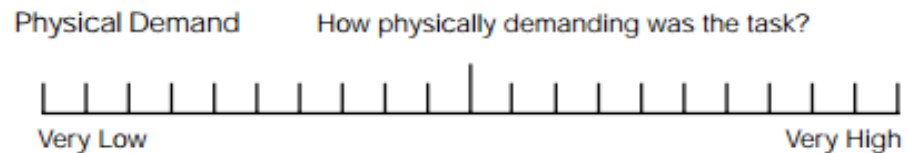
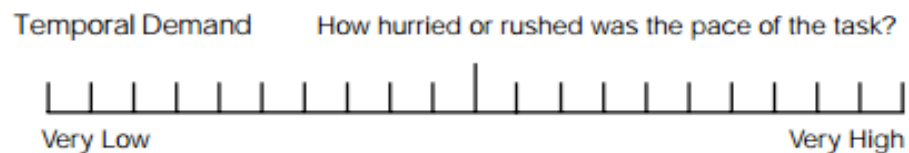
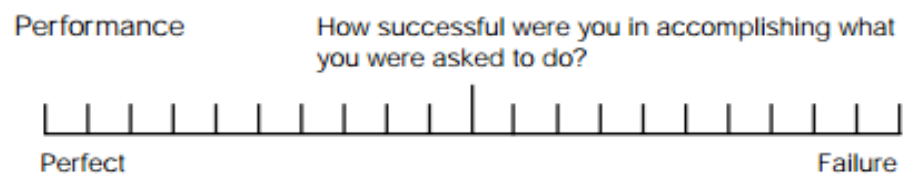
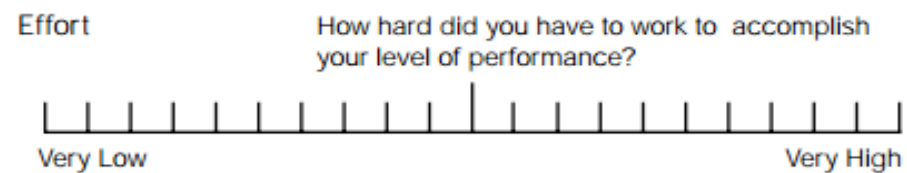
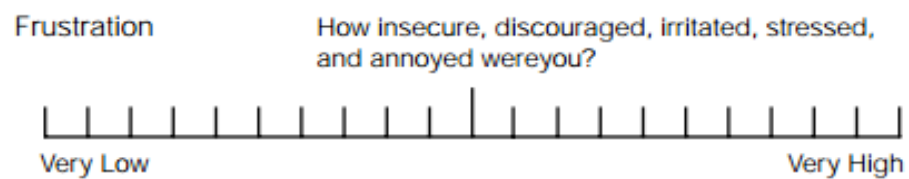
I was comfortable the robot would not hurt me				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

I trusted that the robot was safe to cooperate with				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

The gripper seemed like it could be trusted				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

The speed at which the robot moved made me uneasy				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Thinking about the task you have just complete please consider each question carefully and answer by placing a cross on the scales.



There is a list of words which describe people's moods or feelings. Please indicate how well each word describes how you feel **AT THE MOMENT**. For each word, circle the answer from 1 to 4 which best describes your mood.

	Definitely	Slightly	Slightly Not	Definitely Not
1. Happy	1	2	3	4
2. Dissatisfied	1	2	3	4
3. Energetic	1	2	3	4
4. Relaxed	1	2	3	4
5. Alert	1	2	3	4
6. Nervous	1	2	3	4
7. Passive	1	2	3	4
8. Cheerful	1	2	3	4
9. Tense	1	2	3	4
10. Jittery	1	2	3	4
11. Sluggish	1	2	3	4
12. Sorry	1	2	3	4
13. Composed	1	2	3	4
14. Depressed	1	2	3	4
15. Restful	1	2	3	4
16. Vigorous	1	2	3	4
17. Anxious	1	2	3	4
18. Satisfied	1	2	3	4
19. Unenterprising	1	2	3	4
20. Sad	1	2	3	4
21. Calm	1	2	3	4
22. Active	1	2	3	4
23. Contented	1	2	3	4
24. Tired	1	2	3	4
25. Impatient	1	2	3	4
26. Annoyed	1	2	3	4
27. Angry	1	2	3	4
28. Irritated	1	2	3	4

E.3 Cranfield University Research Ethics System Approval Email

Johnson, Teegan

From: donotreply@infonetica.net
Sent: 02 October 2018 10:45
To: Johnson, Teegan
Subject: CURES Submission: Approved

Dear Teegan

Reference: CURES/6445/2018
Title: Integrating Cognitions of Human Operators in digital Robot Design

Thank you for your application to the Cranfield University Research Ethics System (CURES).

Your proposed research activity has been confirmed as Level 1 risk in terms of research ethics. You may now proceed with the research activities you have sought approval for.

Please remember that CURES occasionally conducts audits of projects. We may therefore contact you during or following execution of your fieldwork. Guidance on good practice is available on the [research ethics intranet pages](#).

If you have any queries, please contact cures-support@cranfield.ac.uk

We wish you every success with your project.

Regards

CURES Team

May we remind you of the importance of addressing health and safety issues in your research. Templates and further guidance are available [here](#).