2015

Investigating the efficacy of different methodological approaches in the development of scientifically valid physical employment studies

Benjamin Lee-Bates

University of Wollongong

INVESTIGATING THE EFFICACY OF DIFFERENT METHODOLOGICAL APPROACHES IN THE DEVELOPMENT OF SCIENTIFICALLY VALID PHYSICAL EMPLOYMENT STANDARDS

A thesis submitted in fulfilment of the requirements for the awards of the degree

MASTER OF PHILOSOPHY (PSYCHOLOGY)

from

UNIVERSITY OF WOLLONGONG

by

BENJAMIN LEE-BATES, BACHELOR OF PSYCHOLOGY

FACULTY OF SOCIAL SCIENCES

2015

ABSTRACT

Current methodological frameworks for the development of a legally and scientifically defensible physical employment standard (PES), relies on the conduct of an efficacious job task analysis (JTA). However, as different JTA methodologies have received little examination in relation to their comparative accuracy and utility, no best practice methodology has been identified in the context of physically demanding occupations (PDO). Subjective methodologies in particular, including surveys and focus groups (FGs) may be underutilized in their ability to characterise physically demanding tasks and may provide a resource efficient alternative to 'gold standard' objective, observational methodologies. The primary aim of this thesis was therefore, to inform the identification of a JTA best practice methodology in the context of a PES by validating subjective methodologies and their ability to accurately describe physical job task parameters. This is accomplished in a series of three papers.

Chapter 2 (Paper 1) provides a systematic review of all existing JTA methodologies conducted within PDOs, and identifies common themes and methodological weakness relating to the psychometric properties of these studies. Results indicate that the majority of studies lacked explicit checks for reliability, validity and bias, highlighting the need for research into the comparative efficacy of different JTA methodologies. From these data a mixed method JTA is proposed that uses the comparative strengths and weakness of both subjective and objective methodologies.

Chapter 3 (Paper 2) provides further examination of subjective JTA methods through an examination of systematic bias that may be innate in these methods. This was accomplished through a survey conducted on a sample of Royal Australian Navy (RAN) volunteers which examined the effects of demographic and job profile characteristics on descriptions of physically demanding job tasks. Results showed no evidence of bias

resulting from participant characteristics; however self-serving bias may have been present in which participants that were more actively involved in a task had an inflated perception of that task's importance. These results have important implications for the identification of bias in commonly used JTA methods and the integration of subjective methods in the development of PESs.

Using data from the same population, Chapter 4 (Paper 3) provides direct comparison of three commonly used JTA techniques, surveys, FGs and task simulations, in their relative ability to accurately describe and rank tasks by their identifying characteristics. Overall, FGs showed a tendency to overestimate ratings of importance and physical effort, but were able to accurately predict vertical and horizontal distance when compared to task-simulation data. By comparison surveys were able to provide similar rankings and estimates of physical effort to task simulation data. From these results a three stage JTA methodology is recommended by which surveys then FGs are used to reduce the reliance on expensive physical demands analyses.

This thesis concludes with a summary of the key findings, consideration of research limitations, and discussion of implications for future research and recommended PES development practices, Chapter 5. Overall, the findings of this thesis address important gaps in literature and have the potential to make significant contributions to the field of organisational psychology and PES development by helping to identify methodological best practice.

ACKNOWLEDGEMENTS

I would like to acknowledge the help and support of my supervisors; Peter Caputi, Kane Middleton and Dan Billing. Their expertise and commitment to this project, has ensured that this thesis was conducted with the upmost scientific integrity. I have learned so much through working with them, and as a result feel much more confident in my ability to conduct effective research both independently and in a team work environment.

I would also like to acknowledge Greg Carstairs and Denise Linnane for their contributions across all stages of this project, particularly data collection, data analysis and revision. Their input has helped to ensure high quality research across all stages of its development, implementation and completion. This project would not have been possible without them.

CERTIFICATION

I, Benjamin Lee-Bates, declare that this thesis, submitted in fulfilment of the requirements for the award of Master of Philosophy, in the School of Psychology, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document is in the style has not been submitted for qualifications at any other academic institution and has been prepared in the journal article compilation style format.

Benjamin Lee-Bates

23 December 2015

PUBLICATIONS INCLUDED AS PART OF THESIS

Lee-Bates, B., Caputi, P., Middleton, K., & Billing, D. (2015). *A Systematic Review of Job Task Analysis within Physically Demanding Occupations: Is there Evidence for Best Practice?* Manuscript submitted for publication.

Lee-Bates, B., Billing, D., Caputi, P., Carstairs, G., Linnane, D., & Middleton, K., (2015). *Comparison of Subjective and Objective Job Task Analysis Methodologies for Physically Demanding Occupations*. Manuscript submitted for publication.

Lee-Bates, B., Billing, D., Caputi, P., Carstairs, G., Linnane, D., & Middleton, K., (2015). *Application of Subjective Job Task Analysis Methodologies in Physically Demanding Occupations: Evidence for the Presence of Self-Serving Bias.* Manuscript submitted for publication.

OTHER CONTRIBUTIONS ARISING FROM THESIS

Middleton, K., Carstairs, G., Billing, D., Carr, A., Lee-Bates, B., Caputi, P., & Linnane, D. (2014). *Development of Physical Employment Standards for the Royal Australian Navy: Validation of Identified Whole-of-ship Tasks*. Melbourne.

Lee-Bates, B., Caputi, P., Middleton, K., Billing, D., Carstairs, G., & Linnane, D. (August, 2015). *Application of Subjective Job Task Analysis Methodologies in Physically Demanding Occupations: Evidence for the Presence of Self-Serving Bias.* Oral presentation at the Second International Conference on Physical Employment Standards, Canmore, AB.

STATEMENT OF CONTRIBUTION OF OTHERS

This statement verifies that the greater part of the work in the previously stated publications/manuscripts are attributed to the candidate. Benjamin Lee-Bates, under the guidance and supervision of his supervisors, took primary responsibility for the design of each study, all data collection and analysis, prepared the first draft of each manuscript, and prepared the papers for submission to relevant journals. Co-authors, who were also supervisors to the candidate, contributed to the thesis by providing guidance on the design and structure of each study, and provided editorial suggestions for every paper.

Benjamin Lee-Bates, Masters Candidate

Associate Professor Peter Caputi, Primary Supervisor

Professor Frank Deane, Head of Postgraduate Studies

TABLE OF CONTENTS

ABSTRACT i
ACKNOWLEDGEMENTS iii
CERTIFICATION iv
PUBLICATIONS INCLUDED AS PART OF THESIS v
OTHER CONTRIBUTIONS ARISING FROM THESIS vi
STATEMENT OF CONTRIBUTION OF OTHERS vii
TABLE OF CONTENTS viii
KEY ABBREVIATIONS x
CHAPTER 1: CONTEXT STATEMENT 1
1.1 Preamble 1
1.2 Literature Review 1
1.3 Research Aims and Projected Outcomes 8
1.4 Thesis Structure 8
1.5 Significance and Originality 10
CHAPTER 2: A SYSTEMATIC REVIEW OF JOB TASK ANALYSIS WITHIN PHYSICALLY DEMANDING OCCUPATIONS: IS THERE EVIDENCE FOR BEST PRACTICE? 17
2.1 Abstract 17
2.2 Introduction 18
2.3 Method 22
2.4 Results 25
2.5 Discussion 29
2.6 Conclusion 36
CHAPTER 3: THE APPLICATION OF SUBJECTIVE JOB TASK ANALYSIS TECHNIQUES IN PHYSICALLY DEMANDING OCCUPATIONS: EVIDENCE FOR THE PRESENCE OF SELF-SERVING BIAS 44
3.1 Abstract 44
3.2 Introduction 45
3.3 Methods 49
3.4 Results 51
3.5 Discussion 59
3.6 Conclusion 64

CHAPTER 4: COMPARISON OF SUBJECTIVE AND OBJECTIVE JOB TASK ANALYSIS METHODOLOGIES FOR PHYSICALLY DEMANDING OCCUPATIONS 72
4.1 Abstract 72
4.2 Introduction 72
4.4 Results 82
4.5 Discussion 87
4.6 Conclusion 93
CHAPTER 5: SUMMARY AND CONCLUSION 101
5.1 Summary of Main Findings 101
5.2 Contributions to Physical Employment Standards Development 103
5.3 Limitations 105
5.4 Recommendations for Future Research 106
5.5 Conclusion 107
APPENDICES 108
Appendix A: Australian Defence Human Research Ethics 109
Appendix B: Literature Review Summary 112
Appendix C: Focus Group Information Sheet 124
Appendix D: Consent Form 127
Appendix E: Survey 128
Appendix F: Focus Group Questions 150
Appendix G: Focus Group Response Sheet 155

KEY ABBREVIATIONS

Job Task Analysis…………………………………………...............…..……JTA

Physical Employment Standard……………………………..…...….……..…...PES

Physically Demanding Occupation………..…………………….…………..……PDO

Subject Matter Expert……………………………………...………..……….SME

Focus Group……………………………..………………...……….…………FG

Royal Australian Navy…………………………………….…................……..RAN

Whole of Ship……………………………………………....………......…….WOS

Analysis of Variance…………….………..………………...…………..……ANOVA

CHAPTER 1: CONTEXT STATEMENT

1.1 Preamble

In 2012 the Australian Defence Force began the development of occupationally specific physical employment standards (PESs) across all of its employment categories. Subsequently the Land Division of the Defence Science and Technology (DST) Group was tasked with developing a scientifically defensible PES for the Royal Australian Navy (RAN) using a two stage approach. Firstly, a whole of ship (WOS) baseline standard was to be developed accurately representing the physical demands of personnel deployed on any RAN vessel, which would be followed by the quantification of physical demands for tasks in individual employment categories. The current master's thesis is aligned specifically with the first stage of this project, that is, the characterisation and identification of representative job tasks by way of a job task analysis (JTA). Prior to the involvement of the current thesis in the study, trade task workshops were conducted with experienced RAN personnel from a range of ship classes in order to establish tasks that were relevant to all personnel and could be considered WOS tasks. Since then the DST Group has extended the scope of this project to address the research questions contained within this thesis with the overarching goal of developing a method of best practice for a PES within the ADF and by extension other physically demanding occupations (PDO).

1.2 Literature Review

Physical Employment Standards

PESs are designed to measure an individual's capacity to meet the demands of a job without undue stress, strain, or injury (Constable & Palmer, 2000; Jamnik, Gumienak, & Gledhill, 2013a; Petersen et al., 2010; Tipton, Milligan, & Reilly, 2013) and are supportive of fair and non-discriminatory work practices (Jamnik, Thomas,

Burr, & Gledhill, 2010; Larsen & Aisbett, 2012; Taylor, 2013). The development of an accurate PES not only ensures that employees are able to meet the physical demands of a job role, but have also been shown to reduce the number of work related injuries and by virtue of such, the rate of health related retirement, absenteeism and staff turnover (Epstein, Yanovich, Moran, & Heled, 2013; Rayson, Holliman, & Belyavin, 2000). PESs is particularly relevant to military organisations where it is acknowledged that an individual's physical capabilities can influence their combat effectiveness (Wilkinson et al., 2011), with injury being the leading health problem in this field (Bilzon, Scarpello, Bilzon, & Allsopp, 2002; Epstein et al., 2013). As a result the military, and a range of PDOs, have invested a large amount of research and funding towards establishing and understanding the link between measures of aerobic and muscular fitness and injury or absence (Hauret, Jones, Bullock, Canham-Chervak, & Canada, 2010; Rayson, 2000). Given the rates of preventable workplace injury in these occupations, PESs are especially valuable for ensuring that employees are able to complete all tasks associated with their job role effectively and without excessive physical stress, strain or injury (Krosnick et al., 2001; Mitchell & Driskill, 1996; Tipton et al., 2013)

Legal Requirements

The necessity of an accurate PES is made apparent by the legal responsibilities of an organisation to protect its employees. Under common law an employer has a responsibility, or 'duty of care', by which they are responsible for protecting their employees against any physical harm that might result from their employment which might be avoided by ensuring they are physically capable of performing the job prior to employment Jamnik et al., 2013a; 2012; Tofari, Laing Treloar, & Silk, 2013). For example, the Australian Disability Discrimination Act (1992) requires that an evidence-based approach is adopted by employees when assessing work suitability for

recruitment, promotion and redundancy purposes, and similarly under Canadian law strict legal criteria have been implemented for PES development. Furthermore, the Meiorin Decision (SCCMD 1999) requires that all physiological job standards are rationally connected to the demands of the job and that these standards are based on both safe and efficient performance of a job role, rather that the characteristics of the individuals in that job role. As a result of these legislations it is widely acknowledged that objective, scientifically established practices are required across all stages of PES development to prevent the implementation of unlawful, discriminatory standards (Jamnik et al., 2013; Payne & Harvey, 2010; Rayson et al., 2000). Therefore, if efforts are made by an organisation to ensure that their PES is scientifically defensible, they are not only creating standards of employment that accurately reflect the demand of the job role but also legally protect themselves against litigation from employees that are unable to meet these standards.

Stages of PES Development

The goal of developing a scientifically defensible PES is however, somewhat vague given the complexity involved in PES development. To resolve this problem the Bona Fide Occupational Requirements Consensus Forum (2007) was conducted in which subject matter experts (SMEs) in public safety and human resources were consulted to determine a framework for the development of a valid PES. The result of this forum was a template of ‘best practices’ that could be used to ensure the efficacy of the final PES in a series of stages, Table 1.1.

Table 1.1

Template for developing PESs in order to qualify as a bona fide occupational requirement

Step	Description
1	Form a project management team including all stakeholders
2	Become familiar with all of the job description and associated requirements
3	Conduct a physical demands analysis
4	Establish a representative rank-ordered subset of the critical physically demanding and frequently occurring on-the-job tasks
5	Characterize the subset of the most critical physically demanding and frequently occurring tasks
6	Develop a draft physiological employment standard based on the critical physically demanding and frequently occurring tasks then pilot test and refine the physiological employment standard with job incumbents
7	Establish a standardized, objective assessment procedure for administering the physiological employment standard
8	Establish the scientific accuracy (validity and reliability) of the physiological employment standard
9	Develop performance standards for the physiological employment standard
10	Evaluate the results of applying the physiological employment standard then address any adverse impact and the possibility
	Implement the physiological employment standard
	Maintain an ongoing review of the effectiveness of the physiological employment standard

Despite the utility of this template, the methods used to complete each step are not comprehensively described. For example, Step 3, the conduct of a physical demands analysis, may be accomplished using a variety of observational techniques from which a range of physiological data can potentially be collected. Similarly the characterisation of the most critical, physically demanding and frequently occurring tasks, Step 4, may be achieved using a range of subjective and objective data collection methodologies (Larsen & Aisbett, 2012; Payne & Harvey, 2010; Tipton et al., 2013). These steps characterise a crucial stage of development that is often overlooked and inconsistent in the area of PES development, that is, the conduct of a JTA. As there is currently no

recognised best practice for JTA in the context of a PES, further exploration of these methodologies is required to inform a practice within the PES development framework, particularly in relation to the application of subjective methodologies for which there is a paucity of information regarding methodological validation in the context of PDOs (Dierdorff & Wilson, 2003; Harvey & Wilson, 2000; Larsen & Aisbett, 2012).

Job Task Analysis

JTA is the systematic process of identifying and documenting the content of a job in terms of its tasks and requirements and is important to almost every aspect of human resources, including assessment, recruitment, training and performance appraisal, and are also an important predictor of high work performance (Morgeson, Spitzmuller, Garza, & Campion, 2014). Examination of JTA methodologies is therefore important to the field of organisational and applied psychology which are primarily concerned with the assessment of an employee's job performance and suitability to a job role. In the scope of PESs, JTA typically involves identifying and characterising tasks that are the most frequently occurring, physically demanding or important, referred to as *criterion job tasks* (Petersen et al., 2010; Rayson, 2000). This process is important in the scope of PESs as it is these criterion job tasks which inform the content and nature of the resulting standards.

Despite the importance of JTA very little empirical evidence exists in relation to the critical appraisal of JTA methodologies in the context of PDOs (Dierdorff & Wilson, 2003; Harvey & Lozada-Larsen, 1988; Robert J Harvey & Wilson, 2000; Larsen & Aisbett, 2012). Specifically, very few studies have provided investigation towards determining the comparative validity and reliability of these methods (Dierdorff & Wilson, 2003; Larsen & Aisbett, 2012) which is concerning given that the methods that are used to conduct a JTA vary widely and may include a range of subjective and

objective methodologies (Harvey, 1991; Jamnik et al., 2013b; Payne & Harvey, 2010; Tipton et al., 2013). Further exploration is therefore needed to determine the comparative efficacy of objective and subjective JTA methods, such that a best practice can be developed and implemented within the PES design framework.

Subjective vs. objective methodologies

Subjective JTA methodologies include those by which subject matter experts (SMEs), defined as an incumbent or supervisor with experience and thorough knowledge of a task (Blacklock et al., 2015), are questioned regarding their perceptions of how a job task is typically completed. These methods include surveys, interviews and focus groups (FGs), and are typically used in the early stages of a PES to gain general understanding about the contents of a job role (Larsen, Graham, & Aisbett, 2013; Payne & Harvey, 2010; Tipton et al., 2013). Although occasionally subjective methods are used to determine specific job task qualities, such as how frequently a task is completed, or how physical demanding it is, these methods are typically used only as a method of shortlisting criterion job tasks. In comparison, objective methodologies comprise of observational techniques such as shadowing incumbent employees and video analysis. Most commonly, objective data collection will include a *physical demands analysis*, by which tasks are simulated to quantify the physical demands of tasks, often through observations of physiological parameters such as heart rate, oxygen consumption and strength demands (Jamnik et al., 2010; Kazmierczak et al., 2006; Payne & Harvey, 2010). Given the ability for these techniques to provide clear, objective job task data, these methods may be considered the current methodological 'gold standard' in the development of a PES. However, these methods are also known for being resource intensive in relation to the time they take to conduct, their financial demand and the risk they pose to the safety of employees undergoing task simulations. By contrast,

subjective methods offer researchers a cheap and flexible method of determining job task parameters and can be used to observe similar task variables in a low risk environment. Furthermore, these methods have tentatively been shown to be a reliable measure of these variables and may therefore provide an acceptable alternative to observational JTA methodologies.

The validation of both subjective and objective JTA methods has received little attention in the literature, thus the capacity of subjective methods to achieve accurate descriptions of physically demanding job task data remains relatively uncertain. Specifically, the validity of subjective methods in their ability to describe physical job task variables including duration, frequency, perceived importance and physical demand is an area of research that is under-developed. Ideally the accuracy of subjective methodologies could be determined by establishing alternative form validity, by which objective and subjective methods are run congruently and compared. However, to this author's knowledge only one study has attempted this comparison (Viikari-Juntura, et al., 1996). Furthermore, the measurement of systematic bias that might be inherent in these subjective measures has received a similar paucity of research, which is particularly concerning given that various job profiles and employee-related characteristics have been shown to affect job task ratings including participants' experience (Landy & Vasey, 1991; Richman & Quinones, 1996; Sanchez & Levine, 2000), length of job tenure (Maurer & Tross, 2000), age (Iddekinge, Putka, Raymark, & Eidson, 2005; Maurer & Tross, 2000) and sex (Iddekinge et al., 2005; Landy & Vasey, 1991; Schmitt & Cohen, 1989).

Overall, the exploration of subjective methodologies is important, as if these methods can be proven as a reliable and valid way of capturing job task related variables they might provide a resource-efficient and more flexible alternative to

objectives techniques, or at least be used more effectively to identify and shortlist criterion job tasks for later stages of PES development.

1.3 Research Aims and Projected Outcomes

Given the scarcity of research regarding the appraisal of subjective JTA methodologies in application to PDOs, this thesis is aligned with determining the accuracy of these methods to inform a JTA industry best practice by addressing the following research questions:

i) How do different data collection and analysis techniques, used during the job task analysis phases of PES development, inform and impact on the final representation of the 'job'?

ii) To what extent do employee characteristics bias 'job' representation and the development of PESs?

The primary aim of this project is to address the research questions above via a series of related studies, with the overarching goal of incrementally improving the scientific rigour of JTA and PES conduct. The applied outcomes for this project may manifest in a variety of forms including: an augmented 'best practice' JTA, an improved methodological approach to the development of PESs, contributing to organisational psychology literature regarding the assessment and recommendation of subjective data capture methodologies, and the development of more efficient and targeted use of resources.

1.4 Thesis Structure

This thesis is presented as a compilation of papers prepared for publication in various journals selected by their relevance to the topic area. The structure of each paper is formatted in the style of the journal for which that paper is written for. To address the aims of this thesis, each paper will look at a different aspect of the design and

implementation of JTA methodologies in the context of PDOs, with the overall goal of improving the PES framework by identifying the relative strengths and weakness of commonly used subjective and objective JTA techniques.

Paper 1 (Chapter 2) presents a systematic review of JTA conducted within PDOs. In this review, the most commonly used objective and subjective JTA methodologies are appraised and compared in terms of their methodological justification and sampling procedures, as well as the reliability and validity present within these methods. Using this information, a best practice JTA is discussed by comparing the relative advantages and disadvantages of each methodology and examining how these methodologies may be combined. Paper 2 (Chapter 3) focuses specifically subjective methodologies through investigation of systematic bias resulting from the application of a survey JTA. This survey, conducted on a sample of Royal Australian Navy (RAN) personnel, examines the extent to which individual characteristics, including those related to job profile and demographic features, influence ratings of 33 job tasks, including perceptions of task frequency, duration, distance, physical effort and importance. Through this study the reliability of survey estimates of physically demanding job tasks is explored. However, as the validity of these estimates cannot be assessed, a third paper was needed to compare these observations to observations from other subjective and objective JTA methods. Paper 3 (Chapter 4) therefore extends on the data of the previous chapter by comparing the same survey data to a series of ten FGs conducted within the same population. This was achieved through comparison of the ratings, and subsequent rankings, of all tasks by their various defining characteristics between these two methods. This study also provides a comparison of subjective ratings of physical demand, duration and distance, from both surveys and FGs, to a subset of task simulation data gathered from a single

ship platform. By doing so this paper aims to determine the relative accuracy of subjective methods in their ability to quantify physical demand and other job task parameters. Finally, Chapter 5 summarises the findings of these three papers, by focussing on practical implications for research in the areas of JTA design and the creation of scientifically rigorous PESs. This chapter also considers the major limitations of the studies and provide recommendations towards future JTA and PES research.

1.5 Significance and Originality

This project contributes to the small, yet growing body of literature targeted towards informing a JTA best practice, and may then be used to inform an internationally recognised best practice across both PES and organisational psychology literature. This will be accomplished through conduct of the first systemic review of existing JTA literature in PDOs. This review is aligned with the identification of consistent JTA themes and patterns, which may then be used to inform a best practice methodology in the scope of PES development.

The second stage of this research aims to determine the extent to which systematic bias is present in subjective JTA methods. This study is the first of its kind to specifically target the presence of systematic bias inherent in subjective JTA methodologies, and is therefore an important step towards determining the overall efficacy of these methods as a valid data collection tool.

The final stage of this thesis will explore a novel method of conducting JTA through the application of FGs, as well as surveys compared to simulation data. As only one previous study has aimed to determine the comparative accuracy of multiple subjective and objective JTA methodologies conducted within the same population, the contrast of these methodologies will provide a unique insight in to their relative

accuracy and validity. By comparing these methods the potential use of subjective JTA methodologies to provide accurate ratings of physical effort is considered by direct comparison to objective, observational methodologies.

References

Bilzon, J. L. J., Scarpello, E. G., Bilzon, E., & Allsopp, A. J. (2002). Generic task-related occupational requirements for Royal Naval personnel. *Occupational Medicine*, *52*(8), 503–510. Retrieved from http://www.scopus.com

Blacklock, R. E., Reilly, T. J., Spivock, M., Newton, P. S., & Olinek, S. M. (2015). Standard Establihsment Through Scenarios (SETS): A new technique for occupational fitness standards. *Work*, *52*(2), 375-383. doi:10.32333/WOR-152128

Constable, S., & Palmer, B. (2000). The Process of Physical Fitness Standards Development. Retrieved from http:/ ntrl.ntis.gov

Dierdorff, E. C., & Wilson, M. A. (2003). A meta-analysis of job analysis reliability. *Journal of Applied Psychology*, *88*(4), 635–646. doi:10.1037/001-9010.88.4.635

Epstein, Y., Yanovich, R., Moran, D. S., & Heled, Y. (2013). Physiological employment standards IV: Integration of women in combat units physiological and medical considerations. *European Journal of Applied Physiology*, *113*(11), 2673–2690. Retrieved from http://www.scopus.com

Harvey, R. J. (1991). Job analysis. In M. D. Dunnette & L. Hough (Eds.), *Handbook of industrial & organizational psychology* (2nd ed., pp. 71–163). Palo Alto, CA: Consulting Psycholoigsts Press.

Harvey, R. J., & Lozada-Larsen, S. R. (1988). Influence of Amount of Job Descriptive Information on Job Analysis Rating Accuracy. *Journal of Applied Psychology*, *73*(3), 457–461. doi:10.1037/0021-9010.73.3.457

Harvey, R. J., & Wilson, M. A. (2000). Yes Virginia, There Is an Objective Reality in Job Analysis. *Journal of Organizational Behavior*, *21*(7), 829–854. doi:10.1002/1099-1379(200011)21

Hauret, K. G., Jones, B. H., Bullock, S. H., Canham-Chervak, M., & Canada, S. (2010). Musculoskeletal injuries description of an under-recognized injury problem among military personnel. *American Journal of Preventive Medicine*, *38*(1), 61–70. doi:10.1016/j.amepre.2009.10.021

Iddekinge, C. H. Van, Putka, D. J., Raymark, P. H., & Eidson, C. E. (2005). Modeling Error Variance in Job Specification Ratings: The Influence of Rater, Job, and Organization-Level Factors. *Journal of Applied Psychology*, *90*(2), 323–334. doi:10.1037/0021-9010.90.2.323

Jamnik, V., Gumienak, R., & Gledhill, N. (2013a). Developing legally defensible physiological employment standards for prominent physically demanding public safety occupations: A Canadian perspective. *European Journal of Applied Physiology*, *113*(10), 2447–2457. doi:10.1007/s00421-012-2569-4

Jamnik, V., Gumienak, R., & Gledhill, N. (2013b). Erratum to: Physiological employment standards II: developing and implementing physical employment standards for safety-related occupations. *European Journal of Applied Physiology*, 1. Retrieved from http://www.scopus.com

Jamnik, V. K., Thomas, S. G., Burr, J. F., & Gledhill, N. (2010). Construction, validation, and derivation of performance standards for a fitness test for correctional officer applicants. *Applied Physiology, Nutrition, and Metabolism*, *35*(1), 59–70. doi:10.1139/H09-122

Kazmierczak, K., Mathiassen, S. E., Neumann, P., Winkel, J., Lund, U. (2006). Observer reliability of industrial activity analysis based on video recordings. *International Journal of Industrial Ergonomics*, *36*(3), 275–282. doi:10.1016/j.ergon.2005.12.006

Krosnick, J. A., Smith, V. K., Moody, W. R., Green, M. C., Conaway, M., Holbrook, A. L., Ruud, P. A. (2001). The Impact of "No Opinion" Response Options on Data Quality. *Public Opinion Quarterly*, *66*(3), 371–403. doi:10.1086/341394

Landy, F. J., & Vasey, J. (1991). Jon Analysis - The Composition of SME Samples. *Personnel Psychology*, *44*(1), 27–50. doi:10.1111/j.1744-6570.1991.tb00689.x

Larsen, B., & Aisbett, B. (2012). Subjective job task analyses for physically demanding occupations: What is best practice? *Ergonomics*, *55*(10), 1266–1277. Retrieved from http://www.scopus.com

Larsen, B., Graham, T., & Aisbett, B. (2013). A survey to identify physically demanding tasks performed during storm damage operations by Australian State Emergency Services personnel. *Applied Ergonomics*, *44*(1), 128–133.

