Reliability and Validity of the WorkHab Functional Capacity Evaluation

Carole L. James

MHSc(OT), BSc(OT), Dip COT

This thesis is submitted in fulfilment of the requirements for the

Degree of Doctor of Philosophy (Occupational Therapy)

School of Health Sciences

University of Newcastle, NSW, 2308

Australia

February 2011

Statement of Originality

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying subject to the provisions of the Copyright Act 1968.

Thesis by publication
Acknowledgement of Authorship

I hereby certify that this thesis is in a form of a series of published papers of which I am a joint author. I have included as part of the thesis a written statement from each co-author, endorsed by the Faculty of Health Assistant Dean (Research Training), attesting to my contribution to the joint publications. These can be found in Appendix 1.

Signed: _____

Carole James

Date: _____

Acknowledgements

I wish to acknowledge and thank all the participants: the injured workers; students and staff of the university; occupational therapists; and other health professionals, who participated in this research without whom, this study would not have been possible. I would also like to thank all the health professionals who contributed along my journey, as participants in the study, and as peers and colleagues in providing advice and expertise.

I acknowledge the financial support provided by the University of Newcastle with an early career researcher grant and an equity fellowship grant which made this study possible.

A special thank you goes to my supervisors Dr Lynette Mackenzie and Professor Mike Capra. Both have been very generous with their time, provided much enthusiasm, scholarly advice, direction and support throughout this journey.

Thank you to the research assistants who have assisted in various stages of this project and also to my colleagues within the School of Health Sciences at the University of Newcastle, for your support and encouragement.

A heart sent thank you goes to my family and friends who have provided much practical assistance and support. A special thank you goes to my parents for their continuous support and faith in all my endeavours; to my husband Bernie, for his practical and emotional support and his belief in my abilities and to Zoe, my daughter, for her patience and understanding during this research journey.

Publications arising from work presented in this thesis

Refereed Journal publications:

1. James, C., Mackenzie, L. and Higginbotham, N. (2007). "Health Professionals' Attitudes and Practices in relation to Functional Capacity Evaluations." Work **29**(2): 81-88.
2. James, C. and Mackenzie, L. (2009). "Health Professional's Perceptions and Practices in Relation to Functional Capacity Evaluations: Results of a Quantitative Survey." Journal of Occupational Rehabilitation **19**(2): 203-211.
3. James, C. and Mackenzie, L. (2009). "The clinical utility of functional capacity evaluations: the opinion of health professionals working within occupational rehabilitation." Work **33**(3): 231-9.
4. James, C., Mackenzie. L. and Capra, M. (2010). "Test-retest reliability of the manual handling component of the WorkHab functional capacity evaluation in healthy adults." Disability and Rehabilitation **32**(22): 1863-1869.
5. James, C., Mackenzie, L. and Capra, M. (early online). Inter and Intra- rater reliability of the Manual Handling component of the WorkHab Functional Capacity Evaluation. Disability and Rehabilitation DOI:10.3109/09638288.2010.548896 (posted online: January 19, 2011).
6. James, C., Mackenzie, L. and Capra, M. (under review). "Content Validity of the WorkHab Functional Capacity Evaluation." Submitted for publication in Disability and Rehabilitation
7. James, C., Mackenzie, L. and Capra, M. (under review). "Physiological correlates of functional capacity evaluations: finding the safe maximal lift." submitted for publication in Archives of Physical Medicine and Rehabilitation

Conference paper presentations:

8. James, C., Mackenzie, L. and Capra, M. (2006). Health Professionals' Attitudes and Practices in relation to Functional Capacity. 4th Congress of the World Federation of Occupational Therapists. Sydney, NSW, Australia.
9. James, C. and Mackenzie, L. (2008). The Clinical Utility of Functional Capacity Evaluations: the opinion of Health Professionals working within Occupational Rehabilitation. OT Australia 23rd National Conference. Melbourne, Australia.
10. James, C., Mackenzie, L. and Capra, M. (2010). Evidence for Functional Capacity Evaluations – the test-retest reliability of the WorkHab FCE. 15th Congress of the World Federation of Occupational Therapists. Santiago, Chile.

Table of Contents

Statement of Originality ii
Thesis by publication Acknowledgement of Authorship iii
Acknowledgements iv
Publications arising from work presented in this thesis v
Refereed Journal publications: v
Conference paper presentations: vi
Table of Contents vii
Table of Figures x
Table of Tables xi
Abbreviations xii
Thesis Abstract 1
Chapter 1 Introduction 3
1.1 The Problem 6
1.2 Aims of the research 7
Usage: 7
Reliability of the WorkHab FCE: 8
Validity of the WorkHab FCE: 8
1.3 Significance of the thesis 10
1.4 Thesis overview 11
Chapter two 11
Chapter three 11
Chapter four 12
Chapter five 13
Chapter six 13
Chapter 2 Literature Review 14
2.1 What is function? 14
2.2 What are Functional Capacity Evaluations? 16
2.3 Models and theoretical frameworks applied to FCEs 19
2.4 Types of Functional Capacity Evaluations 22
2.5 Best practice and Functional Capacity Evaluation 24
2.6 The use of Functional Capacity Evaluations in practice – application to the work environment 28
2.7 The WorkHab Functional Capacity Evaluation 33
2.8 Evidence of psychometric properties of FCEs 36
2.9 A review of published studies related to the reliability and validity of FCEs 41
2.10 Chapter Summary 56
Chapter 3 Current Practice with Functional Capacity Evaluations 58
3.1 Manuscript: Health Professionals' Attitudes and Practices in relation to Functional Capacity Evaluations 62
Abstract 62
3.1.1 Introduction and Literature review 63
3.1.2 Methodology 64
3.1.3 Results and Discussion 66
3.1.4 Conclusion 71
Acknowledgements 72
References 72
Appendix A 75
3.2 Manuscript: Health Professionals' Perceptions and Practices in relation to Functional Capacity Evaluations – Results of a Quantitative Survey 76
Abstract 76
3.2.1 Introduction 78
3.2.2 Method 80
3.2.3 Results 81
3.2.4 Discussion 87
3.2.5 Conclusions 91
Acknowledgements 91
References 91
3.3 Manuscript: The Clinical Utility of Functional Capacity Evaluations: the Opinion of Health Professionals working within Occupational Rehabilitation 95

Abstract: 95
3.3.1 Introduction and Literature review 96
3.3.2 Method 98
3.3.3 Results 99
3.3.4 Discussion 103
3.3.5 Conclusion 107
Acknowledgements 108
References 108
3.4 Chapter Summary 111
Chapter 4 The WorkHab Functional Capacity Evaluation: Reliability 114
4.1 Manuscript: Test - Retest reliability of the Manual Handling component of the WorkHab Functional Capacity Evaluation in healthy adults 119
Abstract 119
4.1.1 Introduction 120
4.1.2 Method 121
4.1.3 Results 123
4.1.4 Discussion 126
4.1.5 Conclusion 129
Acknowledgements 129
References 129
4.2 Manuscript: Inter and Intra-rater reliability of the Manual Handling component of the WorkHab Functional Capacity Evaluation 133
Abstract 133
4.2.1 Introduction 135
4.2.2 Method 136
4.2.3 Results 138
4.2.4 Discussion 145
4.2.5 Conclusion 148
Acknowledgements 149
References 149
4.3 Chapter Summary 152
Chapter 5 The WorkHab Functional Capacity Evaluation: Validity 155
5.1 Manuscript: Content Validity of the WorkHab Functional Capacity Evaluation 163
Abstract 163
5.1.1 Introduction 165
5.1.2 Method 167
5.1.3 Results 169
5.1.4 Discussion 177
5.1.5 Conclusion 183
Acknowledgements 183
References 183
5.2 Manuscript: Physiological correlates of functional capacity evaluations: finding the safe maximal lift 188
Abstract 188
5.2.1 Introduction 189
5.2.2 Methods 190
5.2.3 Results 192
5.2.4 Discussion 195
5.2.5 Conclusions 198
Acknowledgements 198
References 198
5.5 Chapter Summary 202
Chapter 6 Discussion 205
6.1 Key research findings within the thesis 205
Current usage of FCEs 205
Clinical utility 207
Reliability 208
Validity 209
6.2 Limitations within the program of research for the thesis 212
6.3 Implications of thesis findings for clinical practice 214