Maurer, T. J., & Tross, S. A. (2000). SME Committee vs. Field Job Analysis Ratings: Convergence, Cautions, and a Call. *Journal of Business and Psychology*, *14*(3), 489–499. doi:10.1023/A:1022936418249

Mitchell, J. L., & Driskill, W. E. (1996). Military job analysis: A historical perspective. *Military Psychology*, *8*(3), 119–142. doi:10.1207/s15327876mp0803_1

Morgeson, Frederick P., Spitzmuller, M., Garza, A. S., Campion, M. A. (2014). Pay attention! The liabilities of respondent expereince and carelessness when making JTA judgments. *Journal of Management*. doi:10.1177/0149206314522298

Payne, W., & Harvey, J. (2010). A framework for the design and development of physical employment tests and standards. *Ergonomics*, *53*(7), 858–871. Retrieved from http://www.scopus.com

Petersen, A., Payne, W., Phillips, M., Netto, K., Nichols, D., & Aisbett, B. (2010). Validity and relevance of the pack hike wildland firefighter work capacity test:

A review. *Ergonomics*, *53*(10), 1276–1285. Retrieved from http://www.scopus.com

Rayson, M. (1998). *The development of physical selection procedures. Phase 1: Job analysis*. (M. A. Hanson, Ed.) *Contemporary Ergonomics 1998*.

Rayson, M., Holliman, D., & Belyavin, A. (2000). Development of physical selection procedures for the British Army. Phase 2: Relationship between physical performance tests and criterion tasks. *Ergonomics*, *43*(1), 73–105. Retrieved from http://www.scopus.com

Rayson, M. P. (2000). Fitness for work: The need for conducting a job analysis. *Occupational Medicine*, *50*(6), 434–436.

Richman, W. L., & Quinones, M. A. (1996). Task Frequency Rating Accuracy: The Effect of Task Engagement and Experience. *Journal of Applied Psychology*, *81*(5), 512. Retrieved from http://www.psycnet.apa.org

Sanchez, J. I., & Levine, E. L. (2000). Accuracy or Consequential Validity: Which Is the Better Standard for Job Analysis Data? *Journal of Organizational Behavior*, *21*(7), 809–818. doi:10.1002/1099-1379(200011)21

Schmitt, N., & Cohen, S. A. (1989). Internal Analyses of Task Ratings by Job Incumbents. *Journal of Applied Psychology*, *74*(1), 96–104. doi:10.1037/0021-9010.74.1.96

Taylor, N. A. S. (2013). Thematic reviews: Series II - Physiological employment standards. *European Journal of Applied Physiology*, *113*(10), 2433. Retrieved from http://www.scopus.com

Taylor, N. A. S., & Groeller, H. (2003). Work-based physiological assessment of physically-demanding trades: A methodological overview. *Journal of*

Physiological Anthropology and Applied Human Science, *22*(2), 73–81. doi:10.2114/jpa.22.73

Tipton, M. J., Milligan, G. S., & Reilly, T. J. (2013). Physiological employment standards I. Occupational fitness standards: Objectively subjective? *European Journal of Applied Physiology*, *113*(10), 2435–2446. Retrieved from http://www.scopus.com

Tofari, P. J., Laing Treloar, A. K., & Silk, A. J. (2013). A quantification of the physiological demands of the Army emergency responder in the Australian Army. *Military Medicine*, *178*(5), 487–494. Retrieved from http://www.scopus.com

CHAPTER 2: A SYSTEMATIC REVIEW OF JOB TASK ANALYSIS WITHIN PHYSICALLY DEMANDING OCCUPATIONS: IS THERE EVIDENCE FOR BEST PRACTICE?

2.1 Abstract

Employers are legally responsible for ensuring PESs are scientifically defensible and must therefore be based on reliable and valid JTA. As a method of best practice for the conduct of JTA is not recognised in existing PES literature this paper aims to examine the strengths and weaknesses of existing methodologies through a systematic review of all JTA conducted o physically demanding occupations. Objective, subjective and mixed JTA were appraised by their chosen methodologies by examining their justification and psychometric properties. Results indicated that most studies lacked explicit checks for reliability, validity and bias, highlighting the need for research into the comparative efficacy of different JTA methodologies. Although objective, observational techniques may be perceived as the 'gold standard' in the context of PES development, a mixed methodology, including the use subjective methodologies such as FGs and surveys may constitute a JTA best practice.

2.2 Introduction

Physical employment standards and the law

Organisations with physically demanding job roles are challenged with ensuring PESs are developed using scientifically defensible methods so that they are resilient against litigation (Hogan & Quigley, 1986; Jamnik, Gumienak & Gledhill, 2013a; 2013b). This legal responsibility is highlighted by statutes obligating employers to protect their employees against unfair treatment and discrimination. For example, the Canadian Supreme Court's Meiorin Decision (1999) bind employers to a 'duty of care' requiring the use of a scientific, evidence-based approach when assessing work suitability for recruitment, promotion and redundancy purposes (Jamnik, Gumienak & Gledhill, 2013a; Jamnik, Thomas, Burr & Gledhill, 2010). Similarly, The Australian Human Rights Commission Act (1986) may accept exclusion or preference relating to employment only if it is based on the *inherent* of the job. Thus by creating standards that are scientifically defensible, the resulting standards are validly linked to the job and may assist the prevention of unlawful discrimination (Hogan & Quigley, 1986; Payne & Harvey, 2010). This consideration is of particular relevance for physically demanding occupations (PDOs) where the implementation of PESs is important for the safe and efficient completion of a job and reducing the number of work-related injuries, health-related retirement, absenteeism and staff turnover (Constable & Palmer, 2000; Payne & Harvey, 2010). However, despite its importance, the information and guidelines available to those wishing to create a scientifically defensible PES are limited (Payne & Harvey, 2010; Gumieniak, Jamnik & Gledhill, 2013; Payne & Harvey, 2010; Taylor & Groeller, 2003). Specifically, there is a paucity of information regarding best practice methodologies during early stages of PES development, including the collection of job

task data that is later used for the selection of tasks that are representative of the physical demands of the job (Hogan & Quigley, 1986; Larsen & Aisbett, 2012).

Current physical employment standards guidelines

Currently, multiple general frameworks exist for the development of PESs (Gumieniak, Jamnik & Gledhill, 2013; Payne & Harvey, 2010; Taylor & Groeller, 2003). Payne and Harvey (2010) for example, provide a comprehensive framework for PES development in 14 steps, beginning with the identification of relevant job tasks and finishing with a final set of physical employment tests that accurately reflect the physical demands of critical, or 'criterion', job tasks. In this framework they describe that tasks should first be shortlisted and quantified so that critical job tasks can be identified that represent the most crucial and physically demanding components of the job role. Similarly, Rayson (2000) describes that an early step in ensuring that employees are physically fit for work is conducting detailed characterisations of job tasks through consultation with subject matter experts (SMEs) and incumbent staff. This general approach has been used in recent studies (Jamnik, Thomas, Shaw & Gledhill, 2010; Larsen, Graham & Aisbett, 2013) and is generally accepted for the development of valid 'bona fide' occupational standards that comply with Canadian laws (Gumienak, Jamnik & Gledhill, 2011; Jamnik et al., 2013a; Rayson, 1998). This process of characterising job tasks is commonly referred to as a job task analysis (JTA), which is used to determine the characteristics of the safe and efficient completion of a job task and provide a foundation for the development of PESs and other employment related decisions (Truxillo, Paronto, Collins & Sulzer, 2004). Given that the conduct of efficacious JTA is essential in the context of PESs (O'Connor & Warner, 1996; Rayson, 2000; Rouleau & Krain, 1975), robust data collection and analysis techniques must be

used at this stage to ensure that the resulting data are reliable, valid and unbiased such that they can be used to accurately inform the identification of criterion job tasks.

Despite its importance, the information available regarding the conduct of scientifically valid JTA in the context of PDOs is scarce, particularly in regards to the recommendation of specific data collection methodologies. For example, Payne and Harvey (2010) describe a two-step approach for conducting a JTA in which tasks are shortlisted based on their frequency, duration and intensity, and then quantified using a 'physical demands analysis'. However, the authors only go as far as describing that 'a range of techniques' (Payne & Harvey, 2010, p. 859) may be used at this stage. Similarly, the Bona Fide Occupational Requirement Consensus Forum states that a PES should be based on a demands analysis of a representative rank-ordered subset of physically demanding and frequently occurring tasks (Jamnik, Gumienak and Gledhill, 2013a); however the methods used to achieve these outcomes are not described. Finally, Tipton et al. (2013) describe that the elements of a task analysis may be determined through both objective and subjective analysis but do not place emphasis on any one approach.

Objective JTA methodologies

Objective data collection methods are often used to conduct JTA and validate the resulting physical employment tests (Payne & Harvey, 2010; Tipton et al., 2013). These techniques may include any observational method by which incumbent employees are assessed either while completing a job task or task simulation, and commonly include video analysis, direct observation during task performance and measurement of physiological parameters such as heart rate and oxygen consumption (Anderson, Plecas & Segger, 2001; Bos, Mol, Visser & Frings-Dresen, 2004; Jamnik et al., 2010; Larsen & Aisbett, 2012). Objective, observational methods may be used to

measure the frequency, duration and ambulatory distance of tasks, making them a powerful tool in the determination of criterion job tasks. Furthermore, given that objective evidence may be a requisite for the legal defence of a PES, objective, observational methodologies may be considered the industry 'gold standard' (Hogan & Quigley, 1986). However, as few studies have directly compared the accuracy of different subjective and objective JTA methodologies, it cannot be concluded that objective methods are inherently more accurate and reliable than subjective methodologies.

Subjective JTA methodologies

Subjective methods are used to characterise many features of PDO's by drawing on SME knowledge. Task parameters that may be characterised using these methods include perceived importance, difficulty, duration, frequency and ambulatory distance while completing tasks and are therefore useful in determining which tasks are most representative of the job role (Gumienak, Jamnik & Gledhill, 2011; Rodgers, 1988; Taylor, 2013). Ratings of importance, in particular, may be helpful in determining which tasks are critical to the successful completion of a job and cannot be measured through objective techniques. Although the use of subjective methodologies may seem superfluous in comparison to objective methods, it has been suggested all JTA conducted in the development of occupational fitness standards will involve some subjective decisions (Tipton Milligan & Reilly, 2013). Furthermore, subjective methodologies may provide a less resource intensive alternative to objective measurements and cover a larger range of job task categories and related parameters. As such, the potential of these methods to inform the identification of criterion job tasks is important in the context of the PES development framework, and are worthy of further investigation and comparison with objective, observational methodologies.

Scope of review

To the authors' knowledge, only one study has directly compared the efficacy of JTA using different methodological approaches within the same population (Viikari, Juntura et al., 1996). Although some literature suggests that a combination of both objective and subjective methodologies, can be used to successfully conduct a JTA (Tipton et al., 2013; Gumieniak, Jamnik & Gledhill, 2011), the lack of evidence regarding the appraisal of specific JTA techniques suggests that further research is needed to clarify the relative advantages of these methods such that the best methods are used to identify criterion job tasks. Given current methodological inconsistencies, this paper presents a systematic review of JTA conducted within PDOs, determining the ability of these different methods to achieve valid and reliable JTA outcomes. Focus is placed on the comparison of subjective and objective methodologies, with the efficacy of these methods assessed by comparing the reliability, validity and bias present in each method, as well as how the researchers detail and justify the use of their chosen methodology. As there are currently no systematic reviews conducted in this area, this paper provides a comparison of JTA methodologies within the context of PDOs to help determine a best practice methodology for the creation of a scientifically, and by virtue, legally defensible PES.

2.3 Method

Search strategy

The search strategy used for this review was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement which provides an evidence-based protocol for developing systematic reviews and meta-analyses (Moher et al., 2009). The PRISMA statement is used to scaffold the systematic review of research literature such that all relevant sources are systematically identified in a way

that is standardised and consistent with other PRISMA style reviews. This review targeted studies published in English in which a JTA was conducted in any PDO using a sample of SMEs or incumbent employees. In this review a JTA was defined as any attempt to measure or describe any characteristics of a job task or role, physically or otherwise.

The following online electronic databases were searched up until August 2015 based on their relevancy to the fields of physiology and organisational psychology: SAGE Journals (from 1947), SCOPUS (from 1920), Web of Science (from 1920), ProQuest (from 1973), Science Direct (from 1823), Wiley (from 1999), and PsychARTICLES (from 1860). The same general search strategy was used for all searchers with syntax customised to suit each database, Table 2.1. Duplicates within databases were automatically accounted for and removed, and one additional researcher was consulted in defining search categories to ensure no key terms were overlooked. Grey literature was also considered through correspondence with key researchers in the PES field, however no new sources were identified. All data were stored and manipulated on a single PC using EndNote X7.

Table 2.1

Search strategies and syntax

Search	Syntax
1.	((‘job task analysis’ OR ‘job analysis’ OR ‘task analysis’ OR ‘trade analysis’ OR ‘employment standard*’ OR ‘physical selection test*’) AND (‘physical*’ OR ‘PDO*’ OR ‘fitness’)) in abstract/title/keywords
2.	((‘physical standard*’ OR ‘fitness standard*’ OR ‘physically demanding task*’ OR ‘physical performance assessment’)) in abstract/title/keywords
3.	(‘trade analysis’ AND (‘physical’ OR ‘fitness’)) in abstract/title/keywords

Inclusion and exclusion criteria

All articles and conference abstracts identified by the initial search were screened in multiple stages based on inclusion and exclusion criteria, Table 2.2. Articles were first screened based on their year of publication, language and publication format. Articles published before 1950 were excluded as they were considered too old to be compared to existing methodologies, and articles published in newspapers were removed as they did not contain sufficient methodological information. The remaining articles were screened based on the relevancy of the title or journal, and articles whose relevancy could not be determined this way were screened based on the content of the abstract or introductory paragraphs. After all relevant articles were identified the reference lists of these articles were checked for relevant articles.

Table 2.2

Inclusion and exclusion criteria for systematic review of all databases.

Inclusion Criteria	Exclusion Criteria
Studies including methods relating to the conduct of a job task analysis, or job inventory for any PDO or; Studies describing the development of a physical employment standard or test	Articles published before 1950 Newspaper articles Non-English articles

Quality appraisal

All articles were assessed twice, with information summarised for each article based on the following methodological information:

- Participant demographic/industry
- Sampling procedure/sample size
- Data collection methodology and type of data collected
- How data was organised/analysed
- Justification for chosen methodology
- Reliability tested/inherent in the methodology
- Validity tested/inherent in the methodology
- Measures taken to prevent bias

2.4 Results

Literature search and screening

The initial database search returned 2728 results, Table 2.3. Following the removal of 462 duplicates, 2266 articles were screened in two stages as detailed in the PRISMA style flowchart (Figure 2.1). 2040 articles were excluded in the first stage of screening and those remaining were screened based on the content of the abstracts or introductory paragraphs resulting in the exclusion of a further 200 articles. In total, 26 articles met inclusion criteria as a result of the electronic database search, with four additional articles identified through examination of reference lists. In total 30 articles were identified and included in this review.

Table 2.3

Systematic review search result numbers, stratified by database.

Database	Results
SAGE Journals	54
SCOPUS	1113
Web of Science	419
ProQuest	462
Science Direct	611
Wiley	22
PsychARTICLES	47
Total	2728

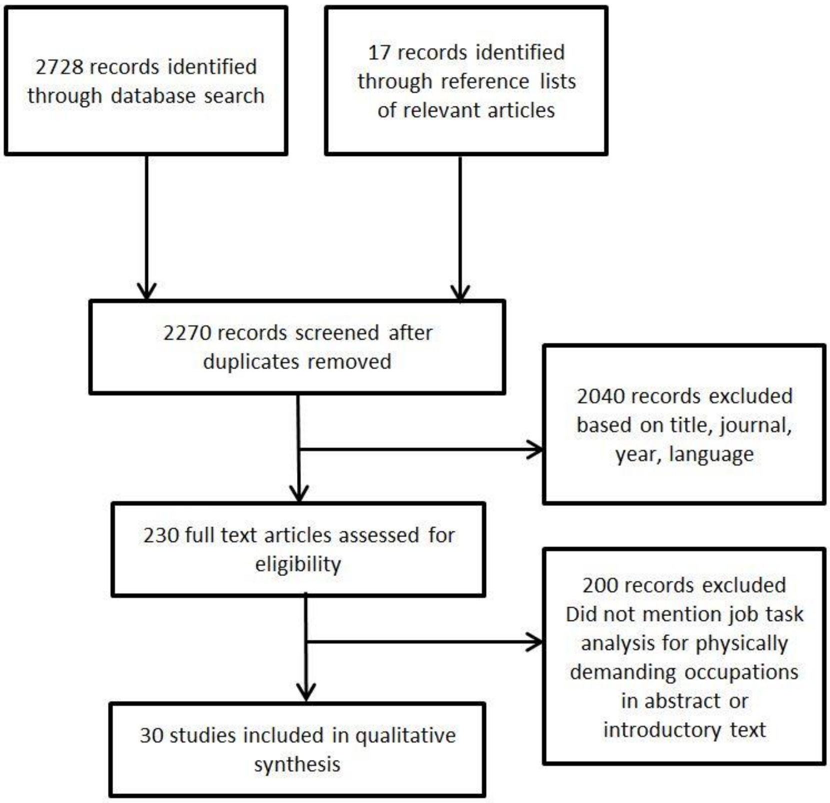

Figure 2.1. PRISMA flow chart for eligible articles reviewed including exclusion and inclusion criteria.

Quality appraisal

Appraisal was based on data extracted from the manuscripts of 30 peer-reviewed journal articles (Appendix B).

Industry and Sample Sizes. JTA were conducted across several industries including: police officers and correctional officers (n = 5), fire fighters (n = 7), military and armed forces (n = 9) and various other physically demanding areas including the forest industry, steel industry, nursing, beach lifeguards, craft jobs, agriculture, SES personnel, gas companies and rubber plant production (n = 9). In approximately half of these studies the sample sizes were not reported. In the majority of these instances a subjective JTA methodology was used. For the remaining studies an average sample size of n = 134, range: 19 - 2756, was observed.

Objective, Subjective and Mixed Methodologies. Most studies used a mixed methodological approach (n = 17), in which subjective and objective methods were both used. In this design, subjective methods were typically used to rank the importance and/or frequency of tasks, followed by the use of a physical demands analysis of the most frequently encountered or important tasks. In almost all instances there was no overlap between the task parameters (e.g. frequency, importance, duration etc.) observed by each methodology. The use of exclusively subjective methods to conduct JTA was common (n = 12), however, these studies often did not describe these methods in detail. Subjective methods included surveys or questionnaires (n = 18), structured interviews or focus groups (FGs) (n = 12), examination of existing records or literature (n = 2), logbook (n = 1), and other unspecified subjective methods (n = 1). Only two studies used predominantly objective observational methods, including the shadowing of incumbent personnel and task simulation. Objective methods included: on-the-job shadowing of incumbent employees (n = 3), video analysis (n = 4), field observation (n = 5) and task simulation (n = 8). Only one study was not described in sufficient detail to determine what methods were used and in half of the studies observed (n = 15) researchers did not provide written evidence or justification that supported their chosen

JTA methodology. Only three studies provided a review of previous JTA literature or research that supported the efficacy of their chosen method, and three articles used methods from previous studies but provided no further explanation as to why these methods were used. Furthermore, only one study contrasted the use of different JTA methodologies conducted within the same population.

Reliability. The forms of reliability considered for task ratings included: test-retest reliability, alternate-form, inter-rater and intra-rater reliability. For the majority of studies reliability was not explicitly tested (n = 16), and of these studies two were not described in sufficient detail to determine if measures of reliability were present. Of those studies that did include explicit checks for reliability, inter-rater reliability (n = 9) was the most commonly observed, occurring in instances where tasks were rated through consultation with SMEs, either in the form of a FG or structured interview. Alternate-form reliability was the second most commonly observed (n = 4), including one study in which two subjective data collection techniques were used concurrently. Test-retest reliability was observed once in which two surveys were administered concurrently on two samples. Internal consistency was observed once through the use of a Pearson's correlation analysis.

Validity. Only three studies made explicit attempts at addressing the validity of JTA results, however all studies contained some form of validity inherent in the methodological approach. For example, all data collected using objective observations of incumbent employees were considered to have ecological validity as they were based on real-time observations of incumbent personnel completing a job task (n = 12). The strength of this validity, however, was dependent on the number of observations, and the accuracy of the simulations. The majority of studies contained content validity (n = 18) in which SMEs, including experienced employees, supervisors, industry specialists

and stakeholders, were involved either directly as a part of the JTA methodology (through FGs and structured interviews), as consultants throughout the study, or through review of the final JTA outcomes. In some instances (n = 5) researchers relied on their own expert knowledge, knowledge of general incumbent personnel or knowledge based on a literature review for accurate task descriptions. These studies were considered to have only face validity given that SMEs were not consulted.

Bias. Although most studies reported participant demographics and considered the ratio of these demographics as a possible limitation, the majority of studies showed no attempts to account for bias and other confounding variables (n = 21). As such only a handful of studies made any adjustments to their methods, analysis and conclusions, based on participant and job profile characteristics. Potential sources of bias that were accounted for primarily included: age (n = 6), sex (n = 5) and measures of experience such as job tenure or seniority (n = 3). Other sources of bias accounted for included department size (n =1), job unit (n = 1) and ethnicity (n = 1), whereas self-serving bias and observer bias were not mentioned as a methodological limitation in any study.

2.5 Discussion

This systematic review explored the comparative strengths and weaknesses of objective and subjective JTA data collection methodologies through appraisal of existing JTA conducted within PDOs. This was achieved by exploring how these methodologies were justified, which methods were most commonly used, and what measures of reliability, validity and bias were tested for, or inherent, in these methodologies. The results of this review indicate that a variety of different methods are used when conducting JTA in PDOs. By comparing the relative strengths and limitations of these techniques, best practice guidelines for JTA methodology may become clearer in the scope of developing a scientifically and legally defensible PES.

Reliability

The results of this study are consistent with Dierdorff and Wilson (2003) who observed inter-rater reliability as the most prevalent form of reliability estimation in the context of JTA. In this study inter-rater reliability was found exclusively in subjective methodologies, usually estimated through the measurement of agreement between SMEs or consultation with supervisory staff. In contrast to Dierdorff and Wilson (2003) very few attempts were found to estimate intra-rater reliability, either through calculation of interclass correlations or Pearson’s correlations, suggesting that consistency of responses between raters was rarely considered. Despite this methodological oversight, studies report that JTA and task inventories typically have good internal consistency, with average reliability estimates of .70 to .90 (Dierdorff & Wilson, 2003; Gael, 1983; Wilson, Harvey & Macy, 1990), thus it could be argued that the reliability inherent in subjective survey JTA methodologies is sufficient, regardless of whether or not it is explicitly tested for. By contrast, studies using objective, observational methodologies seemingly did not consider reliability. Although evidence suggests that physiological measurements such as heart rate and oxygen consumption are consistent when measuring within participant variation (Bar Or & Zwiren, 1975; Brage, Brage, Franks, Ekelund & Wareham, 2005), they may not be stable over time due to random day-to-day variations (Pinna et al., 2007). Furthermore, the only objective, observational studies that reported reliability were those that used alternate-form procedures, but were primarily focussed on determining the validity of other measures, rather than the objective methods themselves, possibly because objective measures are assumed to be valid by nature of observation. Furthermore, observational methods typically relied upon smaller samples, which is problematic given that larger samples may be required to establish good reliability and generalisability (Charter,

1999). Therefore, although observational techniques provide a measurement tool that is stable and valid within the participant, generalisability may be negatively affected if based on small or non-randomly selected samples as seen in the examples in this study. By contrast, subjective methodologies typically use larger samples (surveys in particular), thus it is arguable that these methods are more representative of the population and may be more generalisable.

Given that very few studies provided an explanation of sampling procedures, and in many instances failed to provide a sample size at all, it is clear that greater effort needs to be invested in ensuring appropriate sampling procedures in the context of JTA. The prevalent lack of explicit reporting of reliability is concerning given that the reliability of JTA data is an important precondition for determining consistency in observations, thus a failure to consider reliability may have negative consequences for the resulting PES in terms of its scientific merit (Dierdorff & Wilson, 2003; Gael, 1983). Future studies therefore need to consider the reliability of both objective and subjective methods during both planning and data analysis stages, with future research directed towards determining the consistency of these measures. Specifically, if observational methods are used, more attention should be given towards establishing internal consistency of these measures, by ensuring sample sizes are large enough to be generalised to the larger population.

Validity

Inter- and intra-rater reliability are often used to infer validity under the assumption that the more consistent the rater's responses are, the more factually sound they are (Larsen & Aisbett, 2012). It is however, erroneous to assume that agreement between participants is the equivalent to validity (Dierdorff & Wilson, 2003; Morgeson & Campion, 1997), thus stronger, more detailed measures of validity are needed to

ensure data accuracy. Although not considered a ‘complete’ form of validity, in that it is not sufficient or necessary for the overall validity of an observation (Shadish, Cook & Campbell, 2002), ecological validity was the second most prevalent form of validity, present when tasks were observed using video analysis or task simulation methods. As ecological validity depends on the assumption that observations directly represent everyday life phenomena (Chow, 1987), the validity of task simulation experiences may be criticised on the grounds that conclusions do not represent actual task demands. To the contrary, task simulations may be contrived, as participants are not able to complete tasks as they would in the context of a normal working environment. Given that it relies on direct observation in a natural working environment, video analysis may therefore provide the best ecological validity. However, the accuracy of these observations are heavily dependent on the representation of the observations themselves, which may be affected by several confounding variables including: how representative the participants are of the population, the time in which the observations are made, and the environment in which the tasks are conducted (Tipton et al., 2013). Furthermore, video analyses are unable to provide ratings of physical effort, which are critical for the identification of the most demanding tasks. The Hawthorn effect should also be considered here in relation to all observational methodologies, which describes the unwanted bias of experimental observation on participants’ performance (Parsons, 1975), and is particularly relevant to task simulations which are often conducted in an artificial environment. Therefore, given the limitations and number of confounding variables that need to be considered, ecological validity may not be sufficient for the conduct of accurate JTA.

Content validity, in which outcomes are reviewed by a panel of SMEs, may provide a stronger form of validity. Content validity was the most prevalent form of

validity in this review; however it should be noted that in many instances only one or two SMEs were consulted. Content validity was most frequently reported in subjective JTA where outcomes, usually a ranking of the most physically demanding, frequent, or important tasks, were reviewed through consultation with a panel of SMEs. In the context of PESs, content validity may provide useful defence against litigation. For example The Australian Evidence Act (1995) states that the validity of a test may be approved by the opinion of that of an expert with 'specialised knowledge based on training, study or experience' (s.79). Therefore, establishing content validity of JTA ratings through consultation with SMEs might strengthen the legal integrity of these standards, as well as improve the overall accuracy of the observations.

Finally, convergent construct validity occurred only once in this review possibly due to its resource intensive nature. In this example a survey and logbook were used concurrently to compare duration and frequency of tasks against video observations of the same task parameters (Viikari-Juntura et al., 1996). Given the potential of this approach for ensuring data accuracy, and given the advantages of mixed method designs for complex interventions (Protheroe, Bower & Chew-Graham, 2007), establishing convergent validity may be useful in future JTA designs. Furthermore, examination of the consistency between subjective data and objective methodologies is valuable given the lack of direct contrasts between JTA methodologies.

Bias

The paucity of information relating to effects of individual characteristics on JTA outcomes is reflected in this review, with the majority of studies failing to account for sources of bias attributable to participant characteristics and other employment related variables. This is problematic given that incumbent employee characterises including sex, experience, job tenure and other job profile variables have been shown to

influence a wide range of job task related parameters (Van Iddekinge, Putka, Raymark, & Eidson, 2005; Landy & Vasey, 1991; Larsen & Aisbett, 2012; Maurer & Tross, 2000). When bias was addressed, it was typically in response to sex differences to assess whether a separate set of standards needed to be considered for men and women. Bias is of particular concern to subjective JTA methodologies which are more susceptible to influence from participant characteristics, and self-serving bias, in which tasks are rated differently based on a rater's competency or personal investment in a task (Cucina, Martin, Vasilopoulos & Thibodeuax, 2012; Cucina, Vasilopoulos & Sehgal, 2005; Richman & Quinones, 1996; Smith, 1979). By ignoring potential sources of systematic bias, a JTA might overlook important variations in response across participant subgroups, resulting in a PES that are based on the characteristics of its employees, rather than the characteristics of the job role (Harvey, 1991; Lindell, Clause, Brandt & Landis, 1998). Similarly, objective measurements may be susceptible to the same sources of bias but have the additional concern of experimenter bias in which the researchers' expectations may effect participants' engagement in a task (Sackett, 1979). Given the legal requirements that employment selection must be unbiased and based on the characteristics of the job (Jamnik et al., 2013a; Tipton et al., 2013), it is recommended that future JTA account for bias simply by comparing task ratings between participant subgroups, or by considering bias in the design of the methodology.

Objective vs. subjective methodologies

JTA considered in this study were frequently conducted in two stages, similar to the approach described by Rayson (2000). In this design, tasks are shortlisted based on subjective ratings (importance or frequency), and characterised through conduct of a physical demands analysis, typically measuring oxygen consumption, heart rate and other physiological markers. At face value, this approach is efficacious as it allows the

researcher to gain a comprehensive understanding of job tasks using the strengths of each method. A good example of this approach is provided by Jamnik, Gumienak and Gledhill (2013) who characterised job tasks using surveys to generate a rank-ordered list of the most important tasks, and then described the physical characteristics of tasks through use of simulations and biomechanical analysis. In this example, content validity was established through review of the final PES outcome by SMEs and reliability established through the test, and re-test, of physiological data. It is important to note here however that neither method in isolation w able to provide a complete JTA, given that surveys lack strong validity, and simulations often suffer from poor reliability and generalisability due to smaller sample sizes. Although these issues may be overcome by practices such as the recruitment of larger samples, the collection of objective simulation data is expensive and inflexible in the range of information it can collect, such as ratings of task frequency and importance.

Alternatively, subjective JTA methodologies may provide a resource-efficient alternative to objective methods, and may offer a faster and more convenient method of data collection (Viikari-Juntura et al., 1996). However, as very few studies have directly contrasted the use of objective and subjective JTA methodologies, the potential for surveys and FGs to accurately describe physically demanding job tasks is currently unknown. Given the advantages of subjective JTA techniques, a mixed method JTA may be considered best practice providing that appropriate safeguards for validity and reliability are implemented. For example, subjective methods could be used more effectively to shortlist and characterise the most demanding tasks prior to the conduct of expensive objective methodologies. However despite their potential utility, subjective methods require further validation if they are to be implemented successfully into a PES development framework.

2.6 Conclusion

The need for scientifically and legally defensible PESs requires that the methodologies used in the creation of these standards are valid, reliable and well justified. This review demonstrates that combinations of both objective and subjective methodologies are frequently employed in a two-stage JTA. However, given that both methodologies have issues with reliability and validity, both are limited in their ability to accurately characterise tasks, thus further evidence is needed regarding the comparative efficacy of these methodologies in order to determine JTA methodological best practice. It is important that future research is aimed towards providing clear documentation supporting JTA methods, such that their comparative reliability and validity can be better examined. Furthermore, efforts should be made to ensure appropriate sampling procedures and justification of chosen methodologies. Until such time that sufficient information exists regarding the comparative efficacy of these methods, a two-stage, mixed method approach is recommended for JTA in PDOs, with attention given towards determining the accuracy of data using appropriate tests for inter- and intra-rater reliability and content validity. However, in light of the paucity of information regarding the validity of subjective JTA techniques, these methodologies require further appraisal.

References

Anderson, G. S., Plecas, D., & Segger, T. (2001). Police officer physical ability testing - Re-validating a selection criterion. *Policing-an International Journal of Police Strategies & Management*, *24*(1), 8–31. doi:10.1108/13639510110382232

Australian Evidence Act 1995 (Cth) s. 79 (Austl.). Retrieved from https://www.comlaw.gov.au

Australian Human Rights Commission Act 1968 (Cth) s 1.3 (Austl.). Retrieved from https://www.comlaw.gov.au

Bar Or, O., & Zwiren, L. D. (1975). Maximal oxygen consumption test during arm exercise: reliability and validity. *Journal of Applied Physiology*, *38*(3), 424–426. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed

Bos, J., Mol, E., Visser, B., & Frings-Dresen, M. H. W. (2004). The physical demands upon (Dutch) fire-fighters in relation to the maximum acceptable energetic workload. *Ergonomics*, *47*(4), 446–460. Retrieved from http://www.scopus.com

Brage, S., Brage, N., Franks, P. W., Ekelund, U., & Wareham, N. J. (2005). Reliability and validity of the combined heart rate and movement sensor Actiheart. *European Journal of Clinical Nutrition*, *59*(4), 561–570. doi:10.1038/sj.ejcn.1602118

Chow, S. L. (1987). Science, Ecological Validity and Experimentation. *Journal for the Theory of Social Behaviour*, *17*(2), 181–194. doi:10.1111/j.1468-5914.1987.tb00094.x

Constable, S., & Palmer, B. (2000). The Process of Physical Fitness Standards Development. Retrieved from http://ntrl.ntis.gov

Cucina, J. M., Martin, N. R., Vasilopoulos, L., & Thibodeuax, H. F. (2012). Self-Serving Bias Effects on Job Analysis Ratings. *The Journal of Psychology, 146*(5), 511–531. doi:10.1080/00223980.2012.656155

Cucina, J. M., Vasilopoulos, N. L., & Sehgal, K. G. (2005). Personality-Based Job Analysis and the Self-Serving Bias. *Journal of Business and Psychology*, *20*(2), 275–290. doi:10.1007/s10869-005-8264-2

Dierdorff, E. C., & Wilson, M. A. (2003). A meta-analysis of job analysis reliability. *Journal of Applied Psychology, 88*(4), 635–646. doi:10.1037/001-9010.88.4.635

Gael, S. (1983). *Job analysis: a guide to assessing work activities / Sidney Gael*. San Francisco: Jossey-Bass. Retrieved from http://www.loc.gov/catdir/toc/wiley

Gumienak, R., Jamnik, V., & Gledhill, H. (2011). Physical fitness bona fide occupational requirements for safety related physically demanding occupations; test development considerations. *Health and Fitness Journal of Canada*, *5*(2), 47–52.

Gumieniak, R. J., Jamnik, V. K., & Gledhill, N. (2013). Catalogue of Canadian fitness screening protocols for public safety occupations that qualify as a bona fide occupational requirement. *Journal of Strength and Conditioning Research, 27*(4), 1168–1173. doi:10.1519/JSC.0b013e3182667167

Harvey, R. J. (1991). Job analysis. In M. D. Dunnette & L. Hough (Eds.), *Handbook of industrial & organizational psychology* (2nd ed., pp. 71–163). Palo Alto, CA: Consulting Psycholoigsts Press.

Hogan, J., & Quigley, A. M. (1986). Physical Standards for Employment and the Courts. *American Psychologist*, *41*(11), 1193–1217. doi:10.1037/0003-066X.41.11.1193

Jamnik, V., Gumienak, R., & Gledhill, N. (2013a). Developing legally defensible physiological employment standards for prominent physically demanding public safety occupations: A Canadian perspective. *European Journal of Applied Physiology, 113*(10), 2447–2457. doi:10.1007/s00421-012-2569-4

Jamnik, V., Gumienak, R., & Gledhill, N. (2013b). Erratum to: Physiological employment standards II: developing and implementing physical employment standards for safety-related occupations. *European Journal of Applied Physiology*, 1. Retrieved from http://www.scopus.com

Jamnik, V. K., Thomas, S. G., Burr, J. F., & Gledhill, N. (2010). Construction, validation, and derivation of performance standards for a fitness test for correctional officer applicants. *Applied Physiology, Nutrition, and Metabolism*, *35*(1), 59–70. doi:10.1139/H09-122

Jamnik, V. K., Thomas, S. G., Shaw, J. A., & Gledhill, N. (2010). Identification and characterization of the critical physically demanding tasks encountered by correctional officers. *Applied Physiology Nutrition and Metabolism-Physiologie Appliquee Nutrition Et Metabolisme, 35*(1), 45–58. doi:10.1139/h09-121

Landy, F. J., & Vasey, J. (1991). Job analysis - the composition of SME samples. *Personnel Psychology, 44*(1), 27–50. doi:10.1111/j.1744-6570.1991.tb00689.x

Larsen, B., & Aisbett, B. (2012). Subjective job task analyses for physically demanding occupations: What is best practice? *Ergonomics*, *55*(10), 1266–1277. Retrieved from http://www.scopus.com

Larsen, B., Graham, T., & Aisbett, B. (2013). A survey to identify physically demanding tasks performed during storm damage operations by Australian State Emergency Services personnel. *Applied Ergonomics*, *44*(1), 128–133. http://dx.doi.org/10.1016/j.apergo.2012.05.010

Lindell, M. K., Clause, C. S., Brandt, C. J., & Landis, R. S. (1998). Relationship Between Organizational Context and Job Analysis Task Ratings. *Journal of Applied Psychology*, *83*(5), 769–776. doi:10.1037/0021-9010.83.5.769

Maurer, T. J., & Tross, S. A. (2000). SME Committee vs. Field Job Analysis Ratings: Convergence, Cautions, and a Call. *Journal of Business and Psychology*, *14*(3), 489–499. doi:10.1023/A:1022936418249

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Group, P., & Grp, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ: British Medical Journal*, *339*(7716), 332–336. doi:10.1136/bmj.b2535

Morgeson, F. P., & Campion, M. A. (1997). Social and Cognitive Sources of Potential Inaccuracy in Job Analysis. *Journal of Applied Psychology*, *82*(5), 627–655. doi:10.1037/0021-9010.82.5.627

O Connor, J. S., & Warner, C. (1996). How to develop physical capacity standards. *Personnel Journal*, 8. Retrieved from http://ezproxy.uow.edu.au

Payne, W., & Harvey, J. (2010). A framework for the design and development of physical employment tests and standards. *Ergonomics*, *53*(7), 858–871. Retrieved from http://www.scopus.com

Pinna, G. D., Maestri, R., Torunski, A., Danilowicz-Szymanowicz, L., Szwoch, M., La Rovere, M. T., & Raczak, G. (2007). Heart rate variability measures: A fresh look at reliability. *Clinical Science*, *113*(3-4), 131–140. doi:10.1042/CS20070055

Protheroe, J., Bower, P., & Chew-Graham, C. (2007). The use of mixed methodology in evaluating complex interventions: Identifying patient factors that moderate the

effects of a decision aid. *Family Practice, 24*(6), 594–600. doi:10.1093/fampra/cmm066

Rayson, M. (1998). *The development of physical selection procedures. Phase 1: Job analysis*. (M. A. Hanson, Ed.) *Contemporary Ergonomics 1998*.

Rayson, M. P. (2000). Fitness for work: The need for conducting a job analysis. *Occupational Medicine, 50*(6), 434–436. Retrieved from http://www.occmed.oxfordjournals.org

Richman, W. L., & Quinones, M. A. (1996). Task Frequency Rating Accuracy: The Effect of Task Engagement and Experience. *Journal of Applied Psychology, 81*(5), 512. http://www.psycnet.apa.org

Rodgers, S. H. (1988). Job evaluation in worker fitness determination. *Occupational Medicine (Philadelphia, Pa.), 3*(2), 219–239. Retrieved from http:// http://www.ncbi.nlm.nih.gov

Rouleau, E. J., & Krain, B. F. (1975). Using Job Analysis to Design Selection Procedures. *Public Personnel Management*, *4*(5), 300. Retrieved from http:// http://www.researchgate.net/

Sackett, D. L. (1979). Bias in analytic research. *Journal of Chronic Diseases*, *32*(1), 51–63. doi:10.1016/0021-9681(79)90012-2

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalised causal inference*. Boston: Houghton Mifflin

Smith, J. E. (1979). Convergence among Data Sources, Response Bias, and Reliability and Validity of a Structured Job Analysis Questionnaire. *Personnel Psychology*, *32*(4), 677–692. doi:10.1111/j.1744-6570.1979.tb02340.x

Supreme Court of Canada, 1999; *British Columbia (Public Service Employee Relations Commission) v British Columbia Government Service Employee' Union.* 3 S.C.R. 3 (Meirorin Decision). Retrieved http://scc-csc.lexum.com

Taylor, N. A. S. (2013). Thematic reviews: Series II - Physiological employment standards. *European Journal of Applied Physiology*, *113*(10), 2433. Retrieved from http://www.scopus.com

Taylor, N. A. S., & Groeller, H. (2003). Work-based physiological assessment of physically-demanding trades: A methodological overview. *Journal of Physiological Anthropology and Applied Human Science*, *22*(2), 73–81. doi:10.2114/jpa.22.73

Tipton, M. J., Milligan, G. S., & Reilly, T. J. (2013). Physiological employment standards I. Occupational fitness standards: Objectively subjective? *European Journal of Applied Physiology*, *113*(10), 2435–2446. Retrieved from http://www.scopus.com

Truxillo, D. M., Paronto, M. E., Collins, M., & Sulzer, J. L. (2004). Effects of Subject Matter Expert Viewpoint on Job Analysis Results. *Public Personnel Management*, *33*(1), 33–46. Retrieved from http://ppm.sagepub.com

Viikari-Juntura, E., Rauas, S., Martikainen, R., Kuosma, E., Riihimaki, H., Takala, E. P., & Saarenmaa, K. (1996). Validity of self-reported physical work load in epidemiologic studies on musculoskeletal disorders. *Scandinavian Journal of Work Environment & Health*, *22*(4), 251–259. Retrieved form http://www.ncbi.nlm.nih.gov

Van Iddekinge, C. H., Putka, D. J., Raymark, P. H., & Eidson, C. E. (2005). Modeling Error Variance in Job Specification Ratings: The Influence of Rater, Job, and

Organization-Level Factors. *Journal of Applied Psychology*, *90*(2), 323–334. doi:10.1037/0021-9010.90.2.323

Wilson, M. A., Harvey, R. J., & Macy, B. A. (1990). Repeating Items to Estimate the Test-Retest Reliability of Task Inventory Ratings. *Journal of Applied Psychology*, *75*(2), 158–163. doi:10.1037/0021-9010.75.2.158

CHAPTER 3: THE APPLICATION OF SUBJECTIVE JOB TASK ANALYSIS TECHNIQUES IN PHYSICALLY DEMANDING OCCUPATIONS: EVIDENCE FOR THE PRESENCE OF SELF-SERVING BIAS

3.1 Abstract

The aim of this study was to determine if perceptions of physically demanding job tasks are biased by employee' demographics and employment profile characteristics including: age, sex, experience, length of tenure, rank, and if they completed or supervised a task. Surveys were administered to 427 Royal Australian Navy personnel who characterised 33 tasks in terms of physical effort, importance, frequency, duration and vertical/horizontal distance travelled. Results showed no evidence of bias resulting from participant characteristics, however participants who were actively involved in both task participation and supervision rated these tasks as more important than those involved only in the supervision of that task. This may indicate self-serving bias in which participants that are more actively involved in a task had an inflated perception of that task's importance. These results have important implications for the conduct of JTA, especially the use of subjective methodologies in the development of scientifically defensible PESs.

3.2 Introduction

For a PES to be considered legally 'bona fide', they must accurately represent the critical and physically demanding requirements of a job role (Gumieniak, Jamnik, & Gledhill, 2013; Jamnik, Gumienak, & Gledhill, 2013a). This requirement is made apparent by anti-discrimination legislations, including the Canadian Supreme Court's Meiorin Decision (SCCMD 1999) and the Australian Disability Discrimination Act (1992), which both require employment standards to be scientifically valid to prevent unfair employment related discrimination. To ensure this level of validity, JTA are typically conducted in early stages of PES development to determine how tasks are typically performed and measure their associated physical demands (Anderson, 1994; Harvey, 1991; Rayson, 2000). JTA are typically conducted on a sample of incumbent employees, or subject matter experts, and may involve an array of subjective and objective methods. These methods may include data collection techniques relying on participant's perceptions such as surveys, FGs, and interviews, as well as more objective observations of task performance through task simulation (Larsen & Aisbett, 2012; Taylor & Groeller, 2003; Tipton, Milligan & Reilly, 2013). Data from these methods are then used to identify a set of job tasks that are deemed to be the most physically demanding and critical for the effective and safe performance of the job (criterion tasks) and help inform PES development and other employment related decisions. Conducting a thorough and accurate JTA is therefore critical to the development of valid a PES (Tipton et al., 2013) and an array of other human resource system functions including the development of worker safety protocols and the reduction of work related injuries (Keyserling, Herrin, & Chaffin, 1980; Rayson, 1998; 2000).

Despite the importance of establishing scientifically valid employment standards, there are no specific guidelines surrounding what methodologies should be applied when conducting a JTA, particularly in the context of physically demanding occupations (PDOs). The result of this methodological obscurity is evident in the JTA literature, where array of subjective and objective methodologies are employed to collect various job related information (Jamnik, Gumienak & Gledhill, 2013b; Tipton et al., 2013). As subjective methodologies are commonplace in JTA literature, the paucity of information regarding the measurement of systematic bias, especially in the context of PDOs, is concerning given the legal responsibilities of employers (Grisez, 1996; Jamnik et al., 2013b; Larsen & Aisbett, 2012). In particular, sources of systematic bias relating to subjective ratings of physical effort and other job task parameters has received very little attention throughout the literature (Larsen & Aisbett, 2012).

Systematic Bias

Systematic bias is defined as any repeatable or consistent source of error that is not attributable to chance (Gove, McCorkel, Fain, & Hughes, 1976). In the context of JTA observations this may include any source of variation affecting the capture of accurate job related information that is not related to the conduct of that task. For example, researcher expectations, participant expectations, and the characteristics of the participants themselves are all sources of systematic bias that might effect JTA outcomes (Lindell, Clause, Brandt, & Landis, 1998; Sedgwick, 2014; Smith, 1979). The consideration of participant characterises is particularly relevant to PDOs where the accurate and unbiased characterisation of job tasks is essential in developing PESs that are representative of the entire sample and not disproportionally skewed by a population subgroup (Hogan & Quigley, 1986; Jamnik et al., 2013a). Unfortunately, variations in job task ratings are often overlooked in the conduct of JTA and instead are attributed to

random error (Harvey, 1991). It is therefore common that researchers will conduct an aggregate or *random model* JTA, by which job task data are averaged across participant responses, potentially omitting information relating to the influence of participant characteristics on job task ratings (Maurer & Tross, 2000). Given that one aim of a JTA is to aid in creation of unbiased, non-discriminatory standards, the examination of between group differences resulting from demographic and work related characteristics is worthy of further examination so that potential sources of bias may be accounted for in future designs.

Participant Bias

Harvey (1991) states that job task ratings should be verifiable, replicable and independent of the personal characteristics of the employees that are directly related to an individual's experience, and conduct, of a task. Similarly, Van Iddekinge, Putka, Rayson, and Eidson (2005) state that unreliability of job analysis ratings is typically a result of idiosyncratic rater differences, which may be particularly relevant in the context of subjective data collection methodologies in which information is gathered from experience and perceptions of individuals (Grisez, 1996; Peytcheva & Groves, 2009; Sedgwick, 2014). Examples of bias resulting from participants' demographic and job profile characteristics, including sex, level of experience, length of job tenure, and education can be found in studies across both PDO and non-PDO, however the results of these studies often have mixed and inconsistent results (Grisez, 1996; Larsen & Aisbett, 2012). For example, Arvey, Passino, and Lounsbury (1977) report no differences between sexes for ratings of job perceptions, Schmitt and Cohen (1989) show significant sex differences in job task ratings for the time spent on a job, with females reporting less involvement in a particular subset of tasks. Furthermore, recent studies suggest that sex differences may have a small yet significant effect in relation to

perceptions of job performance and ratings of job importance (Iddekinge et al., 2005; Landy & Vasey, 1991), but go on to describe that these effects may be moderated by other participant characteristics.

By comparison, ratings of task frequency may be dependent on the respondent's level of job experience (Borman, 1992; Landy & Vasey, 1991; Richman & Quinones, 1996) and job tenure (Tross & Maurer, 2000), and may characterise a more consistent effect based on the limited evidence available. Similarly, there is compelling evidence that a participant's level of involvement in a task may affect his or her perception in a form of *self-serving bias*, by which a task is rated as more important depending on the participant's investment in its successful completion (Cucina, Martin, Vasilopoulos, & Thibodeuax, 2012; Cucina, Vasilopoulos & Sehgal, 2005). This theory is supported by Morgeson and Campion (1997) who explain that participants may engage in 'impression management', by which they will attempt to propagate the view that they are good employees by rating the tasks that they are more proficient in as more important or demanding. However, the extent to which this effect is applicable to JTA ratings has received very little examination, thus further evidence is needed.

Aims of Study

Considering the paucity of information surrounding the presence systematic bias in subjective JTA methodologies and the need for an improved JTA methodological framework, the aim of this paper was to determine the extent to which characteristics relating to a participant's demographic characteristics and employment profile impact on perceptions of job task performance. Based on the limited evidence available characteristics relating to a participant's demographic and employment profile characteristics were examined, including age, sex, experience, length of job tenure, rank and whether they were involved with the supervision or completed a task. These

variables were considered in relation to ratings of task frequency, duration, distance, physical effort, and importance.

3.3 Methods

Pilot Survey

Prior to data collection, a pilot survey was developed and administered to identify any formatting and content issues. The selection of relevant job tasks was based on the conduct of a job task inventory with a panel of subject matter experts (Middleton et al., 2014). Participants included 13 highly experienced Royal Australian Navy (RAN) sailors with extensive knowledge of job tasks that could be completed by any staff across all classes of ship, otherwise known as whole of ship (WOS) tasks (age 36.7 ± 9.3 years, range 23 – 54; RAN service 16.5 ± 11.2 years, range 5.5 – 38). A focus group was conducted following completion of the pilot survey to gather feedback regarding the design and content of the survey. This pilot revealed that participants were generally satisfied with the content and design of the survey and agreed that no relevant job tasks had been omitted or were superfluous.

Survey

All participants in this study gave informed, written consent prior to participation (Appendix C), and all data was kept anonymous in accordance with approval granted by the Australian Defence Human Research Ethics Committee and the Human Research Ethics Committee at the University of Wollongong (Appendix A). In total 427 RAN personnel were recruited to take part in this study through communication with an RAN liaison officer. Prior to participating, participants were briefed regarding the history and ethical requirements of the survey and were provided with an information sheet (Appendix D) and consent form (Appendix C). All

participants completed the survey online, using one of 16 password iPads or a desktop computer.

The survey consisted of 52 questions including demographics information and questions relating to WOS tasks performed on the sea-going vessel to which they were last posted to (Appendix D). These WOS tasks included: emergency procedures, fire-fighting, leak-stop and repair, toxic hazard, casualty evacuation, and storing procedures, Table 3.1. Participants were asked to describe WOS tasks that they completed in terms of their frequency (number of times per week), duration (in minutes), and ambulation (both horizontally in meters and vertically in decks) using an open answer response format. Participants also rated the importance and physical effort of tasks on a 7-point likert scale from 1-7, where 1 = not very important at all/very easy and 7 = extremely important/very difficult. Participants were able to leave any general comments and describe tasks that may have been left out. The survey was created online using Qualtrics (http://qualtircs.com, Qualtrics, Provo, UT, USA).