6.4 Future directions for research 218
6.5 Conclusions 219
References 222
Appendices 238
Appendix 1 Co-author statements 239
Manuscript 1: (3.1) Health Professionals' Attitudes and Practices in relation to Functional Capacity Evaluations 240
Manuscript 2: (3.2) Health Professionals' Perceptions and Practices in relation to Functional Capacity Evaluations – Results of a Quantitative Survey 243
Manuscript 3: (3.3) The Clinical Utility of Functional Capacity Evaluations: the Opinion of Health Professionals working within Occupational Rehabilitation 245
Manuscript 4: (4.1) Test - Retest Reliability of the Manual Handling Component of the WorkHab Functional Capacity Evaluation in Healthy Adults 247
Manuscript 5: (4.2) Inter and Intra-rater Reliability of the Manual Handling Component of the WorkHab Functional Capacity Evaluation 250
Manuscript 6: (5.1) Content Validity of the WorkHab Functional Capacity Evaluation 253
Manuscript 7: (5.2) Physiological correlates of Functional Capacity Evaluations: Finding the Safe Maximal Lift. 256
Appendix 2: Poster presentations 259
Appendix 3: Questionnaire - Health Professionals' Attitudes and Practices in relation to Functional Capacity Evaluations 266
Appendix 4: Content validity online survey instrument - Application of an FCE. 277

Table of Figures

Figure 1: Diagram of project plan and studies completed 9
Manuscript 3.2.
Figure 1: Utilisation of All Components of Standardised FCE (n=71*) 84
Figure 2: Utilisation of PARTS of FCE (n=60) 84
Figure 3: Frequency of use of most popular Standardised FCEs..... 85
Figure 4: Influences for choices of FCE..... 86
Manuscript 3.3
Figure 1: Usefulness and Relevance of FCEs (n=77) 100
Figure 2: Adaptability and Flexibility of FCEs (n=77)..... 101
Figure 3: Therapist perceived requirements to administer FCEs (n=77)..... 102
Figure 4: Issues in practice administering FCEs (n=77)..... 103
Manuscript 4.1
Figure 1: Limits of Agreement: Bench to bench, Floor to bench and Bench to shoulder lifts... 125
Manuscript 5.1
Figure 1 – Relationship of Lifting FCE components to DOT physical demands..... 177
Manuscript 5.2
Figure 1: Lift time, muscle involvement and weight progression (N=20)..... 193
Figure 2: RMS as % of MVC for each muscle for each weight. (N=20)..... 193

Table of Tables

Table 1: Summary of findings of functional capacity evaluation psychometric property literature 44

Manuscript 3.1

Table 1: Details of Participants 65

Manuscript 3.2

Table I: Descriptive information on sample (n=77) 82

Table II: Usage by FCE type and Profession (n=77). 83

Table III: Mean score of importance of different qualities/ criteria in choice of FCE (n=77) 86

Manuscript 3.3

Table 1: Features related to the clinical utility of a tool. 97

Manuscript 4.1

Table 1: Results of means, differences, standard deviations, 95% confidence intervals and ICC's for lifts. 124

Table 2: Percentage agreement and Cronbach's alpha for the manual handling score components. 126