Table 3.1

Whole of ship (WOS) tasks performed by RAN personnel

Task	Task description
1	Perform line handling
2	Participate in the breakdown of a pallet of stores while at sea
3	Participate in storing a vessel while alongside
4	Closing up to action stations
5	Closing up to emergency stations
6	Closing up to leaving stations
7	Conduct a single emergency cable run in 5 minutes
8	Lift and carry a fire extinguisher and enter affected compartment within one minute (FAA)
9	Lift and carry a fire extinguisher and enter affected compartment within three minutes (BA-H)
10	Lift and carry a fire hose, attach to water main and enter affected compartment in seven minutes (BA-P)
11	As a nozzleman, participate in sustained use of a charged fire hose
12	As the IC, move and support nozzleman
13	As a hose handler, move with and support nozzleman's charged hose
14	As a hose handler/inductor/hydrant, hold hose for an extended period of time
15	Conduct boundary cooling
16	Conduct a fire overhaul
17	Enter affected compartment within three minutes of the alarm in search of casualties
18	Lift and carry as a team of three, a de-watering pump a distance of x metres in 3 minutes
19	Lift and carry as a team of two, a de-smoking fan a distance of x metres
20	Cut 4x4 oregon timber to size using a hand saw
21	As a team of two, carry timber piece from storage area to required site
22	As a team of two, carry acro shoring from storage area to required site and erect by twisting
23	Hammer wedges into place in order to secure vertical and breast pieces
24	Hammer plugs into place in order to maintain hull integrity
25	Carry a tool/repair bag and conduct a permanent pipe repair
26	As a member of team 1, enter affected compartment and spiral upwards to meet team 2
27	As a member of team 2, enter affected compartment and spiral downwards to meet team 1
28	As a member of team 3, enter gas boundary and evacuate casualty
29	As a member of a Team 4, carry a kit bag and repair and clean up toxic hazard
30	Individually or in a team of 2, perform a fire hose lift
31	Individually or in a team of 2, perform a Res-Q-Mate stretcher lift
32	In a team of 6-8, lift and carry a casualty on a Res-Q-Mate stretcher from site of injury to sick bay
33	In a team of 2, lift and carry a casualty using a RAN safety lift to sick bay

Data Analysis

Demographic data were summarised as means, standard deviations, and frequencies. Frequency and duration data were considered outliers and excluded if they lay outside +/- 3 standard deviations of the mean rating for each task as all of these data were determined to be highly improbable or impossible given the description of the task. Vertical and horizontal distance data were restricted by the length/number of decks respective to each vessel. No limits were applied to importance and physical effort data. Instances which participants had answered with the same value for three or more

consecutive tasks across all task variables (e.g. input a value of 0 for task frequency, duration and distance) were manually identified and removed prior to analysis if these responses were impossible for that task's description.

Task ratings of frequency, duration, horizontal distance, vertical distance, physical effort and importance were treated as the dependent variable for all analyses. Participant characteristics including: age, sex, time served in RAN, time since last at sea, time on current platform, rank, and whether the participant had completed, supervised, or both completed and supervised each task, were treated as the independent variables. Rank, which describes a personnel's seniority within the RAN, was grouped into two categories (seamen and higher ranks including leading seaman, non-commissioned officers, and officers) after consultation with RAN liaison officers. The categorical variable 'platform', consisting of three levels (major war vessels, minor vessels, and submarine) was included as a co-variate in each analysis to account for variance attributable to between-platform differences.

Relationships between dependent variables and all independent variables, excluding task completion/supervision, were examined by including all variables in a multiple linear regression model. In this model ship platform was included using dummy coding, with 'major war vessel' acting as the baseline category. For the variables sex and rank, males and seamen were treated as the baseline category respectively. For these analyses unadjusted beta values, 95% confidence interevals and associated *p* values are presented. Significant relationships were then tested for linearity, normality, and heteroscedasticity to ensure the validity of any significant linear relationships.

Differences between participants that completed task versus those that supervised tasks versus those that completed and supervised tasks were analysed using a univaraite ANOVA model which included all remaining independent variables as co-variates. Welch's corrections were applied for violations of homogeneity of variance and significant were further examined using post-hoc Tukey multiple comparisons.

For these analyses unadjusted beta values, mean differences, 95% confidence intervals and associated *p* values are presented when significant relationships where observed. For all analyses an alpha level of .05 was used as the basis for statistical significance.

3.4 Results

Participants

In total, 468 incumbent RAN personnel participated in the survey; with 41 participants removed based on exclusion criteria. As a result, 427 participants from a range of demographics, ranks, job categories, and experience levels were included in the final analysis, Table 2. Sex ratio, age category, job tenure, and the rank of the current sample were compared to the demographics of the entire permanent RAN staff population. No notable differences were found in the distribution of these demographics between groups through observation of grouped frequencies, thus our sample was considered to be representative of the entire RAN population.

Table 3.2

Survey participant demographics

Variable	n	Min	Max	Mean (SD)	Median
Age	427	18	56	29.61 (8.29)	27
Months since last at sea	421	0	24	3.29 (4.49)	2
Years served in RAN	426	<1year	39	8.22 (6.70)	6
Months in current posting	427	0.5	72	12.31 (9.85)	10

	n (%)
Sex	
Male	361 (84.5)
Female	66 (15.5)
Rank	
Seaman	216 (50.6)
Leading seaman, NCO, officer	211 (49.4)

Sex

Differences resulting from participant sex were observed using between group analyses. These analyses revealed very few significant differences between groups, however examination of mean scores showed a consistent, non-significant relationship between sex and ratings of task importance, with females rating tasks as less important than males. A ceiling effect was also observed for this data in which most tasks were rated as 'very' or 'extremely' important. Other than this trend significant differences were observed between sexes for the physical effort of storing a vessel (β = .02, p = .002), the horizontal distance travelled for a toxic hazard repair team, first aid attack and hands to action stations (β = 11.68 to 21.78, p = .007 to .033), the vertical distance travelled for hands to action stations (β = 1.14, p = .020), and the duration of lifting and carrying de-smoking fan (β = 6.78, p = .028). For all tasks except storing the vessel males rated tasks higher than females, Table 3.

Table 3.3

Significant sex differences using 2-way ANOVA and Kruskal-Wallis tests

Parameter	**Task**	**β**	**95% CI**	**p value**
Duration	Lift and carry de-smoking fan	-6.78	-12.81, -0.75	0.028
Physical Effort	Storing the vessel	0.02	0.01, 0.03	0.002
Horizontal Distance	Team 4 (Repair)	-21.78	-37.58, -5.97	0.007
	FAA	-11.68	-22.43, -0.97	0.033
	Hands to action stations	-17.15	-31.84, -2.47	0.022
Vertical Distance	Hands to action stations	-1.14	-1.85, -0.42	0.002

Rank

Comparisons between ranks revealed several significant differences across various task parameters. Where significant differences were observed, seamen tended to rate tasks as further in horizontal movement (β = 6.49 to 12.14, p = .016 to .036) and occurring more frequently (β = 1.55 to 1.76, p = .013 to.033) than those from higher rank groups, Table 4. The consistency of the tasks for which these differences were observed was however, somewhat disperse across tasks and task parameters. In addition to this, several significant differences were found for ratings of task duration (β = 4.58 to 9.28, p = .006 to .044), with leading seamen and seamen consistently rating tasks as taking longer than commanding officers and non-commissioned officers for seven tasks, Table 4.

Table 3.4

Significant rank differences using 2-way ANOVA and Kruskal-Wallis tests

Parameter	Task	β	95% CI	p value
Importance	Storing the vessel	0.40	0.05, 0.76	0.025
Frequency	Boundary cooling	-1.55	-2.78, -0.33	0.013
	Breakdown pallets	-1.76	-3.36, -0.15	0.033
Duration	Nozzelman	5.32	0.85, 9.79	0.020
	IC	6.06	1.03, 11.09	0.019
	Hose hander	6.09	1.33, 10.86	0.013
	Inductor/hydrant	6.96	1.99, 11.93	0.006
	Boundary cooling	4.58	0.12, 9.03	0.044
	Carry and erect Acro shoring	9.28	0.74, 17.81	0.034
	Conduct permanent pipe repair	8.02	2.11, 13.92	0.008
Horizontal	Storing the vessel	-6.49	-12.56, -0.42	0.036
	Hands to action stations	-15.18	-28.51, -1.85	0.026
	Hands to emergency stations	-12.14	-21.98, -2.31	0.016
Vertical	Fire overhaul	0.89	0.17, 1.61	0.015

Task Completion vs. Supervision

Between group analyses revealed 12 significant differences between participants who ‘completed’, ‘supervised’ or ‘completed and supervised’ a task, seven of which related to differences in importance, Table 5. Multiple comparisons revealed that participants who had both completed and supervised a task rated those tasks as significantly more important than those who had only supervised the task with mean differences ranging from .36 to 1.98 (p = .002 to .035), Figure 1. Furthermore, for all remaining tasks a non-significant trend was observed in which personnel that completed tasks rated them as more important that those who supervised those tasks.

Table 3.5

Significant (p < .05) differences for task supervision vs. completion using 2-way ANOVA and Kruskal-Wallis tests

Task	**Parameter**	**p**	**Direction**	**Mean difference**	**95% CI**	**Tukey p**
Closing up to emergency stations	Importance	0.020	Both>Sup	0.97	0.08, 1.86	0.027
Closing up to leaving ship stations	Horizontal	0.028	Com>Sup	10.02	3.77, 35.63	0.016
Cable run	Physical	0.019	Com>Sup	1.01	0.03, 2.00	0.044
FAA	Vertical	0.037	Com>Both	-0.59	-1.16, -0.01	0.046
BA-P	Importance	0.035	Both>Com	0.36	0.05, 0.72	0.046
Nozzleman	Importance	0.019	Both>Sup	0.71	0.34, 1.38	0.037
Hose handler	Importance	0.028	Both>Sup	0.83	0.22, 1.45	0.008
			Sup>Com	0.61	0.05, 1.16	0.034
Hammer wedges	Importance	0.024	Both>Com	0.75	0.08, 1.41	0.028
			Both>Sup	1.36	0.32, 2.39	0.011
Conduct permanent pipe repair	Importance	0.002	Both>Sup	1.98	0.66, 3.29	0.002
			Per>Sup	1.33	0.12, 2.53	0.027
Team 4 (Repair)	Horizontal	0.013	Both>Sup	30.14	6.36, 53.92	0.013
Team 4 (Repair)	Importance	0.031	Both>Com	0.47	0.02, 0.92	0.042
Carry Res-Q-Mate stretcher	Vertical	0.039	Both>Com	0.91	0.11, 1.71	0.027
			Both>Sup	1.5	0.09, 2.91	0.037

Both = Both completed and supervised task, Sup = Supervised task only, Com = Completed task only

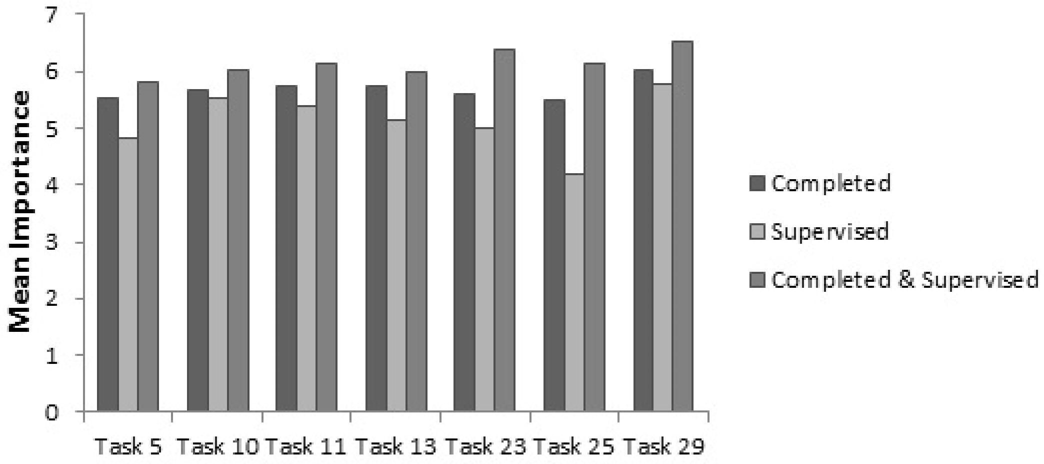

Figure 3.1. Significant differences for mean ratings of importance across all tasks between participants that completed vs. supervised tasks.

Age, Experience, and Job Tenure

Multiple regression analyses showed age as a significant predictor of all task parameters across several different tasks. However, there was little consistency in these relationships across tasks and task parameters with small significant β values observed across all tasks ranging from .02 to 2.32 (p = .001 to .035), Table 6. Similarly, time served in RAN, time on current vessel, and time since last at sea showed inconsistent, non-linear relationships across all tasks and task variable, with β values ranging from .02-2.65 (p = .005 to .049).

Table 3.6

Significant multiple regression beta coefficients

	Task	***β***	**95% CI**	***p***
		Age		
Frequency	1	0.14	0.01, 0.26	0.029
Duration	4	-1.39	-2.57, 0.21	0.021
	5	-0.70	-1.35, -0.05	0.035
	11	-0.37	-0.72, -0.03	0.035
	12	-0.46	-0.85, -0.06	0.023
	21	-0.37	-0.72, -0.01	0.044
	29	-0.59	-1.09. -0.09	0.021
Horizontal Distance	18	-2.19	54.97, 200.32	0.001
	22	-2.41	-4.19, -0.64	0.009
Vertical Distance	18	0.17	0.03, 0.31	0.021
	29	0.09	0.01, 0.17	0.029
		Time served in RAN		
Duration	2	-2.06	-3.54, -0.58	0.007
	29	0.66	0.01, 1.32	0.046
Horizontal Distance	30	0.15	0.01, 0.29	0.038
Vertical Distance	18	0.17	0.03, 0.31	0.022
	19	0.20	0.03, 0.38	0.022
		Time in current position		
Frequency	1	0.44	0.06, 0.82	0.025
	8	0.06	0.001, 0.11	0.046
	9	0.06	0.01, 0.12	0.021
	16	0.06	0.00, 0.16	0.05
	26	0.07	0.01, 0.13	0.026
Duration	3	0.60	0.14, 1.06	0.011
	17	0.24	0.05, 0.42	0.013
Horizontal Distance	8	0.45	0.06, 0.84	0.026
	29	0.62	0.04, 1.19	0.038
	12	0.48	0.002-0.95	0.049
		Time since last posted		
Frequency	4	0.31	0.06, 0.55	0.016
	10	0.14	0.001-0.28	0.049
	11	0.18	0.02, 0.34	0.025
	13	0.20	0.06-0.35	0.060
Duration	2	-1.42	-2.77, -0.07	0.039
	8	-0.33	-0.56, -0.09	0.008
	28	-0.49	-0.95, -0.03	0.036
	30	-0.66	-1.16, -0.15	0.012
Horizontal Distance	24	-2.63	-5.01, -0.18	0.036
Vertical Distance	7	-0.22	-0.42, -0.02	0.037

3.5 Discussion

The aim of this study was to investigate the presence of systematic bias in a JTA survey by examining the relationship between job task perceptions, and participants' demographic, and job profile characteristics. Results indicated that perceptions of task frequency, duration, distance, and physical effort were not consistently influenced by participants' characteristics, including their age, sex,

experience, job tenure, time on current vessel, and time since last at sea. There was however several significant relationships observed between whether the participant had completed or supervised a task, and ratings task importance, in addition to a notable effect of rank on participants' perceptions of task duration. These findings have important implications for the conduct of JTA in the context of PES development and the identification of a scientifically valid best practice methodology.

Participant Demographics

Very few differences were observed between sexes for ratings across all job tasks and variables. The absence of significant sex differences for ratings of physical effort suggests that men and women may perceive tasks as equally demanding, a finding which is consistent with previous research that show negligible effects of sex on job task ratings (Arvey, Passino, & Lounsbury, 1977; Iddekinge et al., 2005; Jamnik, Thomas, Shaw, & Gledhill, 2010; Landy & Vasey, 1991). Interestingly, this evidence contradicts result from physical demand analyses conducted in PDOs that demonstrate substantial differences in physical ability between male and female employees (Arnold, Rauschenberger, Soubel, & Guion, 1982; Arvey, Landon, Nutting, & Maxwell, 1992; Hughes, Ratliff, Purswell, & Hadwiger, 1989) which may suggest that perceptions of physical demand measured using likert scales may not be sensitive enough to detect differences across sexes. Although there is evidence to suggest that subjective ratings of physical effort using 7-point likert type scales are valid, reliable, and related to actual metabolic costs (Hogan & Fleishman, 1979; Hogan, Ogden, Gebhardt, & Fleishman, 1980), these studies were conducted during or immediately after the task completion. As recall typically loses accuracy over time, it is likely that the recall of a task's physical demand is less sensitive in the context of a subjective JTA conducted days, or even

months, after the task was last completed as was the case in the present study. However, the overall paucity of information in this area of research, and towards validation of subjective JTA techniques, makes it difficult to draw definitive conclusions.

Participant's age also showed few significant relationships across tasks, which is consistent with the evidence indicating negligible effects of age on task ratings (Iddekinge et al., 2005; Larsen & Aisbett, 2012; Maurer & Tross, 2000). It is therefore likely that age has no effect on ratings of physically demanding job tasks within the current RAN population.

Job Profile Characteristics

Participants from lower ranks tended to rate tasks as taking significantly shorter durations than participants from senior ranks. One explanation for these differences might be that participants from lower ranks were more likely to complete tasks (rather than supervising them) as part of practice and proficiency training. These participants may therefore have perceived tasks as taking shorter time in comparison to supervisors who may spend longer preparing and recording performance for each task. However, as it was difficult to determine if participants were responding to their experience of the task or what they believed to be 'typical' practice, this interpretation of these data should be treated with caution.

By contrast experience, job tenure, and time since last posted to sea showed only a small number of significant relationships between task parameters. Although 'time on current vessel' did show significant relationships with some tasks, these relationships were both small and non-linear and likely the result of Type I error. This evidence is somewhat divergent from previous literature which indicates significant relationships between experience and ratings of importance (Iddenkinge et al. 2005), frequency

(Landy & Varsey, 1991), and relationships between job tenure and ratings of task frequency (Tross & Maurer, 2000). However, it should also be noted that these studies had very small effect sizes, in some instances accounting for as little as 2% variance in task ratings. It could therefore be argued that the results of the present study are consistent with past literature, suggesting that the length of time in a job role is independent of task perception.

Self-Serving Bias

Between group analyses revealed that participants who were involved in both the completion and supervision of a task, rated those tasks as more important that those who only completed the task, and significantly more important that those who only supervised the task. Although this effect was only found to be significant for seven tasks, all other tasks showed the same non-significant differences between groups. This effect may be explained by the presence of self-serving bias, in which participants with more investment in a task will rate that task as more important. Although limited, previous literature supports the existence of self-serving bias in relation to JTA ratings of importance and perceived task competency (Aguinis, Mazurkiewicz, & Heggestad, 2009; Cucina et al., 2012, 2005). These studies suggest that individuals are likely to report tasks involving traits that they are more proficient in as more important (Cucina et al., 2012). A popular framework for this effect is based on the process of 'impression management' by which individuals will rate the tasks they have a greater investment in as more critical in an effort to create the perception that their job role is more important. As such, it could be argued that those participants involved in both completion and supervision of a task have a tendency to overestimate importance based on the principles of self-serving bias. However, it is also possible that participants involved in both supervision and completion have a greater understanding of WOS tasks and their

importance relative to the operation of that vessel and the broader context of the RAN. It is therefore also possible that personnel only involved in completion or supervision of tasks gave tasks lower ratings as they have less of an understanding of the importance of the task within the scope of the vessel's operation. This finding, in conjunction with differences across ranks, brings to light the importance of considering seniority and rank in the context of JTA, as a participant's role or engagement in a task can evidently influence the incumbent's perception of that task (Cucina et al., 2012, 2005; Ford, Ployhart, Lozzi, & Young, 2004).

Limitations

For some participant subgroups, larger sample sizes may have improved the significance and generalisability of the results. However, due to the number of sub-groups, tasks, and ship platforms this was unavoidable in the current population. Some skewness was also observed across ratings of task frequency and was possibly a result of the way in which questions were phrased as participants were not instructed to describe task frequency in relation to any specific time period. Similarly, some questions lacked specificity in relation to the location and phase of operations for which the tasks were completed. Providing a specific time frame for responses may therefore have improved the consistency and accuracy of the results. The time since tasks were last completed by participants is also a concern for this study and subjective JTA in general, as recall of tasks over time may lack integrity. However, as the information surrounding the recall of this type of episodic task-based memory is limited, determining the time frame needed for greatest accuracy is challenging. Finally as subjective job task ratings are yet to be validated against an objective standard; it is difficult to determine whether differences in task perceptions were the result of true or perceived differences. As such, further qualitative information was needed to determine

if differences represent a variation in the way a task is completed or are the result of error resulting from random variation or measurement.

Implications for PES Development

The results of the present study are generally consistent with previous examinations of the relationship between employees' characteristics and job task ratings. This consistency indicates weak effects of systematic bias for subjective ratings of tasks parameters, observing negligible effects of sex, age, job experience, and job tenure. Greater involvement in a task however may affect its perceived importance, possibly as a result of self-serving bias, and similarly a personnel's rank may influence the way tasks are engaged with and perceived. Although it is beyond the scope of this study, it is likely that these results may be generalised to the broader defence force and other occupations involving routine physically demanding tasks. Future studies should aim towards examining the presence of systematic bias within these populations.

In terms of practical implications, this study firstly creates awareness of systematic bias which is currently unexamined in the PES literature. This is important as bias may have negative implications for the accuracy of JTA information and the representativeness of the resulting PES. Awareness of systematic bias allows us to improve the validity of JTA methodologies such that they can be better integrated in the development of more accurate and legally defensible PES. Through awareness of systematic bias in particular, we may be better able to improve participant selection procedures to account for anticipated differences resulting from individual characteristics. One way this could be accomplished is by ensuring that information relating to known sources of systematic bias is collected such that it may be controlled for in analysis. Given the potential of self-serving bias in particular, subjective JTA should rely less on ratings of importance and more on job task variables that are less

prone to bias when identification of critical job tasks. Alternatively, those wishing to use ratings of importance should control for self-serving bias by including participants' level involvement in a task as a co-variate. Considering the paucity of research in this area, and the ease in which demographic and job profile information can be collected, it is recommended that bias is further examined in the context of JTA in order to add to the growing body of literature surrounding the identification of a best practice PES methodology.

3.6 Conclusion

Despite the lack of strong evidence surrounding the presence of systematic bias, participants' demographics and job profile characteristics should still be considered in the development and analysis of accurate JTA. In particular, incumbent characteristics relating to self-serving bias, including job role and the type and level of engagement in a task, should be considered in the conduct and design of all JTA methodologies. Overall, this study contributes to the emerging body of evidence regarding methodological best practice for the conduct of JTA in PDOs and provides evidence surrounding the use of subjective JTA data collection methodology as a valid data collection tool. However, considering the paucity of research in this area, subjective methods may need further validation regarding the capacity to accurately describe physically demanding job tasks.

References

Aguinis, H., Mazurkiewicz, M. D., & Heggestad, E. D. (2009). Using web-based frame-of-reference training to decrease biases in personality-based job anlaysis: an experimental field study. *Personnel Psycholgy, 62*(2), 405–438. doi:10.1111/j.1744-6570.2009.01144.x

Anderson, C. (1994). Can employees physically do the job? *The Human Resource Professional*, *7*(5), 3. Retrieved from http://www.uow.edu.au

Arvey, R. D., Landon, T. E., Nutting, S. M., & Maxwell, S. E. (1992). Development of Physical Ability Tests for Police Officers: A Construct Validation Approach. *Journal of Applied Psychology*, *77*(6), 996–1009. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed

Arvey, R. D., Passino, E. M., & Lounsbury, J. W. (1977). Job analysis results as influenced by sex of incumbent and sex of analyst. *Journal of Applied Psychology*, *62*(4), 411–416. doi:10.1037/0021-9010.62.4.411

Borman, W. C. (1992). Time-Spent Responses as Time Allocation Strategies: Relations with Sales Performance in a Stockbroker Sample. *Personnel Psychology*, *45*(4), 763. Retrieved from http://www.researchgate.net

Blacker S.D., Wilkinson D.M., Rayson M.P. (2009). Gender differences in the physical demands of British Army recruit training. Military Medicine. 174(8): 811-816. Retrieved from www.ncbi.nlm.nih.gov/pubmed

Cucina, J. M., Martin, N. R., Vasilopoulos, L., & Thibodeuax, H. F. (2012). Self-Serving Bias Effects on Job Analysis Ratings. *The Journal of Psychology*, *146*(5), 511–531. doi:10.1080/00223980.2012.656155

Cucina, J. M., Vasilopoulos, N. L., & Sehgal, K. G. (2005). Personality-Based Job Analysis and the Self-Serving Bias. *Journal of Business and Psychology, 20*(2), 275–290. doi:10.1007/s10869-005-8264-2

Dierdorff, E. C., & Wilson, M. A. (2003). A meta-analysis of job analysis reliability. *Journal of Applied Psychology, 88*(4), 635–646. doi:10.1037/001-9010.88.4.635

Gove, W. R., McCorkel, J., Fain, T., & Hughes, M. D. (1976). Response bias in community surveys of mental health: Systematic bias or random noise? *Social Science and Medicine*, *10*(9), 497–502. doi:10.1016/0037-7856(76)90118-9

Grisez, M. L. (1996). *Potential sources of error in job analysis ratings: A preliminary model of the job analysis rating process*. ProQuest, UMI Dissertations Publishing

Gumieniak, R. J., Jamnik, V. K., & Gledhill, N. (2013). Catalog of Canadian fitness screening protocols for public safety occupations that qualify as a bona fide occupational requirement. *Journal of Strength and Conditioning Research*, *27*(4), 1168–1173. doi:10.1519/JSC.0b013e3182667167

Harvey, R. J. (1991). Job analysis. In M. D. Dunnette & L. Hough (Eds.), *Handbook of industrial & organizational psychology* (2nd ed., pp. 71–163). Palo Alto, CA: Consulting Psycholoigsts Press.

Hogan, J. C., & Fleishman, E. A. (1979). An index of the physical effort required in human task performance. *Journal of Applied Psychology*, *64*(2), 197–204. doi:10.1037/0021-9010.64.2.197

Hogan, J. C., Ogden, G. D., Gebhardt, D. L., & Fleishman, E. A. (1980). Reliability and validity of methods for evaluating perceived physical effort. *Journal of Applied Psychology, 65*(6), 672–679. doi:10.1037/0021-9010.65.6.672

Hogan, J., & Quigley, A. M. (1986). Physical Standards for Employment and the Courts. *American Psychologist*, *41*(11), 1193–1217. doi:10.1037/0003-066X.41.11.1193

Hughes, M. A., Ratliff, R. A., Purswell, J. L., & Hadwiger, J. (1989). A Content Validation Methodology for Job Related Physical Performance Tests. *Public Personnel Management*, *18*(4), 487. Retrieved from http://ppm.sagepub.com

Iddekinge, C. H. Van, Putka, D. J., Raymark, P. H., & Eidson, C. E. (2005). Modeling Error Variance in Job Specification Ratings: The Influence of Rater, Job, and Organization-Level Factors. *Journal of Applied Psychology*, *90*(2), 323–334. doi:10.1037/0021-9010.90.2.323

Jamnik, V., Gumienak, R., & Gledhill, N. (2013a). Developing legally defensible physiological employment standards for prominent physically demanding public safety occupations: A Canadian perspective. *European Journal of Applied Physiology*, *113*(10), 2447–2457. doi:10.1007/s00421-012-2569-4

Jamnik, V., Gumienak, R., & Gledhill, N. (2013b). Erratum to: Physiological employment standards II: developing and implementing physical employment standards for safety-related occupations. *European Journal of Applied Physiology*, 1. Retrieved from http://www.scopus.com

Jamnik, V. K., Thomas, S. G., Shaw, J. A., & Gledhill, N. (2010). Identification and characterization of the critical physically demanding tasks encountered by correctional officers. *Applied Physiology Nutrition and Metabolism-Physiologie Appliquee Nutrition Et Metabolisme*, *35*(1), 45–58. doi:10.1139/h09-121

Keyserling, W. M., Herrin, G. D., & Chaffin, D. B. (1980). Isometric strength testing as a means of controlling medical incidents on strenuous jobs. *Journal of Occupational Medicine: Official Publication of the Industrial Medical*

Association, *22*(5), 332–336. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed

Landy, F. J., & Vasey, J. (1991). Job Analysis - The Composition of SME Samples. *Personnel Psychology*, *44*(1), 27–50. doi:10.1111/j.1744-6570.1991.tb00689.x

Larsen, B., & Aisbett, B. (2012). Subjective job task analyses for physically demanding occupations: What is best practice? *Ergonomics*, *55*(10), 1266–1277. http://dx.doi.org/10.1016/j.apergo.2012.05.010

Larsen, B., Graham, T., & Aisbett, B. (2013). A survey to identify physically demanding tasks performed during storm damage operations by Australian State Emergency Services personnel. *Applied Ergonomics*, *44*(1), 128–133.

Lindell, M. K., Clause, C. S., Brandt, C. J., & Landis, R. S. (1998). Relationship Between Organizational Context and Job Analysis Task Ratings. *Journal of Applied Psychology*, *83*(5), 769–776. doi:10.1037/0021-9010.83.5.769

McLennan, P., Groeller, H., Smith, D. L., & Taylor, N. A. S. (2002). Physically deamdning trades: can women tolerater heavy workloads? In N. A. S. Taylor & H. Groeller (Eds.), *Physiological Bases of Human Performance During Work and Exercise* (pp. 255-259). Philadelphia, PA: Churchill Livingstone Elsevier.

Maurer, T. J., & Tross, S. A. (2000). SME Committee vs. Field Job Analysis Ratings: Convergence, Cautions, and a Call. *Journal of Business and Psychology*, *14*(3), 489–499. doi:10.1023/A:1022936418249

Middleton, K., Carstairs, G., Billing, D., Carr, A., Lee-Bates, B., Caputi, P., & Linnane, D. (2014). *Development of Physical Employment Standards for the Royal Australian Navy: Validation of Identified Whole-of-ship Tasks*. Melbourne.

Morgeson, F. P., & Campion, M. A. (1997). Social and Cognitive Sources of Potential Inaccuracy in Job Analysis. *Journal of Applied Psychology*, *82*(5), 627–655. doi:10.1037/0021-9010.82.5.627

Peytcheva, E., & Groves, R. M. (2009). Using variation in response rates of demographic subgroups as evidence of nonresponse bias in survey estimates. *Journal of Official Statistics*, *25*(2), 193–201. Retrieved from https://rti.pure.elsevier.com

Rayson, M. P. (2000). Fitness for work: The need for conducting a job analysis. *Occupational Medicine*, *50*(6), 434–436. Retreived from http://www.tandf.co.uk

Richman, W. L., & Quinones, M. A. (1996). Task Frequency Rating Accuracy: The Effect of Task Engagement and Experience. *Journal of Applied Psychology*, *81*(5), 512. Retrieved from http://www.psycnet.apa.org

Sanchez, J. I. (1990). The Effects of Job Experience on Judgments of Task Importance. In *Annual Conference of the Society for Industrial and Organizational Psychology*,. Miami.

Sanchez, J. I., & Levine, E. L. (2000). Accuracy or Consequential Validity: Which Is the Better Standard for Job Analysis Data? *Journal of Organizational Behavior*, *21*(7), 809–818. doi:10.1002/1099-1379

Schmitt, N., & Cohen, S. A. (1989). Internal Analyses of Task Ratings by Job Incumbents. *Journal of Applied Psychology*, *74*(1), 96–104. doi:10.1037/0021-9010.74.1.96

Sedgwick, P. (2014). Non-response bias versus response bias. *BMJ (Online)*, *348*(apr09 1), g2573–g2573. doi:10.1136/bmj.g2573

Smith, J. E. (1979). Convergence among Data Sources, Response Bias, and Reliability and Validity of a Structured Job Analysis Questionnaire. *Personnel Psychology*, *32*(4), 677–692. doi:10.1111/j.1744-6570.1979.tb02340.x

Supreme Court of Canada, 1999; *British Columbia (Public Service Employee Relations Commission) v British Columbia Government Service Employee' Union.* 3 S.C.R. 3 (Meirorin Decision). Retrieved form http://scc-csc.lexum.com

Taylor, N. A. S., & Groeller, H. (2003). Work-based physiological assessment of physically-demanding trades: A methodological overview. *Journal of Physiological Anthropology and Applied Human Science*, *22*(2), 73–81. doi:10.2114/jpa.22.73

Tipton, M. J., Milligan, G. S., & Reilly, T. J. (2013). Physiological employment standards I. Occupational fitness standards: Objectively subjective? *European Journal of Applied Physiology*, *113*(10), 2435–2446. doi:10.1007/s00421-012-2569-4

Van Iddekinge, C. H., Putka, D. J., Raymark, P. H., & Eidson, C. E. (2005). Modeling Error Variance in Job Specification Ratings: The Influence of Rater, Job, and Organization-Level Factors. *Journal of Applied Psychology*, *90*(2), 323–334. doi:10.1037/0021-9010.90.2.323

Wilkinson D.M., Blacker S.D., Richmond V.L., Horner F.E., Rayson M.P., Spiess A., Knapik J. J. (2011). Injuries and injury risk factors among British Army infantry soldiers. *Injury Prevention*, *17*(6), 381-7. doi:10.1136/ip.2010.028233

CHAPTER 4: COMPARISON OF SUBJECTIVE AND OBJECTIVE JOB TASK ANALYSIS METHODOLOGIES FOR PHYSICALLY DEMANDING OCCUPATIONS

4.1 Abstract

The application of scientifically valid JTA is essential for the development of a legally defensible PES. However, the critical evaluation of specific JTA methodologies (both subjective and objective) is rarely examined. The purpose of this study was therefore to assess the agreement between subjective and objective JTA methods applied to a sample of the Royal Australian Navy (RAN) personnel. 33 critical job tasks were identified and examined using surveys, focus groups (FGs) and task simulations conducted aboard a single RAN vessel. Perceptions of physical effort, importance, duration, frequency and distance travelled for each task were compared between FGs and surveys by examining the similarity of means and the ranking of these means across all tasks. Subjective ratings of physical effort were compared against observations of heart rate, ratings of perceived exertion (RPE) and likert scale ratings of physical demand for each task. Significant relationships were observed between surveys and FGs for rankings of tasks across all task variables. Similarly, significant positive correlations were observed for mean estimates of all variables. Overall, FGs showed a tendency to overestimate ratings of importance and physical effort, but were able to accurately predict vertical and horizontal distance when compared to task-simulations. Furthermore, estimates of physical effort from surveys and FGs correlated significantly with RPE scores subsequent to task completion, but did not correlate with heart rate. These results have important implications for the role of subjective methods in the PES framework and development of a methodological best practice.

4.2 Introduction

JTA provide the foundation for the development of valid, non-discriminatory employment standards (Payne & Harvey, 2010; Rayson, 2000; Tipton, Milligan, & Reilly, 2013). Physically demanding occupations (PDOs), in particular, require valid JTA to ensure that standards accurately represent the physical requirements of the job, contributing towards minimisation of workplace injury, staff turnover and employment-related discrimination (Constable & Palmer, 2000; Jamnik, Gumienak, & Gledhill, 2013). To ensure a PES is accurately linked to the demands of a job role, a scientific evidence-based approach is required (Evidence Act, 1995; Meirorin Decision, 1999). However, current PES guidelines do not provide clear recommendations for methodological and scientific best practice in the conduct of JTA. Furthermore, there is a paucity of research relating to the validation and comparison of different JTA methodologies, impeding the identification of a methodological best practice (Dierdorff & Wilson, 2003; Larsen & Aisbett, 2012).

Current PES development frameworks describe a range of subjective and objective methodologies may be used to conduct JTA, the results of which are used to identify and characterise job tasks that are most representative of a job role (Larsen and Aisbett 2012; Jamnik and Gledhill 1992; Tipton et al., 2013). Given their direct observational nature, *objective* methodologies, which include physical demands analyses and task simulation, provide a valid and accurate JTA tool and may be considered the industry 'gold standard' for the capture of typical job task performance (Bos et al. 2004; Davis et al. 1982). However, these methods are also resource-intensive and time-consuming. The integration of subjective JTA methodologies to capture physical job task data may therefore be beneficial to the PES development framework, providing that these methods can be shown as reliable and valid. *Subjective*

methodologies include surveys, focus groups (FGs) and interviews with subject-matter experts and are typically used to identify and shortlist the most frequently occurring or important job tasks prior to the conduct of physical demands analysis. Although the reliability of subjective methodologies has been reported as acceptable in the context of JTA conduct (Dierdorff and Wilson 2003; Hogan et al. 1980; Viswesvaran et al. 1996), there is an overall paucity of information regarding JTA best practice, especially in relation to integration of subjective methodologies (Dierdorff and Wilson 2003; Larsen and Aisbett 2012; Larsen et al. 2013). Given that objective methodologies may be considered the current 'gold standard' for the capture of physical job task characteristics, it follows that the validity of subjective methods must be assessed through a direct comparison with objective methods.

Given these limitations, this study directly compares the outcomes of two commonly used subjective JTA methods (surveys and FGs) with one objective JTA method (task simulation). These methods were run concurrently within the same population to determine their ability to accurately capture data commonly used to describe physically demanding job tasks. The primary aim of this study was to examine the concordance of task ratings and rankings between surveys and FGs for six commonly used task variables as to determine whether these methods may be used interchangeably to identify the most critical and demanding job tasks. This study also compares estimates and rankings of physical effort measures against direct observations of physical effort collected via task simulations with the aim of directly determining the accuracy of subjective ratings of physical effort.

4.3 Methods

Pilot Survey

Prior to data collection, a pilot survey was administered to identify any formatting and content issues regarding the final survey and FG questions. The selection of relevant job tasks was based on the conduct of a job task inventory with a panel of subject matter experts (Middleton et al., 2014). Participants included 13 experienced Royal Australian Navy (RAN) sailors deemed SMEs on job tasks that could be completed by any staff aboard any class of ship, otherwise known as *whole of ship* (WOS) tasks (age 36.7 ± 9.3 years, range 23 – 54; RAN service 16.5 ± 11.2 years, range 5.5 – 38). A FG was conducted following completion of the pilot survey to gather feedback regarding the design and content of the survey. This pilot revealed that participants were generally satisfied with the content and design of the survey, and agreed that no relevant job tasks had been omitted or were superfluous.

Participants

All personnel gave informed, written consent prior to participation (Appendix C), and all data were kept anonymous in accordance with approval granted by the Australian Defence Human Research Ethics Committee and the Human Research Ethics Committee at a regional University (Appendix A). In total, 498 RAN personnel were recruited to take part in this study via communication with a RAN liaison officer. Prior to participating, participants were briefed regarding the history and ethical requirements of the study, and were provided with an information sheet (Appendix D) and consent form (Appendix C). In order to capture a representative population sample, participants were sampled across all 10 ship platforms deployed within the RAN. Although participants were collected across 11 platforms using surveys, participants posted on coastal surveying ships were excluded from analyses (n = 18), as FGs were not conducted for this platform.

Survey

The survey consisted of 52 questions; including demographics information and questions relating to WOS tasks and movement around the sea-going vessel to which they were last posted to (Appendix D). These WOS tasks included: emergency procedures, fire-fighting, leak-stop and repair, toxic hazard, casualty evacuation and storing procedures, Table 4.1, and were selected such that a baseline PES that applied to all RAN personnel across all classes of ship could be developed.

In total, 409 participants were asked to describe WOS tasks that they had completed by estimating their frequency (number of times per week), duration (in minutes) and distance (both horizontally in meters, and vertically in decks), using an open-answer response format. Participants also rated the importance and physical effort of tasks on a 7-point likert scale where 1 = not very important at all/very easy and 7 = extremely important/very difficult. Participants were able to leave any general comments and describe tasks that may have been left out. The survey was created online using Qualtrics (http://qualtircs.com, Qualtrics, Provo, UT, USA).

Focus Groups

Four to six participants were selected from the larger pool of survey participants to participate in 1 of 10 FGs. Participants were selected such that a range of ranks and job categories were represented, with participants in each group being homogenous to a single class of ship. All participants were encouraged to contribute to the discussion and were asked to communicate with each other in a way that was not linked to the traditional military hierarchy. Participants were instructed to answer as a group where possible as to obtain a consensus response for that vessel, however, if consensus answers could not be achieved a range of answers were accepted which was later averaged across all respondents.

The same moderator administered all FGs to ensure the consistency of delivery and data collection. The content and response format of the FG questions was identical to the survey, with participants asked to describe the same 33 tasks using the same six defining task variables (Appendix F & G). Unless a task was not performed by any participants, the moderator would move through each task one at a time until data were collected for all tasks relevant to that class of ship. At the end of each FG, participants were asked to describe any tasks that were not identified and could be considered a WOS task. Responses were recorded by the moderator using pen and paper, however a portable video camera and digital voice recorder were also used to record all FGs so that they could be transcribed and checked by an external agent.

Table 4.1 *Whole of ship (WOS) tasks performed by RAN personnel*

Task	Task description
1	Perform line handling
2	Participate in the breakdown of a pallet of stores while at sea
3	Participate in storing a vessel while alongside
4	Closing up to action stations
5	Closing up to emergency stations
6	Closing up to leaving stations
7	Conduct a single emergency cable run in 5 minutes
8	Lift and carry a fire extinguisher and enter affected compartment within one minute
9	Lift and carry a fire extinguisher and enter affected compartment within three minutes
10	Lift and carry a fire hose, attach to water main and enter affected compartment in seven
11	As a nozzleman, participate in sustained use of a charged fire hose
12	As the IC, move and support nozzleman
13	As a hose handler, move with and support nozzleman's charged hose
14	As a hose handler/inductor/hydrant, hold hose for an extended period of time
15	Conduct boundary cooling
16	Conduct a fire overhaul
17	Enter affected compartment within three minutes of the alarm in search of casualties
18	Lift and carry as a team of three, a de-watering pump a distance of x metres in 3 minutes
19	Lift and carry as a team of two, a de-smoking fan a distance of x metres
20	Cut 4x4 oregon timber to size using a hand saw
21	As a team of two, carry timber piece from storage area to required site
22	As a team of two, carry acro shoring from storage area to required site and erect by twisting
23	Hammer wedges into place in order to secure vertical and breast pieces
24	Hammer plugs into place in order to maintain hull integrity
25	Carry a tool/repair bag and conduct a permanent pipe repair
26	As a member of team 1, enter affected compartment and spiral upwards to meet team 2
27	As a member of team 2, enter affected compartment and spiral downwards to meet team 1
28	As a member of team 3, enter gas boundary and evacuate casualty
29	As a member of a Team 4, carry a kit bag and repair and clean up toxic hazard
30	Individually or in a team of 2, perform a fire hose lift
31	Individually or in a team of 2, perform a Res-Q-Mate stretcher lift
32	In a team of 6-8, lift and carry a casualty on a Res-Q-Mate stretcher to sick bay
33	In a team of 2, lift and carry a casualty using a RAN safety lift to sick bay

Task Simulation

29 WOS tasks (excluding tasks 2, 5, 27, and 29) were simulated over four days, aboard a single Guided Missile Frigate (FFG) vessel. 12 RAN personnel were selected through correspondence with a liaison officer such that a range of ranks and job categories were represented. Participants were instructed to complete each task as if it were a real exercise, at a natural pace and while wearing the appropriate protective equipment. Horizontal and vertical distances travelled for each task were

predetermined through collaboration with an experienced subject-matter expert and simulation participants. For fire-fighting tasks (tasks 11-14) and boundary cooling tasks (task 15), task duration was set to 10 and 5 minutes, respectively, to reflect a typical exercise aboard that platform.

Prior to simulation all participants completed a beep test and a maximum push up and sit up test, to obtain resting and maximum heart rate. Participants were allowed to rest in between simulated tasks until close to resting heart rate was achieved. Heart rate was observed using a heart rate transmitter attached to each participant, which recorded continuously during task simulations in 5-second intervals. The duration of each task was recorded manually using synchronised watches, and the vertical and horizontal distance was measured by counting decks, and through use of a trundle wheel. The perceived physical effort was determined individually for all participants immediately following the completion of each task. This was measured using the rating of perceived exertion (RPE) scale (Borg, 1998) and through ratings of task difficulty using the same 7-point likert scale implemented in FGs and surveys. A summary of the sources of data collected and analysed in this study are presented in Figure 4.1.

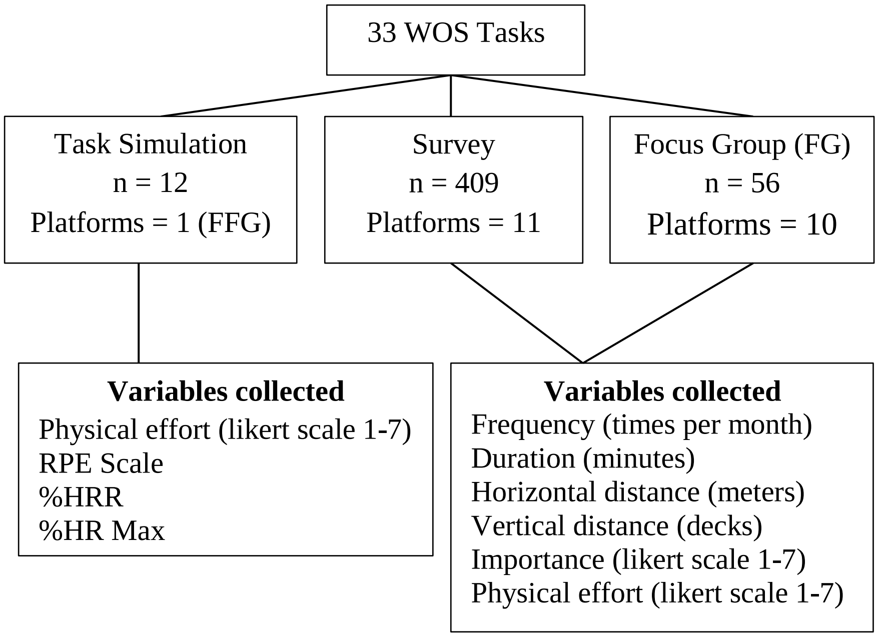

Figure 4.1 Methods and sources of data collection

Statistical Analysis

Data preparation. Participant's data were removed if they had not been to sea in more than 24 months, reported zero months posting on their current platform, did not consent to participate or were not Navy personnel. As large outliers were observed for frequency data, responses were excluded if they exceeded ±3 standard deviations from the mean. Horizontal and vertical distance data were restricted to the length and number of decks respective to each vessel, and one task (Task 4) was removed as an outlier for the analysis of frequency data. Finally, repetitions in data indicating that a participant had not responded thoughtfully to a question (for example responding with the same answer for all tasks) were also removed prior to analysis.

Descriptive statistics were calculated for all survey data across all tasks, and were averaged across participants from 10 ship platforms. For FGs, mean scores were calculated for the same task variables and tasks, by averaging consensus answer

responses across all FG responses. If consensus answer responses were not available, the average score of individual responses was used instead. For all analyses, an alpha level of .05 was used for statistical significance.

Focus groups and Surveys. To observe the agreement between surveys and FGs, means estimates across all tasks were correlated using Pearson's correlation, resulting in six correlations for each variable; frequency, duration, vertical distance, horizontal distance, importance and physical effort. Bland-Altman plots were also generated for each task variable to examine the relationships between mean scores and mean differences between FGs and survey across all tasks.

Given that rankings of tasks are frequently used in PES development to determine those tasks that are most critical or demanding (Anderson, Plecas, & Segger, 2001; Callison & Nussbaum, 2012; Dey & Mann, 2010), tasks were ranked from highest to lowest for all variables (e.g., longest to shortest, most frequent to least frequent, most demanding to least demanding). The percentage of agreement between methods for the top ten ranked tasks for each variable was observed, and overall agreement between rankings was determined using Spearman's correlation analyses given that data were ordinal.

Physical effort and task simulations. Heart rate data were used to calculate two measures of task intensity. Percentage of heart rate reserve (%HRR) was calculated as:

$$\%HRR = \frac{HR_{mean}}{(HR_{max} - HR_{rest})}$$

where %HRR is the percentage of heart rate reserve, HR_{mean} is the mean heart rate across the task, HR_{max} is maximum heart rate achieved during the beep test and HR_{rest} is the minimum resting heart rate. Percentage of max heart rate (%MaxHR) was calculated by dividing the maximum heart rate for a given task by the participant's maximum heart rate achieved in the beep test. Both %HRR and %MaxHR were averaged across

participants for each task and variable. Similarly, RPE and likert ratings of physical effort taken following task completion were also averaged across all participants.

The relationship between mean heart rate, RPE and likert scores, as well as ratings of physical effort from surveys and FGs was examined using Pearson's correlation analyses. The relationship between rankings of tasks by physical demand for FGs and surveys were examined using Spearman's correlations given the ordinal nature of this data. Mean likert scale ratings from task simulations were also directly compared to survey and FG ratings through observation of mean differences between tasks, and mean distances (both vertical and horizontal) and duration were compared between FGs, surveys, and task simulations. Significant differences between measures were examined using repeated measures analysis of variance (ANOVA) tests and Tukey post-hoc multiple comparisons.

4.4 Results

Focus Group vs. Survey

Correlation of means and ranks. Spearman's correlation analyses revealed significant positive relationships between FGs and surveys across tasks, for all variables. When tasks were ranked by their means, significant relationships were observed between methods for all task variables, with task frequency having the strongest relationship between methods. By comparison, rankings of physical effort and duration data yielded moderately strong correlations between methods, while distance and importance data yielded small to moderate correlations, Table 4.2.

When the top ten ranked tasks for each task variable were examined between methods, 80% agreement was observed between methods when ranked by frequency. By contrast, 60% agreement was observed when tasks were ranked by physical effort, 60% agreement was observed for rankings of task frequency, 40% for duration and 50%

for both vertical and horizontal distance, Table 4.2. It should be noted that for FG responses, almost half of the tasks were rated as 'extremely important'; therefore task rankings could not be reliably examined between methods.

Table 4.2

Two-tailed correlations between FGs and surveys for task ratings and rankings for across all task variables

	Physical Effort	Frequency	Duration	Horizontal Distance	Vertical Distance	Importance
Rank						
Spearman's Coefficient	.66**	.81**	.63**	.53*	.57*	.36*
Mean						
Pearson's Coefficient	.79**	.79**	.83**	.60*	.55*	.46*
% agreement in top ten	80%	60%	40%	50%	50%	60%

$^{**}p < .001$, $^{**}p < .05$

Bland-Altman Plots. These plots revealed that FGs tended to rate tasks as more physically demanding, important, and frequently occurring than surveys, however the differences for task frequency were relatively small. By contrast, surveys tended to rate tasks as longer in duration as well as further in both horizontal and vertical distance (Figure 4.2). Notably, two tasks were rated as having considerably longer durations and showed large difference between methods. These tasks included the breakdown of a pallet and storing a vessel while alongside, and were both rated as having considerably longer durations by the FGs compared to the survey. The relationships between the mean difference between methods, and mean estimates of surveys and FGs were observed using a 'line of best fit'. These lines revealed a negative relationship between mean differences and mean task estimates between methods for all variables and revealed that FGs tended to inflate perceptions of tasks rated as higher across both methods. For example, tasks that were perceived as more physically demanding tended

to be rated even more demanding by FGs. This effect was particularly prominent for perceptions of task importance and physical effort.

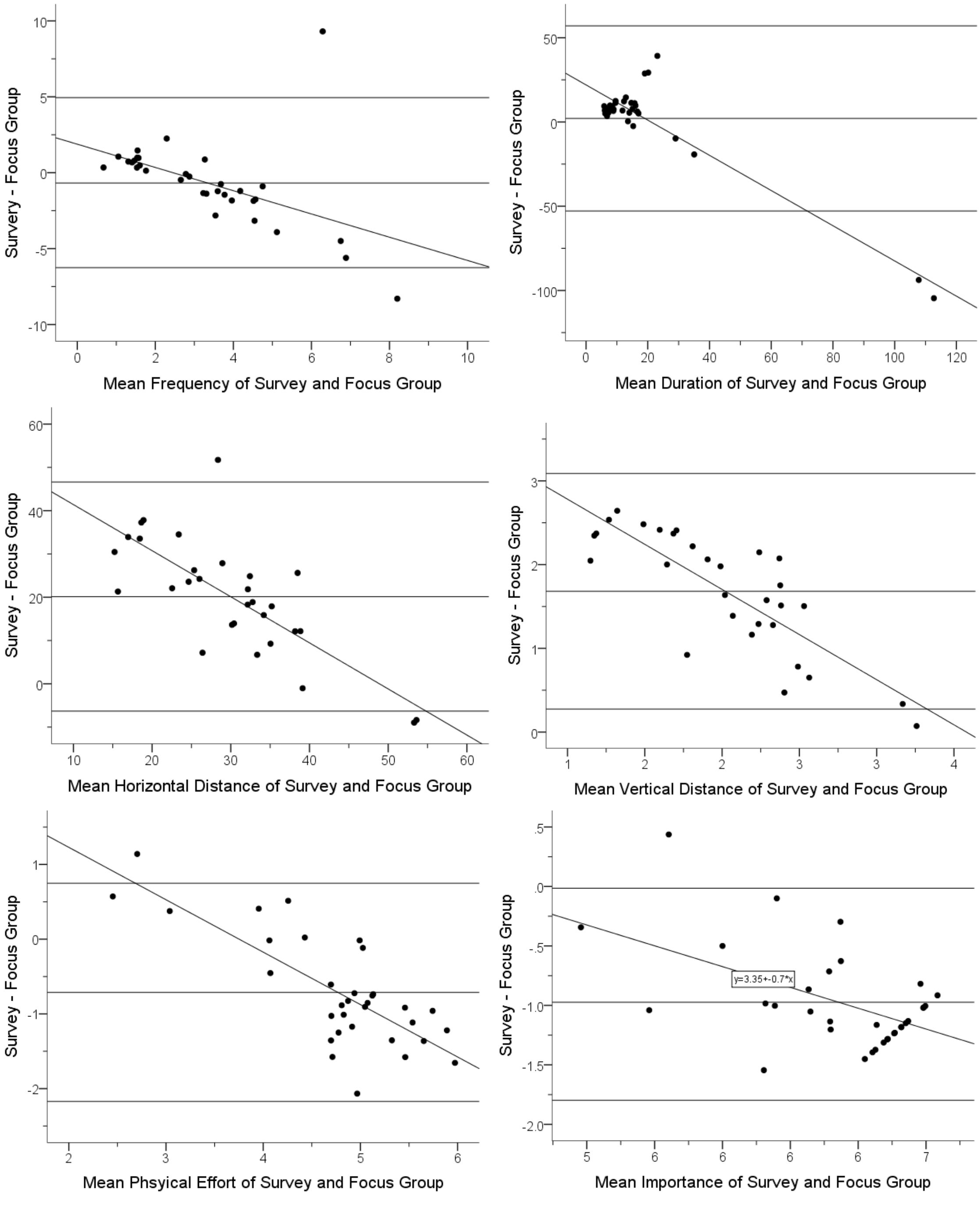

Figure 4.2. Bland-Altman plots comparing FGs and surveys means across all tasks for mean ratings of frequency, duration, distance, physical effort and importance.