Manuscript 4.2

Table 1: Intra-rater agreement: ICC for Manual Handling components (N=14). 139

Table 2: Intra-rater agreement: Manual Handling Score Results: means, difference, standard deviations 95% confidence intervals and ICC for each injured worker (subject). (N=14). 140

Table 3: Intra-rater agreement-Manual Handling Component Results: mean difference (time1: time 2), standard deviation, 95% confidence intervals and ICC for each injured worker (subjects) (N=14). 142

Table 4: Inter-rater agreement: ICC for Manual Handling components (N=17). 143

Table 5. Inter-rater agreement for manual handling components for each injured worker (N=17). 144

Manuscript 5.1

Table 1: Relevance of WorkHab FCE components 170

Table 2: Difficulty of WorkHab FCE components 173

Table 3 Percentage agreement between WorkHab FCE components and the DOT physical demands (n=56). 176

Manuscript 5.2

Table 1: Differences between lifting load up and down, before and after SML and with increasing weight (N=20) 194

Abbreviations

AWP	Assessment of Work Performance
BOH	Bench to Overhead
BS	Bench to Shoulder
BTE	Baltimore Therapeutic Evaluation
CLBP	Chronic Low Back Pain
DOT	Dictionary of Occupational Titles
DVD	Digital video disc
EMG	Electromyography
FAST	Functional Assessment Screening Test
FB	Floor to Bench
FCE	Functional Capacity Evaluation
GAPP-FCE	Gibson Approach to Functional Capacity Evaluation
Hz	Hertz
ICC	Intra-class Correlation
ICF	International Classification of Functioning, Disability and Health
IDR	Instrument for Disability Risk
IWS	Isernhagen Work Systems
LBP	Low Back Pain
mV	Millivolts
MVC	Maximum Voluntary Contraction
NIOSH	National Institute for Occupational Health and Safety
NSW	New South Wales (Australia)
OT	Occupational Therapist
PDI	Pain Disability Index
PILE	Progressive Isoinertial Lifting Evaluation
PWPE	Physical Work Performance Evaluation
RMS	Root Mean Square
RTW	Return to Work
RWL	Recommended Weight Limit
SD	Standard Deviation
SEMG	Surface Electromyography
SML	Safe Maximal Lift
UK	United Kingdom
WRULD	Work-related Upper Limb Disorder
WWS	Work Well Systems Functional Capacity Evaluation
VAS	Visual Analogue Scale

Thesis Abstract

Functional Capacity Evaluations (FCEs) are part of practice in occupational rehabilitation, and are designed to define an individual's functional abilities in the context of safe, productive work tasks. The WorkHab Functional Capacity Evaluation is one of many currently available FCEs. It is commonly used in Australian occupational rehabilitation: however, there is a lack of evidence of its psychometric properties. This thesis reports on research that investigated reliability and aspects of validity of the WorkHab FCE.

The current practice of FCE use in the Australian occupational rehabilitation context was investigated. Qualitative and quantitative methodology were used to study the perceptions and practices of health professionals about the use and clinical utility of FCE's. Results found health professionals use more than one FCE, with the WorkHab FCE the second most commonly used in NSW Australia. There was consistency and similarities in FCE use in practice, with participants adapting FCEs to suit the situation and completing parts rather than the whole of a FCE.

Four studies subsequently investigated the measurement properties of the WorkHab FCE. The manual handling components were evaluated, including test-retest reliability in healthy adults, and intra-rater and inter-rater reliability using DVD footage of injured workers FCEs. Content validity was evaluated using a cross sectional survey of health professionals who use FCEs in practice. Construct validity of the bench to shoulder lift was explored using Electromyography (EMG) to study muscle activity in the upper body.

Results found substantial levels of test-retest reliability and intra-rater and inter-rater reliability for the lifting components of the WorkHab FCE. The findings support content validity for the WorkHab FCE specifically in relation to manual work and vocational retraining; however, construct (convergent) validity of the safe maximal lift of the bench to shoulder lift of the WorkHab FCE was unable to be established using EMG physiological parameters.

Future directions for research of the WorkHab FCE and implications for clinical practice are discussed.

Chapter 1 Introduction

Overview: This chapter introduces the topic of Functional Capacity Evaluations and the WorkHab FCE. It gives the rationale and aims of the project, the project plan and an overview of the thesis.

Functional capacity evaluation (FCE) is a performance measure that is used in occupational and vocational rehabilitation to make decisions about a worker's capacity in relation to work abilities. FCEs are used to screen potential employees as pre-employment assessments, to assess physical rehabilitation needs, to determine work readiness and job placement following injury, to facilitate return to work and to determine a person's functional capacity for compensation or litigation reasons (King et al. 1998; Lee et al. 2001b; Schonstein and Kenny 2001; Strong et al. 2004a; Gouttebarge et al. 2006; Innes 2006; Gross et al. 2007).

FCEs were developed as a compilation or battery of tests predominantly used by occupational and physical therapists to measure the ability of a person to work. A FCE can include a medical history, physical examination and a variety of work-related performance tests; however, there are general inconsistencies in the terminology used to describe a FCE. Terms used include: physical work assessment, physical tolerance assessment, functional assessment, and functional capacity evaluation. There are also various procedures and report formats used across these FCEs (Reneman and Dijkstra 2003). Related to this are differences in the definitions of various aspects of FCEs. In an international Delphi survey, 63% of experts in occupational rehabilitation agreed on the definition of FCE as: "A FCE is an evaluation of capacity of activities that is

used to make recommendations for participation in work while considering the person's body functions and structures, environmental factors, personal factors and health status" (Soer et al. 2008, pg 394). Throughout this thesis, the above definition of functional capacity evaluation is used.

FCEs are commonly used with individuals who have suffered work-related injuries and can be used in the rehabilitation process for a range of injury and disease types. Research has focussed on the use of FCEs with individuals with low back pain (Reneman et al. 2002c; Gibson et al. 2005; Gouttebarge et al. 2006; Reneman et al. 2006a; Reneman et al. 2006b), as well as a range of work-related musculo-skeletal injuries (Deen et al. 2002; James and Mackenzie 2009b).

Health professionals who work in occupational rehabilitation use a variety of FCEs, which are broadly categorised into three groups according to the purpose of the assessment. These purposes include: i) to assess people with a job, ii) to assess people without a specific job; and iii) as a job or work capacity evaluation (Innes and Straker 2003a; Jones and Kumar 2003). Some standardised assessments are commercially available, whereas others were developed by specific clinics or therapists. Non-standardised assessments are often related to local needs and are often specific to certain work environments or situations (James and Mackenzie 2009b).

A standardised FCE can be defined as one that is commercially available, has acceptable measurement/psychometric properties and is conducted using standardised procedures and protocols (Innes and Straker 2003b; Reneman et al. 2006a). A non-standardised FCE can be defined as a self- designed,

internally adaptable assessment tool that may be specifically designed for a particular industry, and may include components of commercially available tools (Chappell et al. 2006).

Standardised functional capacity evaluations and work-related assessments require evidence of reliability and validity, both of which are seen as important pre-requisites for accurate and meaningful measurement (Portney and Watkins 2009). Ideally, these properties should be established in the development stage of an assessment tool. However, for many FCEs there is a shortage of research to confirm the reliability and validity of assessments used to evaluate the rehabilitation needs of workers and the assessment of their work capacity (King et al. 1998; Innes and Straker 1999a; Innes and Straker 1999b; Ekbladh et al. 2004; Gouttebarge et al. 2004).