Objective vs. Subjective Ratings/Rankings. Pearson's correlations showed that RPE and task simulation ratings were significantly and positively correlated with %HR Max and %HRR data. Significant positive correlations were also found between Survey/FG data and RPE/likert scale data, Table 4.3. Significant correlations were not observed between measures of heart rate and mean ratings of physical effort estimated by surveys and FGs.

When tasks were ranked by their means, comparison of the top ten tasks revealed good consistency between methods, Table 4.4. For example, tasks 10, 11, 28 and 33 occurred in the top 10 most physically demanding tasks for all six measures of physical effort, and 80% similarity was observed between surveys and simulation likert data. When all three mean likert scale responses were compared between methods (simulation, FG and survey), both surveys and FGs showed a tendency to over rate the physical effort of a task, relative to task simulation likert scale responses, Table 4.4.

Table 4.3

Pearson's and Spearman's correlations of physical effort ratings and rankings between survey, FGs and task simulations for measures of physical effort (FFG only)

	%HR Max		**%HRR**		**RPE**		**Likert**		**Survey**	
	Mean	**Rank**	**Mean**	**Rank**	**Mean**	**Rank**	**Mean**	**Rank**	**Mean**	**Rank**
% HRR	.89**	.93**								
RPE	.63**	.56*	.54*	.45*						
Likert	.54*	.50*	.45*	.38*	.97**	.97**				
Survey	.09*	.23	.06	.17	.62**	.77**	.65**	.78**		
FG	.17	.21	.14	.14	.61*	.71**	.64**	.73**	.80**	.73**

$^{**}p < .001$, $^{**}p < .05$

Table 4.4

Comparisons physical effort ratings for top 10 ranked tasks across survey, FGs and task simulations (FFG)

	%HR Max	% HRR	RPE	Likert (n = 12)		Survey (n = 68)		Focus Group (n = 6)	
Rank	**Task**	**Task**	**Task**	**Task**	**Mean**	**Task**	**Mean**	**Task**	**Mean**
1	10	11	11	11	3.45	33	5.93	33, 30, 11, 28, 13	7
2	9	4	32	32	4.75	30	5.92	-	-
3	11	10	33	13	4.23	32	5.91	-	-
4	6	9	13	33	4.58	11	5.68	-	-
5	4	33	28	28	4.50	31	5.56	-	-
6	8	6	10	22	4.33	19	5.44	10, 3	6
7	28	28	30	10	4.15	28	5.42	-	-
8	32	8	31	31	3.90	10	5.40	*	5
9	26/27	26/27	9	19	4.08	13	5.36	-	-
10	33	20	22	9	4.24	14	5.35	-	-

*Tasks 19, 14, 22, 17, 21, 20, 12, 9, 26, 27, 24, 1, 25, 7 had equal ratings

Comparisons of mean estimates of physical effort, duration, vertical and horizontal distance between methods using ANOVAs and multiple comparison tests revealed significant differences for all variables. These comparisons revealed that surveys rated tasks as significantly more demanding and longer, in terms of horizontal distance, vertical distance and duration. By contrast FGs rated tasks as both significantly more demanding and longer than simulations, but had statistically similar ratings of vertical and horizontal distance, Table 4.5.

Table 4.5

Mean physical effort, duration and distance between methods across all tasks (FFG)

	Task Simulation	Survey	Focus Group	ANOVA *p*	Multiple Comparisons
Physical Effort	3.49	4.90	5.07	<.001	Sim < Survey, *p* <.001 Survey < FG, *p* = .319 FG > Sim, *p* <.001
Duration	3.08	14.97	10.48	<.001	Sim < Survey, *p* <.001 Survey > FG, *p* = .050 FG > Sim, *p* <.001
Horizontal Distance	13.40	47.52	11.07	<.001	Sim < Survey, *p* <.001 Survey > FG, *p* = <.001 FG < Sim, *p* = .851
Vertical Distance	0.76	2.62	0.74	<.001	Sim < Survey, *p* <.001 Survey > FG, *p* = <.001 FG < Sim, *p* =.998

4.5 Discussion

The aim of this study was to examine the agreement between subjective and objective JTA methodologies conducted on a population of RAN personnel. These comparisons revealed that subjective ratings of physical effort taken from surveys and FGs correlated significantly with perceived ratings of physical effort taken immediately following task simulations, but were not correlated with objective heart rate measurements. Furthermore, FGs were able to provide accurate estimates of task simulation distance. When surveys and FGs were compared by ratings and rankings of physical effort, frequency, duration, distance and importance, statistically significant correlations were also observed. However, large differences were observed between the top ranked tasks for distance and duration. Overall, these results show that there is scope

for the implementation of subjective methods in a mixed method JTA approach; however some variables may be better examined through objective measurement.

Physical Effort

Correlation analyses revealed that FGs and surveys were significantly related to measures of perceived physical effort following task simulation, including RPE and likert scale ratings. Furthermore, considerable agreement was observed across rankings of physical effort across all measures, including RPE scores and heart rate measures. Surveys in particular had remarkably similar lists of the 10 most physically demanding tasks in comparing task simulation rankings, supporting the accuracy of this method in its ability to determine the relative demand of job tasks. However, there may be some concern that task rankings were not identical; considering the ranking of tasks can play a large role in the identification of criterion job tasks, with even small discrepancies between methods potentially having large implications for the final PES.

By comparison, FGs had a tendency to overestimate the physical demands of tasks relative to simulations, resulting in a ceiling effect by which 12 out of the 33 tasks were rated as ‘extremely difficult’. This made it difficult to create a list of rank-ordered tasks that clearly identified the most demanding tasks, indicating that likert scale ratings may not have been appropriate in the context of FGs. Furthermore, Bland-Altman plots revealed that FGs had a tendency to exaggerate the physical demands of harder tasks, and rated tasks as more demanding than surveys on average. This overestimation of physical effort, may be explained by a groupthink bias in which participants will try to maintain group harmony by minimising within-group disagreement (Turner & Pratkanis, 1998). As a result, participants may have avoided underestimating the

difficulty of critical job tasks in the company of their peers given that all tasks were at least somewhat difficult (Turner & Pratkanis, 1998).

Similarly, surveys also overestimated physical demand compared to task simulations. One explanation may be that participants responded to tasks differently between methods, which is evident in estimates of duration and distance which revealed that both surveys and FGs tended to rate tasks as significantly longer than task simulations. Furthermore surveys, but not FGs, tended to estimate tasks as significantly further in both vertical and horizontal distances, with very large differences observed for some tasks. Although significant agreement was observed between rankings by physical demand, these discrepancies may highlight one limitation of the survey methodology; that it is impossible to ascertain whether participants are responding to the exact task identified by the survey.

Although significant correlations were observed between FGs, surveys, and RPE scores, these relationships were not observed for mean heart rate data. However, this may not indicate poor validity of the subjective methods as heart rate is typically used to measure cardiovascular endurance rather than muscular strength (Bos et al., 2004; Petersen et al., 2010). Given that the majority of the tasks did not involve any sustained whole-body movements, heart rate may not be the most efficacious method of measuring physical demand. Alternatively, localised measures of muscular work may be more appropriate given that whole body exertion may not be the key limiting factor to task performance in the context of the tasks included in this study. Furthermore, given that RPE scores have established validity (Borg, 1970; Borg, 1998) and may provide a relatively more accurate measure of physical demand, significant correlations between RPE scores, and FGs and surveys provide evidence for the validity of these subjective measures.

The comparison of observational and subjective JTA assists in informing appropriate application of these methods to physically demanding job roles. Given the ability of surveys to provide comparatively accurate rankings of tasks by their physical demand, it is worth considering inclusion of this method in the conduct of JTA within PDOs. Specifically, surveys may be used as a method to shortlist the most physically demanding tasks, which may expedite the conduct of JTA by reducing the number of tasks required in task simulations and physical demands analyses.

Vertical and Horizontal Distance

Ratings of horizontal and vertical distance resulted in a similarity of 50% in the top ten ranked tasks, with low to moderate correlations observed between subjective measures for mean estimates and overall task rankings. Similarly Bland-Altman graphs revealed that survey participants consistently rated tasks as further in both vertical distance and horizontal distance. These results support previous literature describing that distances to familiar locations are typically overestimated (McCormack, et al., 2008), suggesting that surveys of JTA have a tendency to overestimate distance. This is further evident regarding comparison of the subset of task simulation data, which revealed that surveys estimated both vertical and horizontal distance as significantly greater than simulations. The use of surveys may therefore be unsuitable for the estimation of distances travelled during tasks, possibly because this type of information may be more difficult to accurately recall (Brown, 1985; Unge et al., 2005; ViikariJuntura et al., 1996). By contrast, FGs had statistically similar estimations of vertical and horizontal distance when compared to task simulation data. One possible explanation is that FGs encouraged participants to consider tasks in greater detail, prompting more accurate recall.

Task Frequency

Similar to past studies that have found support for the efficacy of subjective ratings of frequency (Landy & Vasey, 1991; Dierdorf & Wislon, 2003; Richman & Quinones, 1996), the current study observed significant correlations between subjective methods for ratings and rankings of task frequency. Although there are no objective standards to assist with the validation of subjective task frequency, the high agreement between subjective methods may support the accuracy of task frequency data. However, the low agreement in the top 10 ranked tasks is concerning, especially given that ranking tasks by their relative frequency is a method commonly used in previous literature (Anderson et al., 2001; Callison & Nussbaum, 2012; Doolittle & Daniel, 1989; Hughes, Ratliff, Purswell, & Hadwiger, 1989; Lusa, Louhevaara, & Kinnunen, 1994). Discussions in FGs revealed that task frequency varied considerably depending on the ship's 'phase of operation'. For example, if a ship was in its 'work-up phase', tasks might be completed several times a day, compared to once a month during normal operation. The opportunity to consider this variable in FGs might therefore explain the differences between methods for task rankings. Given this evidence, it is important that future subjective JTA describe the timeframe of tasks in sufficient enough detail to allow participants to respond accurately to the question. By doing so, variables that influence the frequency at which a task is encountered, such as the phase of operation, may be accounted for.

Given the current lack of feasible objective standards to assist in capturing task frequency, this study supports the capture of this data through the use of subjective methods. Surveys in particular may be useful for shortlisting tasks in regards to their frequency prior to conducting a physical demands analysis, providing contextual variables, in particular those pertaining to phase of operation, are considered.

Importance

Ratings of importance showed the lowest consistency between surveys and FGs, with the majority of tasks rated by FGs as ‘extremely important’. Similarly, Bland-Altman plots revealed that FGs inflate ratings of more important tasks relative to surveys. This tendency for FGs to overestimate task importance (and to a lesser extent physical effort), might be the result of self-serving bias, in which participants overstate the importance of tasks that they are personally involved with in order to bolster their perception of self-importance (Cucina, Martin, Vasilopoulos, & Thibodeuax, 2012; Cucina, Vasilopoulos, & Sehgal, 2005). This phenomenon may be reinforced by the groupthink effect, resulting in discussions in which individual participants would not underestimate the importance of a task, so as to minimize disharmony within the group. Consequently, the resulting FG consensus answer of importance will be higher relative to the survey environment in which these phenomena do not operate. It is also worth noting that almost all tasks in this study were considered ‘combat survivability tasks’, and were therefore all essential to the operation of the vessel.

Overall, ratings of importance using likert scales may not have provided a sensitive enough measure in the context of this population. As there are objective methods of estimating task importance, those wishing to determine critical job tasks by identifying their relative importance to the job role, it may be more appropriate to ask participants to rank tasks themselves, rather than estimate their importance, given that a ceiling effect is likely occur.

Recommended Approach

Although the advantages of objective observations are clear, namely the validity in which observations are made, surveys and FGs may provide a more resource efficient alternative in terms of both time and financial costs. Overall, this study has shown that surveys may provide a valid tool for capturing physical effort data when benchmarked

against task simulations, while FGs were useful for gaining greater clarity regarding the context in which tasks are completed and were capable of providing accurate estimates of task distances. By contrast, duration data may be less suitable in the context of subjective JTA methods, given the lack of agreement between methods observed for both ratings and rankings of tasks, while frequency data yielded good consistency between methods and may be accurately estimated using either method. Finally, likert ratings of importance showed little consistency between methods, and may be less useful in identifying critical job tasks within this population.

The results of this study indicate that a mixed method JTA approach, by which subjective methodologies are used to minimise the number of tasks required for resource-heavy observational methodologies and task simulations, is recommended. For example, surveys could be conducted on a representative sample to shortlist the most physically demanding and frequently occurring tasks. Once the tasks that may be included in FGs have been identified, an approximation of how these tasks are completed, such as distance travelled and the environment in which they are typically performed, can be determined. This information could then be used to design a series of task simulations to be performed by incumbent personnel in order to gain information regarding duration and physiological demands. Through this approach, the advantages of subjective methodologies are maximised, reducing resources spent on task simulations and other time-consuming observational methods.

4.6 Conclusion

The need for employees to establish a scientifically defensible PES has brought to light the scarcity of information surrounding the efficacy and comparison of different JTA data collection methodologies. This study has shown that subjective data collection methodologies, including FGs and surveys, may be a viable option for the collection of

specific task-related variables. Although certain variables may be more accurately quantified using observational techniques, subjective methodologies provide a useful and resource-efficient tool for estimating certain task parameters. In particular the ability for surveys to accurately describe a task's relative physical demand, and FGs ability to estimate task distance, may be underestimated, and should be used more readily as a method of shortlisting and characterising tasks for later stages of PES development. Further exploration towards the potential of these methods will enhance the current PES framework, given that subjective methodologies may facilitate a valid and more resource-efficient JTA methodology.

References

Anderson, G. S., Plecas, D., & Segger, T. (2001). Police officer physical ability testing - Re-validating a selection criterion. *Policing-an International Journal of Police Strategies & Management*, *24*(1), 8–31. doi:10.1108/13639510110382232

The Australian Human Rights Commission Act 1986 (Austrl.) Retrieved from https://www.humanrights.gov.au

Bilzon, J. L. J., Scarpello, E. G., Bilzon, E., & Allsopp, A. J. (2002). Generic task-related occupational requirements for Royal Naval personnel. *Occupational Medicine*, *52*(8), 503–510. Retrieved from http://www.scopus.com

Borg, G. (1970). Perceived exertion as an indicator of somatic stress. *Scandinavian Journal of Rehabilitation Medicine*, *2*(2), 92–98. Retrieved from http://www.ncbi.nlm.nih.gov

Borg, G. (1998). *Borg's Perceived exertion and pain scales*. Champaign, IL: Human Kinetics

Bos, J., Mol, E., Visser, B., & Frings-Dresen, M. H. W. (2004). The physical demands upon (Dutch) fire-fighters in relation to the maximum acceptable energetic workload. *Ergonomics*, *47*(4), 446–460. Retrieved from http://www.scopus.com

Brown, S. W. (1985). Time perception and attention: The effects of prospective versus retrospective paradigms and task demands on perceived duration. *Perception & Psychophysics*, *38*(2), 115–124. doi:0.3758/BF03198848

Callison, M. C., & Nussbaum, M. A. (2012). Identification of physically demanding patient-handling tasks in an acute care hospital. *International Journal of Industrial Ergonomics*, *42*(3), 261–267. Retrieved from http://www.scopus.com

Constable, S., Palmer, B. (2000). The Process of Physical Fitness Standards Development. Retrieved from http://dtic.mil/cgi-bin/GetTRDoc?AD=ADA49534

Cucina, J. M., Martin, N. R., Vasilopoulos, L., & Thibodeuax, H. F. (2012). Self-Serving Bias Effects on Job Analysis Ratings. *The Journal of Psychology*, *146*(5), 511–531. doi:10.1080/00223980.2012.656155

Cucina, J. M., Vasilopoulos, N. L., & Sehgal, K. G. (2005). Personality-Based Job Analysis and the Self-Serving Bias. *Journal of Business and Psychology*, *20*(2), 275–290. doi:10.1007/s10869-005-8264-2

Curton, E. D., & Lordahl, D. S. (1974). Effects of attentional focus and arousal on time estimation. *Journal of Experimental Psychology*, *103*(5), 861–867. Retrieved from http://www,psycnet.apa.org/psycinfo

Davis, P. O., Dotson, C. O., & Santa Maria, D. L. (1982). Relationship between simulated fire fighting tasks and physical performance measures. *Medicine and Science in Sports and Exercise*, *14*(1), 65–71. doi:10.1249/00005768-198201000-00013

Dey, A. K., & Mann, D. D. (2010). A complete task analysis to measure the workload associated with operating an agricultural sprayer equipped with a navigation device. *Applied Ergonomics*, *41*(1), 146–149. doi:10.1016/j.apergo.2009.06.005

Dierdorff, E. C., & Wilson, M. A. (2003). A meta-analysis of job analysis reliability. *Journal of Applied Psychology*, *88*(4), 635–646. doi:10.1037/001-9010.88.4.635

Doolittle, T. L., & Daniel, B. (1989). Physical Demands of Bakery Workers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *33*(11), 682–686. doi:10.1177/154193128903301112

Evidence Act 1995 s.79 (Cth) (Austl.). Retrieved from http://www.austlii.edu.au/

Fine, S. A., & Cronshaw, S. F. (1999). *Functional job analysis: a foundation for human resources management*. London: Lawrence Erlbaum Associates

Gledhill, N., & Jamnik, V. K. (1992). Characterization of the physical demands of firefighting. *Canadian Journal of Sport Sciences, 17*(3), 207–213. Retrieved from http://www.scopus.com

Hicks, R. E., Miller, G. W., Gaes, G., & Bierman, K. (1977). Concurrent Processing Demands and the Experience of Time-in-Passing. *The American Journal of Psychology, 90*(3), 431–446. Retrieved from http://uow.summon.serialssolutions.com

Hogan, J. C., Ogden, G. D., Gebhardt, D. L., & Fleishman, E. A. (1980). Reliability and validity of methods for evaluating perceived physical effort. *Journal of Applied Psychology*, *65*(6), 672–679. http://doi.org/10.1037/0021-9010.65.6.672

Hughes, M. A., Ratliff, R. A., Purswell, J. L., & Hadwiger, J. (1989). A Content Validation Methodology for Job Related Physical Performance Tests. *Public Personnel Management*, *18*(4), 487. Retrieved from http://ezproxy.uow.edu.au

Jamnik, V., Gumienak, R., & Gledhill, N. (2013). Developing legally defensible physiological employment standards for prominent physically demanding public safety occupations: A Canadian perspective. *European Journal of Applied Physiology*, *113*(10), 2447–2457. doi:10.1007/s00421-012-2569-4

Jamnik, V. K., & Gledhill, N. (1992). Development of fitness screening protocols for physically demanding occupations. *Canadian Journal of Sport Sciences*, *17*(3), 222–227. Retrieved from http://www.ncbi.nlm.nih.gov

Kirtland, K. A., Porter, D. E., Addy, C. L., Neet, M. J., Williams, J. E., Sharpe, P. A., Ainsworth, B. E. (2003). Environmental measures of physical activity supports:

Perception versus reality. *American Journal of Preventive Medicine, 24*(4), 323–331. doi:10.1016/S0749-3797(03)00021-7

Landy, F. J., & Vasey, J. (1991). JOB ANALYSIS - THE COMPOSITION OF SME SAMPLES. *Personnel Psychology, 44*(1), 27–50. doi:10.1111/j.1744-6570.1991.tb00689.x

Larsen, B., & Aisbett, B. (2012). Subjective job task analyses for physically demanding occupations: What is best practice? *Ergonomics, 55*(10), 1266–1277. Retrieved from http://www.scopus.com

Lindell, M. K., Clause, C. S., Brandt, C. J., & Landis, R. S. (1998). Relationship Between Organizational Context and Job Analysis Task Ratings. *Journal of Applied Psychology*, *83*(5), 769–776. http://doi.org/10.1037/0021-9010.83.5.769

Lusa, S., Louhevaara, V., & Kinnunen, K. (1994). Are the Job Demands on Physical Work Capacity Equal for Young and Aging Firefighters. *Journal of Occupational and Environmental Medicine*, *36*(1), 70–74. Retrieved from http://journals.lww.com

McCormack, G. R., Cerin, E., Leslie, E., Du Toit, L., & Owen, N. (2008). Objective Versus Perceived Walking Distances to Destinations: Correspondence and Predictive Validity. *Environment and Behavior*, *40*(3), 401–425. doi:10.1177/0013916507300560

Middleton, K., Carstairs, G., Billing, D., Carr, A., Lee-Bates, B., Caputi, P., & Linnane, D. (2014). *Development of Physical Employment Standards for the Royal Australian Navy: Validation of Identified Whole-of-ship Tasks*. Melbourne.

Mitchell, J. L., & Driskill, W. E. (1996). Military job analysis: A historical perspective. *Military Psychology*, *8*(3), 119–142. doi:10.1207/s15327876mp0803_1

Payne, W., & Harvey, J. (2010). A framework for the design and development of physical employment tests and standards. *Ergonomics*, *53*(7), 858–871. Retrieved from http://www.scopus.com/

Petersen, A., Payne, W., Phillips, M., Netto, K., Nichols, D., & Aisbett, B. (2010). Validity and relevance of the pack hike wildland firefighter work capacity test: A review. *Ergonomics*, *53*(10), 1276–1285. Retrieved from http://www.scopus.com

Rayson, M. P. (2000). Fitness for work: The need for conducting a job analysis. *Occupational Medicine*, *50*(6), 434–436.

Reed, J. A., Ainsworth, B. E., Wilson, D. K., Mixon, G., & Cook, A. (2004). Awareness and use of community walking trails. *Preventive Medicine*, *39*(5), 903–908. doi:10.1016/j.ypmed.2004.03.013

Richman, W. L., & Quinones, M. A. (1996). Task Frequency Rating Accuracy: The Effect of Task Engagement and Experience. *Journal of Applied Psychology*, *81*(5), 512. Retrieved from http://www.psycnet.apa.org

Supreme Court of Canada, 1999; *British Columbia (Public Service Employee Relations Commission) v British Columbia Government Service Employee' Union.* 3 S.C.R. 3 (Meirorin Decision). Retrieved http://scc-csc.lexum.com

Tipton, M. J., Milligan, G. S., & Reilly, T. J. (2013). Physiological employment standards I. Occupational fitness standards: Objectively subjective? *European Journal of Applied Physiology*, *113*(10), 2435–2446. Retrieved from http://www.scopus.com

Turner, M. E., & Pratkanis, A. R. (1998). Twenty-five years of groupthink theory and research: lessons from the evaluation of a theory. *Organizational Behavior and Human Decision Processes*, *73*, 105–115. doi:10.1006/obhd.1998.2756

Unge, J., Hansson, G. Å., Ohlsson, K., Nordander, C., Axmon, A., Winkel, J., & Skerfving, S. (2005). Validity of self-assessed reports of occurrence and duration of occupational tasks. *Ergonomics*, *48*(1), 12–24. Retrieved from http://www.scopus.com

ViikariJuntura, E., Rauas, S., Martikainen, R., Kuosma, E., Riihimaki, H., Takala, E. P., & Saarenmaa, K. (1996). Validity of self-reported physical work load in epidemiologic studies on musculoskeletal disorders. *Scandinavian Journal of Work Environment & Health*, *22*(4), 251–259. Retrieved from http://www.scopus.com

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative Analysis of the Reliability of Job Performance Ratings. *Journal of Applied Psychology*, *81*(5), 557–574. doi:10.1037/0021-9010.81.5.557

CHAPTER 5: SUMMARY AND CONCLUSION

5.1 Summary of Main Findings

The overarching aim of this thesis was to improve the current PES development framework through examination of commonly used subjective JTA methodologies. The second Chapter of this thesis was aligned with the exploration of common JTA methodologies with the purpose of determining which methodology was most prevalent within PES literature and which method might be considered ‘best practice’. This was addressed by conducting a systematic review of all existing JTA studies related to physically demanding occupations. The results of this review revealed an absence of measures for reliability and validity, with a lack of justification for their chosen methodologies for the majority of studies. Furthermore, very few studies attempted to control for sources of bias that may influence the reliability and generalisability of JTA outcomes. Most notably, a mixture of subjective and objective JTA methods were used, which may be considered best practice in the context of a PES considering the unique advantages and limitations of each methodology. However, given the current paucity of information regarding the validation of comparison of these methodologies it is difficult to conclude which method, or combination of methods, constitutes best practice. Specifically, subjective methods lack safeguards for validity, despite their prevalence throughout JTA literature, while objective observational methods lack evidence for generalisability due to small sample sizes, possibly because of the resource demanding nature of these methods and the demand it plays on employee’s time and safety. Given the potential utility of subjective methods as a resource efficient JTA tool, it was concluded that future research should assess the comparative accuracy of these methodologies. Specifically, surveys and FGs should be validated against a more objective data sample, obtained through task simulations, such that the validity of these

methodologies can be established and used to help improve the current PES development framework.

Chapter 3 provided a closer examination of a subjective survey JTA methodology, through the detection of systematic bias resulting from the participant's demographic and job profile characteristics. Results from this study demonstrated the importance of considering and adjusting for a participant's role within a job task, which was shown to influence the way in which a task is performed or perceived by incumbent personnel. Specifically, a participant's greater involvement with a task may inflate his or her perception of that task's importance. This effect may be the result of self-serving bias in which participants attempt to create a positive self-image by rating tasks they are more invested in as more important to the job role. Although the potential of self-serving bias poses a challenge to future research, this effect may be overcome using appropriate sampling procedures, and careful survey design. Despite this effect, the results of Chapter 3 showed an absence of systematic bias resulting from participant's demographic and employment profile characteristics overall. These findings provide encouraging support for the validity of surveys in application to JTA, which may be used to obtain perceptions of job tasks that are independent of employee characteristics and more closely related to the job role. However, although this paper demonstrates a lack of bias, it does not guarantee the validity of these estimations.