Reliability is defined as the extent to which a measure is consistent, free from error, and demonstrates the reproducibility or dependability of the assessment over time (Portney and Watkins 2009). Validity is defined as the extent to which an instrument measures what it is intended to measure (Portney and Watkins 2009). Reviews have highlighted the limited evidence for reliability and validity of FCEs (King et al. 1998; Innes and Straker 1999a; Innes and Straker 1999b; Gouttebarge et al. 2004; Innes 2006). However, more recently, applicable research for some commercially available assessment tools has begun to appear in the literature (Ting et al. 2001; Matheson et al. 2002b; Durand et al. 2004; Pransky and Dempsey 2004; Reneman et al. 2004).

The WorkHab Functional Capacity Evaluation is an Australian standardised assessment used in occupational rehabilitation and work environment settings

to determine functional capacity. The WorkHab FCE is based on objective physiological measures, including heart rate, observations of biomechanics and reported pain and perceived exertion (effort) ratings (Bradbury and Roberts 1998). This FCE was identified as the most widely used commercially available assessment tool in the Australian occupational rehabilitation environment. Thirty-six percent of Occupational Therapists in an Australian study identified the WorkHab FCE as the FCE most in use (Deen et al. 2002). Furthermore, in studies of rehabilitation providers in New South Wales (Australia), the WorkHab FCE was identified by 17% and 52% of respondents respectively as the most frequently used commercially available FCE (Cotton et al. 2006; James and Mackenzie 2009b). Despite this, an investigation into the reliability and validity of several FCEs by Innes and Straker (1999a and 1999b) found there is limited evidence on the psychometric properties of this particular assessment tool related to its use in the work setting. Apart from the studies presented in this thesis, no further publications that relate to the reliability and validity of the WorkHab FCE were located.

1.1 The Problem

There is a lack of psychometric properties of specific FCE tools, despite the need for health professionals to use evidence to support interventions. Studies have been conducted on the current practice and attributes of work assessments in general (Innes and Straker 2002b; Innes and Straker 2003b). However, investigation has not been documented specifically into FCE use, including the perceptions and experiences and the clinical utility of FCEs by health professionals in the Australian environment. The WorkHab FCE is a commonly used FCE in the Australian occupational rehabilitation arena;

however, there is a lack of published literature about its psychometric properties. The need for evidence about the psychometric properties of assessment tools used in occupational rehabilitation led to the development of this project to investigate the usage of FCEs in the Australian occupational rehabilitation arena, and the reliability and validity of the WorkHab FCE. The specific aims of the research are discussed below.

1.2 Aims of the research

The aims of the research described in this thesis were: i) to evaluate the level of usage of the WorkHab FCE in occupational rehabilitation practice in Australia; and ii) to determine the reliability and validity of the WorkHab FCE.

Specific research questions to address these aims were:

Usage:

What are the attitudes, perceptions and behaviours of Australian health professionals in relation to Functional Capacity Evaluation use?

What are the factors that influence the selection of any FCE tool by health professionals?

How does the usage of the WorkHab FCE compare with other FCEs in Australia?

What factors influence health professionals' clinical judgements and professional reasoning when providing results and recommendations for the individual being assessed?

Reliability of the WorkHab FCE:

What is the inter-rater reliability of health professionals conducting the WorkHab FCE?

What is the intra-rater reliability of health professionals conducting the WorkHab FCE?

What is the test re-test reliability of health professionals conducting the WorkHab FCE?

Validity of the WorkHab FCE:

What is the level of content/face validity demonstrated by the WorkHab FCE?

What is the level of construct validity demonstrated by the WorkHab FCE?

Figure 1 provides an overview of the project plan and the studies completed as part of this thesis.

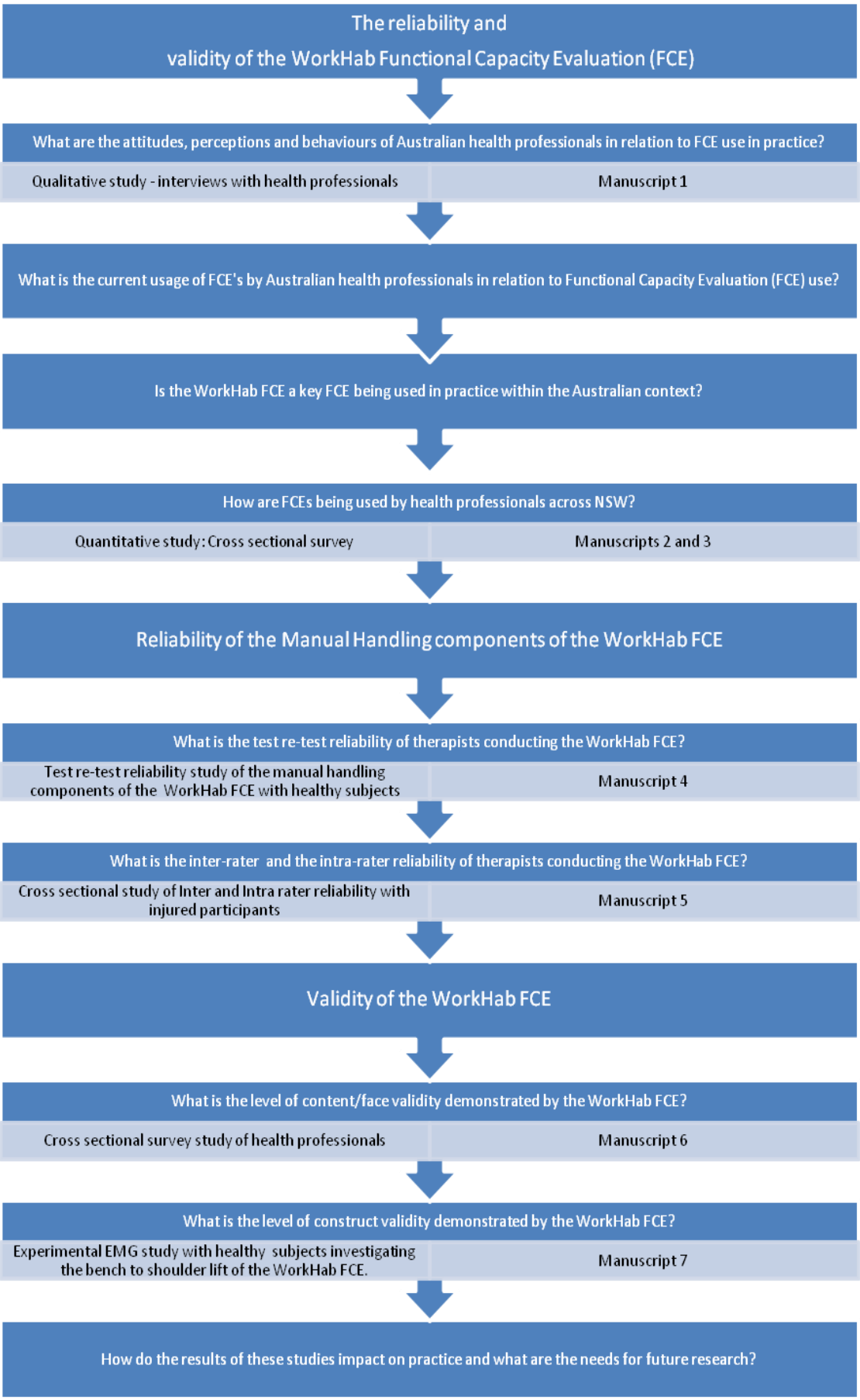


Figure 1: Diagram of project plan and studies completed

1.3 Significance of the thesis

The investigation into the current practice of FCE use in the Australian occupational rehabilitation environment attempts to identify what specific FCE tools are used and why health professionals chose particular tools in practice. Analysis of the perceptions of health professionals provides a greater understanding of practice and the clinical utility of these tools. This helps to inform future practice in this area. Investigation of the psychometric properties of the WorkHab FCE produces evidence to support it's use in practice and to inform health professionals with regards to the efficacy of the tool. The test-retest, inter-rater and intra-rater reliability determines evidence for the reliability of this tool in the studied populations. The content validity study attempts to determine if the content of the WorkHab FCE is appropriate for use in the occupational rehabilitation setting and the construct validity tries to provide quantitative evidence for determination of a safe maximal lift in this tool.