Chapter 4 extends upon the methods used in Chapter 3 by directly contrasting data from two subjective JTA methods (surveys and FGs) against objective task simulation data, which may be considered the current 'gold standard' in relation to the collection of accurate job task data. . This study examined the relative accuracy of these subjective methods and their ability to accurately quantify and rank tasks by a range of job task variables. Comparison of FGs and surveys revealed moderate, significant

agreement for both ratings and rankings by task frequency and physical effort, and small significant relationships for ratings of duration and distance. Most notably, ratings of perceived physical effort taken immediately following task simulations showed significant agreement with survey data, suggesting that subjective JTA methods may be used to quantify the relative physical demands of tasks. Furthermore, ratings of task distance were remarkably similar when compared between FGs and simulations. The results of this study overall support the use of a mixed method JTA approach by which surveys and FGs are used to shortlist the most demanding and frequently occurring tasks and then used to inform the development of physical demands analysis. These findings have important implications for the way in which subjective JTA methods are integrated within the conduct of a scientifically defensible PES.

5.2 Contributions to Physical Employment Standards Development

By demonstrating their prevalent use throughout PDOs (Chapter 2) and providing evidence for their ability to provide a largely unbiased description of physical demanding job tasks (Chapter 3), this thesis supports the use of subjective data collection tools as both a valid and useful JTA methodology. Furthermore, these subjective methods may be able to accurately describe some physical job task characteristics relative to the current 'gold standard' of task simulations (Chapter 4). These findings may be useful in relation to the early stages of PES development as they demonstrate that subjective methods may be used to help identify criterion job tasks and possibly reduce the number of tasks that are subjected to lengthy, expensive and potentially hazardous physical demands analyses. This might be accomplished though implementation of a mixed method, multi-stage JTA by which objective and subjective methods are used selectively to minimise the use of resources, and maximise the accuracy of job task descriptions. Considering the resource intensive nature of

observational JTA methodologies, the identification and adoption of valid subjective JTA methods may save considerable time and money. Furthermore, these methods may provide a safer and more flexible data collection tool that is better able to take in to account the context in which tasks are completed relative to task simulations which often lack ecological validity. If subjective methods can be shown as scientifically valid, they may be used to replace, or reduce the dependence on traditional objective, observational methods that are currently considered the industry 'gold standard' in relation to the capture of physical job task data.

This thesis also discusses whether systematic bias is prevalent in perceptions of commonly collected job task characteristics. Identifying potential sources of systematic bias allows for future employment standards projects to improve participant selection procedures to account for anticipated differences resulting from individual and job related characteristics. Doing so will allow for the development of a PES that is more closely related to the characteristics of a job role, rather than characteristics of personnel within that job role. This could be achieved either statistically, by controlling for differences between participant subgroups for which differences are observed, or by creating standards based only on tasks that are performed similarly between groups and may be applied to both objective and subjective methodologies. This contribution may also extend towards the development of employment standards within non-PDOs which may also be influence by sources of systematic bias including self-serving bias.

Finally, by addressing gaps in the literature surrounding systematic bias, and by exploring the accuracy of commonly used subjective JTA methods, this thesis provides valuable information regarding the scientific defensibility of these methods. Consequently, the outcomes of this study provide evidence for the development of a legally defensible PES and may have a variety of implications for how subjective

methods are treated as scientifically defensible evidence in instances of litigation pertaining to application of a PES.

5.3 Limitations

While there is much to be learned from the findings of this thesis, it is important to address several methodological shortcomings. One major limitation was sample size, which proved to be a recurring obstacle for analyses in Chapters 3 and 4. Considering Data from 11 classes of ship were examined in this study and some classes of ship had very small sample sizes, between group analyses were unable to account for the effect of this variable, and many analyses were forced to treat participants as homogenous in ship class in order to meet sufficient sample sizes. This was particularly relevant in Chapter 4 where means scores and their rankings were based on data from all participants and were not sensitive to difference across ship platform. Sample size was also a limitation for the comparison of objective and subjective data, with only 12 participants from just one ship platform observed for task simulations. Although there is intention to collect more data, this could not be achieved prior to the completion of this thesis. As a result, any conclusions based on comparison between any objective and subjective data must be treated with caution as they may not generalise to the broader RAN population. Further research is therefore needed to confirm any significant effects and improve generalisability of these findings to other physically demanding job roles. Secondly, FGs may have benefitted from a moderator with greater experience with WOS tasks and greater understanding of the RAN work environment. This sometimes posed a problem for dialogue with participants and the interpretation of participants' responses, especially in instances where a consensus answer was not reached. Although this was generally overcome through later examination of video and audio recordings, it

is recommended that all future FGs be run by an experienced moderator or SME to ensure the consistency of responses for ease of data interpretation and analysis. Finally, the wording of some survey questions may have been improved by giving participants more specific details about the nature of the task and a clear time frame for which that task was completed. Asking participants to describe tasks in relation to a more specific time frame may have improved data accuracy by removing variation resulting from the phase of that vessel's operation. Similarly, there may have been some confusion over what task, or component of a task, participants were to respond to, which was evident in comparisons of task distance and duration between survey and task simulation data. Greater clarity regarding task descriptions is therefore recommended for future studies to ensure that participants are responding to the exact task identified by the researcher.

5.4 Recommendations for Future Research

Based on the results of this thesis and gaps in the literature, it is recommended that future research be directed towards assessing and improving the validity and reliability of subjective JTA techniques. As demonstrated in Chapter 4, this could be achieved is by applying a mixed method JTA to a range of PDOs with the aim of validating subjective methods and perceptions of job task variables by comparing these data to an objective standard through task simulations or video analysis of incumbent employees. Specifically, perceptions of task duration and frequency require further validation given the mixed evidence in the literature regarding their accuracy relative to surveys and FGs. Surveys in particular may be beneficial to improving the PES development framework given their demonstrated accuracy in rating and ranking tasks by their physical demand, however the extent to which surveys may be used to describe other physical job task variables such as distance may be useful. Similarly, FGs also

require further investigation as a valid JTA data collection tool as they may assist in the identification of criterion job tasks, and aid in development of more representative task simulations and physical demands analysis.

5.5 Conclusion

This thesis addresses methodological gaps relating to the validation of subjective JTA methodologies and identification of best practice methodology in the context of PES development. This was accomplished through systematic review, and direct comparison of existing subjective and objective JTA outcomes, in addition to detailed exploration of systematic bias that may be inherent in these methods. The findings of this study have contributed towards a better understanding of strengths and weakness of subjective JTA methods, particularly surrounding the use of surveys as a valid measurement tool for ratings of perceived physical effort. By continuing the exploration of valid and unbiased JTA tools, including surveys and FGs, these methods may be adopted more readily and effectively within the development of PESs across a range of PDOs. The results of this research will help to foster the development of a PES methodology with stronger scientific and legal defensibility that is more closely related to the physical demand of a job role, benefitting both the employer and employee.

APPENDICES

Appendix A: Ethics Approval……………………………………….…….……..…...108

Appendix B: Literature Review Summary…………………………..……......….…111

Appendix C: Consent Form……………………………..…..………......…..…...…123

Appendix D: Participant Information Sheet……………….……….……..………..……126

Appendix E: Survey……………………………….…………...……...…...……..……127

Appendix F: Focus Group Questions……….………….…....….........................……149

Appendix G: Focus Group Responses……….……..…..……………….……..…..….…154

Appendix A: Australian Defence Human Research Ethics

JOINT HEALTH COMMAND

ADHREC, CP3-3-036, Campbell Park Offices, PO Box 7912, Canberra BC ACT 2610

2007/1090182
ADHREC/OUT/2014/R17126702

Dr Jace Drain
Human Performance Scientist
Human Protection and Performance Division (HPPD)
Defence Science and Technology Organisation (DSTO)
506 Lorimer St
FISHERMANS BEND VIC 3207

Dear Dr Drain,

AUSTRALIAN DEFENCE HUMAN RESEARCH ETHICS COMMITTEE (ADHREC) ADHREC REQUEST FOR AMENDMENTS, AND RESEARCHER REQUEST FOR MODIFICATIONS TO PROTOCOL 491-07 – ADF PHYSICAL EMPLOYMENT STANDARDS

1. Thank you for submitting your protocol to ADHREC for approval. At the 2 December 2013 meeting, ADHREC had requested further amendments to your protocol including:

   a. updating the participant information sheet to include the potential risk of adverse skin reactions to the use of Fixomull tape; and

   b. clearly articulating the risks and contra-indications associated with the use of gastro-intestinal radio pills.

2. As part of your resubmission to ADHREC, you also included a request to modify your protocol to remove six associate investigators no longer working on the study, and include replacement investigators.

3. ADHREC have considered and approved the amendments and your modifications. Your protocol meets the requirements of the National Health and Medical Research Council's *National Statement on Ethical Conduct in Human Research*, and the Committee wishes you well in continuing with your research. Please note that your next progress report is due on **1 June 2014.**

3. If you have any further queries or concerns, please contact the ADHREC Secretariat.

Yours sincerely,

D.L. BRENNAN
A/Deputy Director Research Ethics
ADHREC

Tel (02) 6266 3807
E-mail: ADHREC@defence.gov.au

7 February 2014

Australian Government
Department of Defence
Defence Science and
Technology Organisation

Minute

Land Division
506 Lorimer Street
Fishermans Bend
VIC 3207

20th March 2014

Executive Secretary ADHREC

ADHREC Protocol 491/07: ADF Physical Employment Standards

The Chief Investigators of the previously approved ADHREC Protocol 491/07 wish to amend the investigators associated with the Physical Employment Standards project;

1. Remove
   - Dr Amelia Carr (Chief investigator)

2. Add
   - WGCDR Marcelle Mitting (Chief investigator)
   - Associate Professor Peter Caputi (Associate investigator)
   - Mr Lachlan Hingley (Associate investigator)
   - Ms Heather Bowes (Associate investigator)
   - Mr Ben Lee-Bates (Associate investigator)
   - Mr Gavin Lenton (Associate investigator)

3. Edit
   - Mr John Sampson to Dr John Sampson (Associate investigator)

Please find attached the modified ADHREC Protocol 491/07. Changes to the original approved protocol are highlighted in yellow in the updated version for ease of review.

If you have any questions please do not hesitate to contact me on 03 9626 8875 or jace.drain@dsto.defence.gov.au.

Kind Regards,

Jace Drain
Human Performance Scientist
Land Division
DSTO

Appendix B: Literature Review Summary

Author, Year	Population	Objective/ Subjective/ Mixed	Method (n)	Method Justification	Variables and Response Format	Methods used for the Selection of Criterion Tasks	Reliability	Validity	Bias Prevention
Anderson et al., 2001	Police Officers	Mixed	Survey (n=67) Shadow (n=121)	Validation of self-report data vs. researcher observations	**Survey** Necessity Frequency Duration Effort Critical incidents **Shadow** Necessity Frequency Duration Critical incidents	Mean, median and ranges as well as % of completion. Tasks ranked by parameters from officers/ researchers. Critical incidents analysed qualitatively	Alternate form	Ecological	-Random Sampling -Demographics compared to population
Arnold et al., 1982	Steel Workers	Subjective	Structured Interviews (n =??)	None	**Interview** Weight – Not Described Distance – Not Described Frequency – Not Described Duration – Not Described	Common themes (e.g. Strength) were identified such that a battery of strength tasks were designed to reflect task demands Tasks were identified that tapped underlying successful performance of important tasks, were safe and were quick and inexpensive to administer	Inter-rater	Content validity	None
Arvey et al., 1992	Police Officers	Mixed	Survey (n=50)	None	**Survey** Importance – 5 point scale Frequency – 5 point scale	Test events were developed as rough representations of tasks indicated as important in the job analysis. Other tasks were included based on tradition or	None	Content validity	None

						practicality for implementation. No analysis of JTA results is presented.			
Ayoub, 1987	Air force	Mixed	Survey 1 (n=??) Survey 2 (n=??) Field interview Task Observation	None	**Survey 1** Physical Demand – 9 point scale **Survey 2** Weights Forces Posture Frequency Duration Strength – 9 point scale Endurance – 9 point scale **Workplace Observation** Weights Forces	Survey 1: Mean rankings were taken for each task and rank ordered Survey 2: 25 shortlisted tasks. Means (SDs) and ranges for weights, duration, frequency and force for resulting 25 tasks were obtained and plotted on histograms A subset of 13 tasks were derived based on frequency distributions	Internal consistency (Pearson's correlations (.23-.57))	Face Validity	None
Bilzon, Scarpello & Allsopp, 2002	Navy	Mixed	Task Analysis (n = ??) Task Simulation (n = 30)	Objective criteria used must be fair valid and justifiable criteria	**Task Analysis** Methods not described **Task Simulation** Duration	Methods used to derive criterion tasks from all tasks not described.	None	None	None
Bos, Mol, Visser & Frings-Dresen, 2007	Fire Fighters	Objective	Shadow (n=222)	None	**Shadow** Heart Rate Reserve Frequency in 24hrs Duration	For tasks performed by more than 5 individuals mean/range number of incidents and duration of incidents reported, Weighted means and extreme values of %HRR during tasks and duration of tasks were taken by looking at the means and 90^{th} percentile	None	Ecological validity	Informed consent

						Top 6 frequently performed tasks reported			
Callison & Nussbaum, 2012	Nurses	Mixed	Shadow (n= Data collected by 7 students) Survey (n=148)	None	**Shadow** Frequency Duration Number of nurses involved **Survey** Physical demand – top 10	Patient handling task ordered by frequency from observations. Completion time, number of nurses involved presented. Most frequent task identified, included in survey and ranked by perceived physical demand. Tasked compared using ANOVA.	Inter-rater (shadowing only)	Ecological validity	Informed consent Age, job title, unit, years of experience, nurses/RNs. Age and years of experience were included as co-variates
Celentano, Nottrodt & Saunders, 2007	Military	Mixed	Survey (n=??) Interviews (n=??) Task Simulation (n=??)	None	Not described	Not described what data was collected in survey or how representative tasks were selected from the large list of physically demanding tasks “The completed questionnaires were reviews, categorized and checked for repeated responses…” pp.482	Insufficient detail	Insufficient detail	Insufficient detail
Considine et al., 1976	Fire Fighters	Subjective	Survey (n=??)	To ensure construct validity of resulting employment test	Not described	Not described “Construct validity of this physical performance battery was derived through procedural steps that included :an extensive task analysis of the fire fighter position and statistical analysis of	Insufficient detail	Insufficient detail	Insufficient detail

						potential test items." pp.7			
Dey & Mann, 2010	Agricultural Spray Operators	Mixed	Survey (n=13) Field Observation (n=3)	Heart rate variability is an indicator of mental workload	**Survey** Importance – Ranked **Field Observation** HR Variability	Tasks were ranked by importance and then monitored using a polar hear-rate monitor to determine mental workload for 10 minutes	None	Content validity (incumbent workers)	None
Doolittle et al., 1998	Army	Mixed	Survey (n=?? SMEs/incumbents) Task Observation/ Videotape Analysis (n=??)	None	**Survey** Difficulty - ?? Frequency – Open Duration – Open **Task Observation** Force used Pace Frequency Movement type Metabolic energy costs	Task grouped in to families based on movement types Tasks deemed by consensus to be the highest difficulty with at least moderate frequency that were critical to the job were considered physically limiting tasks to be included in PET development	Alternate form	Ecological validity Content validity	None
Gledhill & Jamnik, 1992	Fire Fighters	Mixed	Survey1 (n=57 incumbents) Survey 2 (n=60 incumbents) Task Simulation (n=2-12 incumbents)	Literature review of previous methods	**Survey 1** Physically demanding task identified from a list of 57 **Survey 2** Physically demanding tasks ranked **Task Simulation** Objects lifted, object weight, height of lift, body position/action, number of reps, distance of carry, time, frequency, HR, VO2, blood lactate	Tasks were ranked by incumbents by physical effort after a representative list of physically demanding tasks was derived. A ranked ordered list of 8 commonly experienced physically demanding tasks was established Tasks were reviewed by stakeholders Physical demands characterised for each task and	Inter-rater	Content validity Ecological Validity	Randomly selected Informed consent

						average results were documented			
Hughes, 1989	Correctional Officers	Subjective	SME Interviews (n=?) Survey developed after draft (n=886 incumbents)	Most direct evidence is from real time observation, however this is not feasible thus a survey was used	**Survey:** Frequency – 7 Point Scale Duration – 7 Point Scale Average Intensity – 7 Point Scale Peak Intensity – 7 Point Scale	Tasks ranked by frequency, intensity and demand. 8 tasks identified as being physically demanding job required activities Results confirmed conclusions initially made by existing records, observations and interviews with SMEs Data compared against that of interviews, reports and observations	Inter-rater	Content validity	Representative sample of pop. Responses compared by age, gender, ethnic origin, job classification and institution
Jamnik & Gledhill, 1992	Gas Company	Objective	Task Simulation (n=?? incumbents)	Similar approach to previous study (Fleishman, 1979)	**Time-Motion** weight lifted height of lift duration repetitions distance carried **Physical demand characterisation** Posture, angles of force application, VO2, heart rate, average and max number of repetitions, distance and weights of objects	Not described if all or just some tasks were included as a result of the physical demands analysis in the time-motion study Average HR and VO2 max/min was presented for each task Task were categorised based on the frequency, demand and nature of the task as well as the expertise needed and the environment based on results of the PDA	None	Content validity Ecological Validity	Informed consent

						Feedback from supervisors was used to fine tune the categorisation of job tasks			
Jamnik et al., 2010	Correctional Officers	Mixed	Delphi Survey after draft (n=190 incumbents) Task simulation (n=74 incumbents)	A valid framework for conducting PDA	**Survey** Importance – 5 point scale Physical demand – 5 point scale Frequency - 5 point scale **Task Simulation** VO2 Strength Endurance Flexibility HR RPE	Tasks ranked by importance, physical demand, frequency based on average scores from survey. Most important tasks were selected for simulation while meeting min. thresholds for frequency and effort Task then simulated and measured. Differences between task parameters examined using ANOVA	Inter-rater	Content validity Ecological Validity	Informed consent Random selection Rep. sample Age/years of work experience, institution size, disability compared Gender compared (for observations and survey)
Keyserling, 1980	Rubber Plant Production	Subjective	SME Interview (n=??)	None	**Task Simulation** Basic Description Posture maintained Force exerted Locations of hands respective to feet	Biomechanical analysis was carried out on strength tests that represented job tasks based on descriptions in the JTA.	None	Face validity	None
Larsen, Graham & Aisbet, 2013	SES Personnel	Subjective	SME Interview (n=??) Survey (n=362)	Review of previous attempts of JTA- Surveys can capture large number of perspectives	**SME Interview:** Importance – 7 Point Scale **Survey:** Importance - 7 Point Scale Physical demand - 7 Point Scale Frequency – Open Answer Duration – Open Answer Movements – Categorical Choice	Only tasks with physical demand mean, median and mode values greater than 5 were selected for detailed analysis= 12 tasks Tasks were also ranked on importance.	None	Content validity	Informed consent Random Sampling

					Fitness Components – Categorical Choice				
Lusa, Louhevaara & Kinnunen, 1994	Fire Fighters	Subjective	Survey (n=243)	Based on method of previous study (Chahal et al.)	**Survey** Physical Demand - Ranked Frequency - Ranked	Rating of tasks calculated according to votes of respondents among the three most demanding tasks in each dimension of physical work capacity Frequency based on proportion who estimated they had done the a task once in 3 months Chi squared test used to evaluate effects of age/department size on responses	None	Face validity	Effects of age, department size accounted for (no sig. differences found)
Marcinik et al., 1993	U.S Navy Divers	Mixed	Survey (n=72 incumbents) Structured Group Interviews (n=72) Video of Incumbents (n=?)	None	**Survey** Weight Distance Time Examples **Interviews** Most representative tasks - Qualiatitve **Video** Major Muscle groups Skill required	Survey was administered to identify physically demanding tasks performed by fleet divers. Group interviews with divers were conducted to identify a subset of tasks representative of the physical strength and endurance demands of diver's work. Representative tasks were videotaped and analysed in order to select tasks for job performance test construction. Selected tasks were objectively measured	Inter-rater reliability	Ecological validity Content validity	Informed consent

						to develop specific procedures for job performance assessment.			
Metriveir, Gauthier & Gaboriault, 1982	Police Officers	Subjective	Interview with SME (n=2)	None	**Interview** Importance - ??	Tasks were picked and ranked y importance, no more detail given as to how importance was measured.	Inter-rater reliability (weak)	Content Validity (weak)	None
Mueller & Belcher, 2000	Fire Captains	Subjective	Two surveys with incumbents (n=31)	Review of previously used job analysis methods	**Survey 1** Importance - 5 point scale Frequency - 5 point scale Needed at entry - yes/no **Survey 2** Needed at entry - 3 point scale Importance – 4 point scale	Scores combined to form a Criticality Index each survey from for which tasks were then ranked separately for each survey (fire captains vs. supervisors) Rankings and the relationship between rankings of two groups was then tested using Kendall Correlations to form a measure of agreement between SME groups Results found that there was substantial agreements between evaluations of fire captain tasks between incumbents and supervisors	Inter-rater reliability (Strong)	Content validity	Differences between age, time in rank, length of employment were different but not sufficient explanations
Nottrodt & Celentano, 1987	Military	Mixed	Survey (n=?? incumbents) Follow up Interviews (n=??) Observations (n=??)	Follows methods used in past to ensure predictive test validity	**Survey /Interview** Physically demanding tasks identified for a small number of tasks **Observations** Not described	Study does not describe how survey/interview data was treated or what observational data was collected. Results showed that lifting strength was the most important	None	Content Validity	None

						factor in determining job tests.			
Phillips et al., 2012	Fire Fighters	Subjective	Six semi-structured interviews (n=31)	Previously used model of JTA that hasn't been applied to fire fighters	**Interview** Importance - 5 point scale Frequency - Open Duration - Open Distance - Open Action category – Multiple choice Activity category – Multiple choice	Most frequent response for each task domain (as well as range and mode) was presented and ranked Operational importance of the three most physically demanding tasks was then ranked while the duration, frequency, dominant actions and distance of these tasks was characterised.	Inter-rater reliability	Content validity Comparison of results Gledhill and Jamnik study (1992) = Concurrent valdidity	None – Male only sample
Rayson, 1998	Army`	Mixed	Survey (n=??) Interviews (n=??) Video Observations/Biomechanical (n=127)	Tests must be predictive, quantitative, reliable, safe, practicable and non-discriminatory	**Survey** Most physically demanding tasks **Simulation/Video** Frequency Vertical distance Horizontal distance Load mass Max HR VO2	Survey was used to identify most physically demanding tasks using unreported method of quantification. 64 of the 86 tasks identified were simulated and observed through videotapes of tasks performance The distribution of tasks parameter frequencies was documented and the mean/mode values were calculated	None	Content validity	None
Reilly, Zedeck &	Craft Jobs	Subjective	Survey (n=58) Surveys (n=91)	None	**Survey (1 & 2)**	Two surveys were conducted the	Test-retest reliability	Content validity	Common sources of bias

Tenopyr, 1979					Importance - 7 Point Scale Frequency - 7 Point Scale Difficulty - 7 Point Scale	second one being confirmatory of the first. No differences were found between the two surveys SMEs were consulted at every stage of the study. "Based on job analysis data, an initial set of nine physical measures was identified…" pp.264 No further information given.			were considered but not tested
Reilly, Wooler & Tipton, 2006	Beach Lifeguards	Mixed	Structured Interview (n=91 incumbents) Theoretical Analysis Task simulation (n=28)	None	**Structured Interview** Identify most physical demanding, essential and generic activities – Not described **Theoretical Analysis** Duration – Required time **Simulation** Duration Distance	Data were analysed using simple descriptive statistics Once physically demanding tasks were identified, theoretical analysis was conducted to determine restraints on duration. Most demanding strength based and endurance based tasks were reported as well as the average distance run/swam as well as time taken for all simulated activities	None	Ecological Validity Content Validity	None
Sothmann et al., 2004	Fire Fighters	Subjective	Survey (n=353)	Methodology to validate the minimally acceptable muscular strength and endurance levels	**Survey** Importance – Point distribution Frequency – Point distribution **Simulation** Duration - Open	Task ranked by time and importance by distributing 100 point across 11 tasks. Important tasks were simulated and the duration measured	None	Content validity	Informed consent Results compared across gender and age groups by observing differences in

				for effective fire suppression		with consultation with SMEs. Rating and observations of performance were used to characterise criterion job tasks and their necessary parameters			regression slopes
Stevenson et al., 1985	Canadian Armed Forces	Mixed	Literature Survey Task Simulation / Video Analysis(n=132)	None	**Task Simulation** Duration Heart rate Distance	7 key tasks were identified from the literature/ job files and described in terms of their duration and distance, weight carried etc. Tasks were simulated and mean times were recorded. Performance correlated with a battery of fitness tests.	None	Face Validity Ecological Validity	Gender
Viikari-Juntura et al., 1996	Forest Industry	Mixed	Survey (n=2756) Logbook (n=36) Observation (n=36)	Questionnaires and logbooks offer a quick and less costly method with which to collect Information.	**Survey/Logbook** Weight - Open Duration - Open Frequency - Open **Observation** Posture Weight Duration Frequency Motion	Duration or frequencies observed for each task was multiplied by the number of repetitions during one workday. Observations performed by an OP using the portable ergonomic observation method. Spearman correlations for task parameters between methods (survey/logbook and direct observations) was taken	Alternate form reliability	Concurrent validity Ecological Validity	Response format bias

						Correlations higher between logbook and observations than survey and observations			
Vogel, Wright, Patton & Sharp, 1980	U.S Army	Mixed	Task Description Task Simulation (n=200?)	“standards based objectively on actual physiological demands are preferable to subjective determinations of task demands” pp.3	**Task Description** Weights lifted Heights lifted to Distance Estimated caloric costs **Task Simulation** Caloric costs Weights lifted Distance moved VO2 Distance	Selection and description for physically demanding tasks were provided by army service school and then clustered based fitness demands. 4-6 of the most physically demanding tasks in each cluster of tasks was selected by evaluating their physical difficulty determined by observations of task parameters. Strength and stamina needs were estimated for each task based on literature data. Tasks were then simulated so that distance, caloric costs, weights and duration could be confirmed. Demands on task simulation were then converted to a series of aerobic and strength tests.	Alternate form reliability	Ecological Validity Face Validity	States results are gender free, but differences in performance between genders was not examined

Appendix C: Focus Group Information Sheet

Information Sheet

ADF Physical Employment Standards Focus Group

The purpose of this information sheet is to describe the ADF Physical Employment Standards project and to invite you to participate in the study.

Brief description of the study

The Physical Employment Standards research project has been established as part of a broad strategy to manage the issues and costs associated with the high injury rates experienced by ADF personnel. The purpose of the Physical Employment Standards study is to develop objective and valid physical employment assessments for ADF employment categories. The study is conducted over four phases:

Phase 1 involves a survey or a series of trade-task workshops that are directed towards identifying the most physically demanding and high risk trade-tasks for a given ADF employment category.

Phase 2 involves a series of staged in-role field observations that allow the most physically demanding and high risk trade-tasks, identified in Phase 1, to be observed, analysed and quantified. You may be asked to participate in a survey to subjectively analyse and quantify trade-tasks.

Phase 3 involves a series of criterion-task workshops to refine and verify the proposed employment trade-specific physical employment assessment concepts and protocols.

Phase 4 is designed to determine entry and maintenance performance standards for the employment category-specific physical employment assessments.

Your part in the study

You are invited to participate in Phase 1 of this study, the trade-task workshop. These workshops will last for approximately 1 hour and will consist of small groups of about 6 people. You will be asked to participate by answering a series of open ended questions designed to identify the most difficult and important tasks that you perform as part of your duties. There are no right or wrong answers in the discussion but the central

theme will be around which whole of ships tasks are the most physically demanding. For example you may be asked to identify tasks that you'd find physically demanding that you also carry out on a daily basis.

It is important for you to note that your involvement in this study is entirely voluntary and if you choose not to participate there will be no detriment to your career or future health care. Finally, if you choose to participate and later change your mind and wish to withdraw, you may do so without any detriment to your career or future health care.

Risks of participating

It is important to point out to you that there will be a number of risks associated with participation in this study. However, as you would expect, a range of safeguards have been put in place to make sure that these risks will be minimised.

The first risk is that you feel that you are being coerced or forced to participate in this study. In order to minimise the potential for coercion, recruitment of participants will be conducted by a person who is not in your direct chain of command. As mentioned above, you will also be formally notified of your freedom to withdraw at any time should you change your mind about participating in this study.

A video record will be taken during the trade tasks workshop however all participants will remain anonymous in the process of analysing and writing up the results of the study.

Statement of Privacy

There is a separate risk associated with protecting your privacy. There is a risk that the data collected may be used inappropriately within Defence or within the wider community. Examples of this may include using a photo of you without your permission or quoting your individual results in a Defence report. These risks will be reduced by the following:

1. You will be given a code number specific to this study and all data will be 'de-identified' whereby your name will be removed from any sets of records that are used for analysis and reported on to Defence or distributed in the wider community.

2. The information that links your name to your code will be held in confidence by the civilian Principal Researcher.

3. Only group data summaries will be used in any reports

4. Any videos or pictures that are included in the reports will be 'de-identified' by blurring your face or the Civilian Chief Investigator will seek your written permission to use the original image if this is considered desirable.

5. All original data will be kept under lock and key at the Defence Science & Technology Organisation (DSTO) for a period of at least five years.

6. Secure information disposal methods will be used such as document shredding.

7. The data will only be used for the purposes outlined above with your express permission.

Appendix D: Consent Form

CONSENT FORM
ADF Physical Employment Standards

I……………………………………… give my consent to participate in the project mentioned in the subject information sheet on the following basis:

I have had explained to me the aims of this research project, how it will be conducted and my role in it.
I understand that I am participating in this project in a voluntary capacity and can withdraw at any time without penalty or detriment to my career or future health care.
I understand that, as an ADF member, I will be considered to be 'on duty' during participation in the study.
I understand the risks involved as described in the subject information sheet.

I am co-operating in this project on condition that:

- The information I provide will be kept confidential.
- The information will be used only for this project.
- The research results will be made available to me at my request and any published reports of this study will preserve my anonymity.

Video clips and still shots may be used for reports and presentations, therefore if these clips are used you may be identifiable. Please sign and date one of the following options:

OPTION 1: I GIVE permission for the researchers to use video clips or still shots that may identify me.

………………………… ……………………………

Sign Date

OPTION 2: I DO NOT give permission to use video clips or still shots that identify me whether pixelated or not.

………………………… ……………………………

Sign Date

I have been given a copy of the information sheet, consent form and ADHREC's *Guidelines for Volunteers*.

I have signed the consent form, which is co-signed and retained by the chief investigator.

……………………………………… ……………………………………

Participant's signature Principal Researcher's signature

……………………………………… ……………………………………

Printed name Printed name

……………………………………… ……………………………………

Date Date

Appendix E: Survey

Physical Employment Standards for the Royal Australian Navy: Task Quantification Survey

CONSENT

I give my consent to participate in the project mentioned in the subject information sheet on the following basis:

I have had explained to me the aims of this research project, how it will be conducted and my role in it.

I understand that I am participating in this project in a voluntary capacity and can withdraw at any time without penalty or detriment to my career or future health care.

I understand that, as an ADF member, I will be considered to be 'on duty' during participation in the study.

I understand the risks involved as described in the subject information sheet.

I am co-operating in this project on condition that:

- The information I provide will be kept confidential.
- The information will be used only for this project.
- The research results will be made available to me at my request and any published reports of this study will preserve my anonymity.

Video clips and still shots may be used for reports and presentations, therefore if these clips are used you may be identifiable.

Please tick the appropriate box below.

- ◯ I GIVE permission for the researchers to use video clips or still shots that may identify me.
- ◯ I DO NOT give permission to use video clips or still shots that identify me whether pixelated or not.

I have been given a copy of the information sheet, consent form and ADHREC's Guidelines for Volunteers.

Please tick the appropriate box below.

- ◯ I **give** my consent to participate in the project
- ◯ I **do not give** my consent to participate in the project

You are invited to participate in this study, conducted by the Defence Science and Technology Organisation (DSTO) in collaboration with the University of Wollongong (UOW) for the Royal Australian Navy (RAN).

You are one of a select group of personnel that have been chosen to take part in this study. The results from this survey will enable us to understand the requirements of RAN personnel when performing job tasks, identify tasks that are the most demanding, and how these tasks differ across platforms.

All of the following questions throughout this survey relate ONLY to your typical duties on-board the vessel you are CURRENTLY posted to and in your CURRENT role. They apply to operations, exercises and daily activities while at sea. The questions DO NOT relate to training courses or past vessels.

Please answer each question accurately and complete all sections of the survey. Completion of the survey should take approximately 30 minutes.

Demographics

The questions below are designed to gather further information about the participants in this study, for scientific purposes only. Your information will be stored in a secure location, will be de-identified, and will remain confidential.

PMKeys

8579117

Are you male or female?

Male ○ Female ◉

What is your age (in years)?

37

To which category do you belong?

Acoustics Warfare Analyst
Aerospace Engineer
Aviation Support
Aviation Technician Aircraft
Aviation Technician Avionics
Boatswain's Mate
Chaplain
Chef
Clearance Diver
Combat System Operator

What is your rank?

Captain

To which platform are you currently posted?

- (•) FFG
- ( ) FFH
- ( ) Hydrograph
- ( ) Landing Ship (LSH, LSD, LCH)
- ( ) Minehunter
- ( ) Patrol Boat
- ( ) Submarine
- ( ) Support (Sirius, Success)

To which vessel are you currently posted?

- (•) HMAS Darwin
- ( ) HMAS Newcastle
- ( ) HMAS Melbourne
- ( ) HMAS Sydney

What is the phase of operational cycle of the vessel to which you are posted?

- (•) Preparation
- ( ) Work-up
- ( ) Operation
- ( ) Reconstitution
- ( ) Other

How long has it been since you were at sea (in months)?

1

How long have you served in the RAN (in years)?

1

How long have you served in in your current posting (in months)?

1

All of the following questions throughout this survey relate ONLY to your typical duties on-board the vessel you are CURRENTLY posted to and in your CURRENT role. They apply to operations, exercises and daily activities while at sea. The questions DO NOT relate to training courses or past vessels.

If you have feedback regarding a specific question, a comments space is provided at the end of this survey.

Movement on Vessel

The following questions relate to your movement while performing your most common duty on board the vessel to which you are currently posted.

How many times are you typically on watch in a 24-hour period?

1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○

How many decks would you traverse during your most common activity while on watch?

1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○

How many times would you traverse this many decks per watch?

Overall, how long would it take you to traverse this many decks (in minutes)?

In order to complete your most common activity, approximately how far would you need to walk (in metres)?

What type of ladder would you predominantly use when traversing decks?

Sloped ladders	Vertical ladders	Sloped and vertical ladders	Stairwells	Other
○	○	○	○	○

If any, what type of hatch would you predominantly need to open when completing your most common task?

Hatch (manual)	Hatch (spring/hydraulic assisted)	Door	Not necessary to open hatches
○	○	○	○

Replenishment at Sea (RAS)

Have you participated in RAS?

- (•) Yes
- ( ) No

Line Handling

Have you participated in line handling?

- (•) Yes
- ( ) No

Breakdown of a pallet of stores

Have you participated in the breakdown of a pallet of stores?

- (•) Yes
- ( ) No

How many times would you participate in line handling per month?

☐

When performing line handling, how many people are most commonly on a line at any one time?

☐

Are the demands of the task the same for each person handling the line?

- ○ Yes
- ○ No - Higher demand for people at the front of the line
- ○ No - Higher demand for people in the middle of the line
- ○ No - Higher demand for people at the back of the line

What would be the most common duration that you would perform line handling (in minutes)?

☐

How would you rate the importance of line handling relative to mission fitness?

1 - Not at all important	2 - Low importance	3 - Slightly important	4 - Neutral	5 - Moderately important	6 - Very important	7 - Extremely important
○	○	○	○	○	○	○

How would you rate the physical effort of line handling?

1 - Very easy	2 - Easy	3 - Somewhat easy	4 - Neutral	5 - Somewhat difficult	6 - Difficult	7 - Very difficult
○	○	○	○	○	○	○

What posture would you adopt when line handling?

- ○ Sitting
- ○ Kneeling (two knees)
- ○ Kneeling (one knee)
- ○ Squatting
- ○ Standing

How many times would you participate in breaking down a pallet of stores per month?

[ ]

How many people are most commonly involved in the breakdown of a pallet of stores?

[ ]

What would be the most common duration that you would be involved in the breakdown of a pallet of stores (in minutes)?

[ ]

How many items would you most commonly handle in the time frame specified above?

[ ]

What is the most common mass of an item you would hold during the breakdown of a pallet of stores (in kilograms)?

[ ]

When given an item, you are required to:

- ○ Pass the item
- ○ Carry and pass the item
- ○ Carry the item to the storage location

How would you rate the importance of breaking down a pallet of stores relative to mission fitness?

1 - Not at all important	2 - Low importance	3 - Slightly important	4 - Neutral	5 - Moderately important	6 - Very important	7 - Extremely important
○	○	○	○	○	○	○

How would you rate the physical effort of breaking down a pallet of stores?

1 - Very easy	2 - Easy	3 - Somewhat easy	4 - Neutral	5 - Somewhat difficult	6 - Difficult	7 - Very difficult
○	○	○	○	○	○	○

Storing a Vessel (Alongside)

Have you participated in storing the vessel to which you are currently posted?

- (•) Yes
- ( ) No

How many times would you participate in storing the vessel to which you are currently posted per month?

[ ]

How many people are most commonly involved in storing the vessel?

[ ]

What would be the most common duration that you would be involved in storing the vessel (in minutes)?

[ ]

How many items would you most commonly handle in the time frame specified above?

[ ]

What is the most common mass of an item you would hold when storing the vessel (in kilograms)?

[ ]

When given an item, you are required to:

- ( ) Pass the item
- ( ) Carry and pass the item
- ( ) Carry the item to the storage location

How would you rate the importance of storing the vessel relative to mission fitness?

1 - Not at all important	2 - Low importance	3 - Slightly important	4 - Neutral	5 - Moderately important	6 - Very important	7 - Extremely important
( )	( )	( )	( )	( )	( )	( )

How would you rate the physical effort of storing the vessel?

1 - Very easy	2 - Easy	3 - Somewhat easy	4 - Neutral	5 - Somewhat difficult	6 - Difficult	7 - Very difficult
( )	( )	( )	( )	( )	( )	( )

Combat Survivability

The questions that follow relate to tasks that have been identified as Whole of Ship tasks through focus groups with experienced RAN personnel. Please review the following tasks and rate them as per below.

Frequency
What is the average number of times you perform this task per month?

Duration
What is the most common duration for which the task is performed (in minutes)?
Although some task description include Fleet Standard Times, please specify the actual duration.

Vertical Distance
How many decks would you most commonly traverse during this task?

Horizontal Distance
How far would you most commonly traverse during this task (in metres)?

Importance
We recognise that almost every task is important, but we want you to consider how important this task is relative to mission fitness.

Physical Effort
How much physical effort is required to perform this task?

Emergency Situations

Have you performed or supervised the following tasks in your current role on the platform to which you are currently posted (please specify)?

	Yes (Performed)	Yes (Supervised)	Yes (Performed and Supervised)	No
Closing up to Action Stations	○	○	○	○
Closing up to Emergency Stations	○	○	○	○
Closing up to Leaving Stations	○	○	○	○
Conduct a single emergency cable run	○	○	○	○

Closing up to Action Stations

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

Closing up to Emergency Stations

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

Closing up to Leaving Stations

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

Conduct a single emergency cable run

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

If any of the above tasks are completed differently to how they are described above, please specify below.

Firefighting

What is the most common location that you would perform firefighting tasks on the vessel to which you are currently posted?

Have you performed or supervised the following tasks in your current role on the platform to which you are currently posted (please specify)?

	Yes (Performed)	Yes (Supervised)	Yes (Performed and Supervised)	No
Lift and carry a fire extinguisher a distance of x metres and enter affected compartment within one minute of the alarm being raised (FAA).	○	○	○	○
Lift and carry a fire extinguisher a distance of x metres and enter affected compartment in three minutes (BA-P).	○	○	○	○
Lift and carry a fire hose a distance of x metres, attach to water main and enter affected compartment in seven minutes (BA-H).	○	○	○	○
As a nozzleman, participate in sustained use of a charged fire hose.	○	○	○	○
As the IC, move and support nozzleman.	○	○	○	○
As a hose handler, move with and support nozzleman's charged hose.	○	○	○	○
As a Hose Handler/Inductor/Hydrants, hold hoses for an extended period of time.	○	○	○	○
Conduct boundary cooling.	○	○	○	○
Conduct fire overhaul.	○	○	○	○

Lift and carry a fire extinguisher a distance of x metres and enter affected compartment within one minute of the alarm being raised (FAA).

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort
1	1	1	1	Not very important at all	Very easy

Lift and carry a fire extinguisher a distance of x metres and enter affected compartment within three minutes (BA-P).

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort
1	1	1	1	Not very important at all	Very easy

Lift and carry a fire hose a distance of x metres, attach to water main and enter affected compartment in seven minutes (BA-H).

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort
1	1	1	1	Not very important at all	Very easy

As a nozzleman, participate in sustained use of a charged fire hose.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort
1	1	1	1	Not very important at all	Very easy

As the IC, move and support nozzleman.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort
1	1	1	1	Not very important at all	Very easy

As a hose handler, move with and support nozzleman's charged hose.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort
1	1	1	1	Not very important at all	Very easy

As a Hose Handler/Inductor/Hydrants, hold hoses for an extended period of time.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort
1	1	1	1	Not very important at all	Neutral

Conduct boundary cooling.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

Conduct fire overhaul.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

Shoring

What is the most common location that you would perform shoring tasks on the vessel to which you are currently posted?

Have you performed or supervised the following tasks in your current role on the platform to which you are currently posted (please specify)?

	Yes (Performed)	Yes (Supervised)	Yes (Performed and Supervised)	No
Enter affected compartment within three minutes of the alarm being raised in search of casualties.	○	○	○	○
Lift and carry as a team of three, a de-watering pump a distance of x metres in 3 minutes.	○	○	○	○
Lift and carry as a team of two, a de-smoking fan a distance of x metres.	○	○	○	○
Cut 4x4 Oregon timber to size using a hand saw.	○	○	○	○
As a team of two, carry timber piece from storage area to required site.	○	○	○	○
As a team of two, carry acro shoring from storage area to required site and erect by twisting.	○	○	○	○
Hammer wedges into place in order to secure vertical and breast pieces.	○	○	○	○
Hammer plugs into place in order to maintain hull integrity.	○	○	○	○
Carry a tool/repair bag and conduct a permanent pipe repair.	○	○	○	○

Enter affected compartment within three minutes of the alarm being raised in search of casualties.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

Lift and carry as a team of three, a de-watering pump a distance of x metres in 3 minutes.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

Lift and carry as a team of two, a de-smoking fan a distance of x metres.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

Cut 4x4 Oregon timber to size using a hand saw.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

As a team of two, carry timber piece from storage area to required site.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

As a team of two, carry acro shoring from storage area to required site and erect by twisting.

	Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

Hammer wedges into place in order to secure vertical and breast pieces.

	Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

Hammer plugs into place in order to maintain hull integrity.

	Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

Carry a tool/repair bag and conduct a permanent pipe repair.

	Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

If any of the above tasks are completed differently to how they are described above, please specify below.

Toxic Hazard

What is the most common location that you would perform toxic hazard (casualty search/evacuation, cleanup) tasks on the vessel to which you are currently posted?

Have you performed or supervised the following tasks in your current role on the platform to which you are currently posted (please specify)?

	Yes (Performed)	Yes (Supervised)	Yes (Performed and Supervised)	No
As a member of Team 1 (Search) and carrying two spare ELSRDs, enter affected compartment and spiral upwards to meet Team 2 placing ELSRD on first casualty within four minutes.	○	○	○	○
As a member of Team 2 (Search) and carrying two spare ELSRDs, enter gas boundary and spiral downwards to meet Team 1 placing ELSRD on first casualty within four minutes.	○	○	○	○
As a member of Team 3 (Casualty Evacuation) and carrying two spare ELSRDs, enter gas boundary and evacuate casualty.	○	○	○	○
As a member of a Team 4 (Repair Team), carry a kit bag with tools and repair and clean up toxic hazard.	○	○	○	○

As a member of Team 1 (Search) and carrying two spare ELSRDs, enter affected compartment and spiral upwards to meet Team 2 placing ELSRD on first casualty within four minutes.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

As a member of Team 2 (Search) and carrying two spare ELSRDs, enter gas boundary and spiral downwards to meet Team 1 placing ELSRD on first casualty within four minutes.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

As a member of Team 3 (Casualty Evacuation) and carrying two spare ELSRDs, enter gas boundary and evacuate casualty.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

As a member of a Team 4 (Repair Team), carry a kit bag with tools and repair and clean up toxic hazard.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

If any of the above tasks are completed differently to how they are described above, please specify below.

Casualty Evacuation

What is the most common location that you would perform casualty evacuation tasks on the vessel to which you are currently posted?

Have you performed or supervised the following tasks in your current role on the platform to which you are currently posted (please specify)?

	Yes (Performed)	Yes (Supervised)	Yes (Performed and Supervised)	No
Individually or in a team of 2, perform a fire hose lift.	○	○	○	○
Individually or in a team of 2, perform a Res-Q-Mate/Paraguard stretcher lift.	○	○	○	○
In a team of 6-8, lift and carry a casualty on a Res-Q-Mate/Paraguard stretcher from site of injury x metres to first aid post/sick bay.	○	○	○	○
In a team of 2, lift and carry a casualty using a RAN Safety Lift (i.e. fore-aft carry) from site of injury x metres to first aid post/sick bay.	○	○	○	○

Individually or in a team of 2, perform a fire hose lift.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

Individually or in a team of 2, perform a Res-Q-Mate stretcher lift.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

In a team of 6-8, lift and carry a casualty on a Res-Q-Mate stretcher from site of injury x metres to first aid post/sick bay.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

In a team of 2, lift and carry a casualty using a RAN Safety Lift (i.e. fore-aft carry) from site of injury x metres to first aid post/sick bay.

Frequency (per month)	Duration (minutes)	Vertical distance (decks)	Horizontal distance (metres)	Importance	Physical Effort

If any of the above tasks are completed differently to how they are described above, please specify below.

We would like to collect additional information on your **sleep behaviour** and **exercise** habits while at sea. This will enable us to understand possible factors that need to be considered when developing Physical Employment Standards and Assessments.

Sleep Patterns

In a 24 hour period at sea, how many times do you go to sleep?

[ ]

On average, how many hours of sleep do you get each time you go to sleep at sea?

[ ]

The amount of sleep you get while at sea is...

- ○ Perfectly adequate
- ○ Adequate
- ○ Slightly adequate
- ○ Neutral
- ○ Slightly inadequate
- ○ Inadequate
- ○ Totally inadequate

The quality of sleep you get while at sea is...

- ○ Very good
- ○ Good
- ○ Fair
- ○ Poor
- ○ Very poor

In a 24 hour period at home, how many times do you go to sleep?

[ ]

On average, how many hours of sleep do you get each time you go to sleep at home?

[ ]

Physical Exercise

Do you participate in structured physical activity (in the presence of a PT or MFL) while at sea?

- ◯ Yes
- ◯ No

Do you participate in self-directed physical activity (without a PT or MFL present) while at sea?

- ◯ Yes
- ◯ No

Thank you for participating in this research. Once you click 'next', your responses will be submitted and you will not be able to amend any of your answers.

Please provide any general comments below prior to submission.

We thank you for your time spent taking this survey. Your response has been recorded.

Appendix F: Focus Group Questions

1	**Emergency Situations**	Closing up to action stations.					
		F	D	VD	HD	Imp	PE
2		Closing up to emergency stations.					
		F	D	VD	HD	Imp	PE
3		Closing up to leaving ship stations.					
		F	D	VD	HD	Imp	PE
4		Conduct a single emergency cable run in 5 minutes.					
		F	D	VD	HD	Imp	PE
5	**Replenishment at Sea**	Perform line handling.					
		F	D	VD	HD	Imp	PE
6		Participate in the breakdown of a pallet of stores while at sea.					
		F	D	VD	HD	Imp	PE
7	**Storing**	Participate in storing a vessel while alongside.					
		F	D	VD	HD	Imp	PE

<table>
<tr><td rowspan="3">8</td><td rowspan="22">**Firefighting**</td><td colspan="6">Lift and carry a fire extinguisher a distance of x metres and enter affected compartment within one minute of the alarm being raised (FAA).</td></tr>
<tr><td>F</td><td>D</td><td>VD</td><td>HD</td><td>Imp</td><td>PE</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">9</td><td colspan="6">Lift and carry a fire extinguisher a distance of x metres and enter affected compartment within three minutes (BA-P).</td></tr>
<tr><td>F</td><td>D</td><td>VD</td><td>HD</td><td>Imp</td><td>PE</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">10</td><td colspan="6">Lift and carry a fire hose a distance of x metres, attach to water main and enter affected compartment in seven minutes (BA-H).</td></tr>
<tr><td>F</td><td>D</td><td>VD</td><td>HD</td><td>Imp</td><td>PE</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">11</td><td colspan="6">As a nozzleman, participate in sustained use of a charged fire hose.</td></tr>
<tr><td>F</td><td>D</td><td>VD</td><td>HD</td><td>Imp</td><td>PE</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">12</td><td colspan="6">As the IC, move and support nozzlemen.</td></tr>
<tr><td>F</td><td>D</td><td>VD</td><td>HD</td><td>Imp</td><td>PE</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">13</td><td colspan="6">As a hose handler, move with and support nozzleman’s charged hose.</td></tr>
<tr><td>F</td><td>D</td><td>VD</td><td>HD</td><td>Imp</td><td>PE</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">14</td><td colspan="6">As a Hose Handler/Inductor/Hydrants, hold hoses for an extended period of time.</td></tr>
<tr><td>F</td><td>D</td><td>VD</td><td>HD</td><td>Imp</td><td>PE</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td colspan="6">Conduct boundary cooling.</td></tr>
</table>

15		F	D	VD	HD	Imp	PE
16		Conduct fire overhaul.					
		F	D	VD	HD	Imp	PE
17	**Shoring**	Enter affected compartment within three minutes of the alarm being raised in search of casualties.					
		F	D	VD	HD	Imp	PE
18		Lift and carry as a team of three, a de-watering pump a distance of x metres in 3 minutes.					
		F	D	VD	HD	Imp	PE
19		Lift and carry as a team of two, a de-smoking fan a distance of x metres.					
		F	D	VD	HD	Imp	PE
20		Cut 4x4 Oregon timber to size using a hand saw.					
		F	D	VD	HD	Imp	PE
21		As a team of two, carry timber piece from storage area to required site.					
		F	D	VD	HD	Imp	PE
22		As a team of two, carry acro shoring from storage area to required site and erect by twisting.					
		F	D	VD	HD	Imp	PE
23		Hammer wedges into place in order to secure vertical and breast pieces.					
		F	D	VD	HD	Imp	PE

24		Hammer plugs into place in order to maintain hull integrity.					
		F	D	VD	HD	Imp	PE
25		Carry a tool bag and conduct a permanent pipe repair.					
		F	D	VD	HD	Imp	PE
26	**Toxic Hazard**	As a member of Team 1 (Search) and carrying two spare ELSRDs, enter affected compartment and spiral upwards to meet Team 2 placing ELSRD on first casualty within four minutes.					
		F	D	VD	HD	Imp	PE
27		As a member of Team 2 (Search) and carrying two spare ELSRDs, enter gas boundary and spiral downwards to meet Team 1 placing ELSRD on first casualty within four minutes.					
		F	D	VD	HD	Imp	PE
28		As a member of Team 3 (Casualty Evacuation) and carrying two spare ELSRDs, enter gas boundary and evacuate casualty.					
		F	D	VD	HD	Imp	PE
29		As a member of a Team 4 (Repair Team), carry a kit bag with tools and repair and clean up toxic hazard.					
		F	D	VD	HD	Imp	PE
30	**Casualty Evacuation**	Individually or in a team of 2, perform a fire hose lift.					
		F	D	VD	HD	Imp	PE
		Individually or in a team of 2, perform a Res-Q-Mate stretcher lift.					
		F	D	VD	HD	Imp	PE

31							
		In a team of 6-8, lift and carry a casualty on a Res-Q-Mate stretcher from site of injury x metres to first aid post/sick bay.					
		F	D	VD	HD	Imp	PE
32							
		In a team of 2, lift and carry a casualty using a RAN Safety Lift (i.e. fore-aft carry) from site of injury x metres to first aid post/sick bay.					
		F	D	VD	HD	Imp	PE
33							
Other Tasks (Describe)							
		F	D	VD	HD	Imp	PE
		F	D	VD	HD	Imp	PE
		F	D	VD	HD	Imp	PE
		F	D	VD	HD	Imp	PE

Appendix G: Focus Group Response Sheet

Questions

- What is the average number of times you perform this task per month?
- What is the most common duration for which the task is performed (in minutes)?
- How many decks would you most commonly traverse during this task?
- How many meters would you most commonly traverse during this task?
- On a scale of 1 to 7 how would you rate the importance of this task relative to mission fitness?

1	2	3	4	5	6	7
Not very important at all	Low importance	Slightly important	Neutral	Moderately important	Very important	Extremely important

- On a scale from 1 to 7 how would you rate the physical effort required of this task?

1	2	3	4	5	6	7
Very easy	Easy	Somewhat easy	Neutral	Somewhat difficult	Difficult	Very difficult