In the current occupational rehabilitation climate, evidence for the reliability and validity of specific assessment tools is essential when assessment results are relied upon in litigation processes. Furthermore, this evidence also underpins the use of a FCE tool as an outcome measure to justify rehabilitation interventions and to provide a measure of financial accountability for those who pay for FCE services. Appropriate clinical and professional reasoning decisions are dependent on accurate findings from a FCE and are therefore essential for successful occupational rehabilitation programs. The consumers of occupational rehabilitation programs also need confidence that a FCE is reliable, so that results will be consistent regardless of different evaluators in different parts of the country (Gardener and McKenna 1999).

1.4 Thesis overview

This thesis is presented in publication style. Each manuscript was written in the conventional publication style for the journal to which it was submitted. However, in this thesis a word document of each manuscript is presented and a consistent referencing style, (the generic Author - Date style), is used throughout.

Chapter two

This chapter consists of an overview and discussion of the general literature related to FCEs. Further relevant literature is also presented in subsequent chapters and within each individual manuscript presented in the thesis.

Chapter three

Information on the current practice of FCE use in the Australian occupational rehabilitation context is presented in this chapter. The chapter presents two studies that were completed and published.

The first study was a small qualitative study that used a phenomenological approach to explore why health professionals chose a particular FCE, and to identify what factors influenced the decisions made by health professionals when providing results and recommendations for the individual being assessed. (See **Manuscript 1 (3.1)**: "*Health professionals' attitudes and practices in relation to functional capacity evaluations."* published in 2007 in Work: A Journal of Prevention and Rehabilitation, 29(2), 81-88).

The second study was informed by the qualitative study and further investigated the perceptions and practices of Australian health professionals about the use of FCEs and views about the clinical utility of FCEs in general. This study used

a cross- sectional study design to survey health professionals who conducted FCEs. (See **Manuscript 2 (3.2):** *"Health professionals' perceptions and practices in relation to functional capacity evaluations – results of a quantitative survey",* published in 2009, in the Journal of Occupational Rehabilitation, 19(2) 203-211 and **Manuscript 3 (3.3):** *"The clinical utility of functional capacity evaluations: the opinion of health professionals working within occupational rehabilitation"* published in 2009, in Work: A Journal of Prevention and Rehabilitation, 33(3) 231-239).

Chapter four

The reliability of the WorkHab FCE is discussed in this chapter, specifically in relation to the manual handling component, and presents manuscripts of two further published studies.

The first study investigated the test re-test reliability of the WorkHab FCE with healthy adults. This study focussed on the test results of the manual handling component of the WorkHab FCE, specifically the floor to bench lift, the bench to bench lift and the bench to shoulder lift. (See **Manuscript 4 (4.1)**:*"Test-retest reliability of the manual handling component of the WorkHab functional capacity evaluation in healthy adults,"* published in 2010, in Disability and Rehabilitation, 32(22): 1893-9).

The second study evaluated the intra-rater and inter-rater reliability of the manual handling component of the WorkHab FCE using a cross sectional study design. Video footage of the manual handling component of four injured workers' FCEs was divided into lifting segments, which were randomly assigned on a DVD. Health professionals who were trained and accredited in the use of

the WorkHab FCE then rated these lifting segments by means of the WorkHab manual handling scoring system and identified if the lift was a safe maximal lift. (See **Manuscript 5 (4.2):** *"Inter and intra-rater reliability of the manual handling component of the WorkHab functional capacity evaluation"* published early online (January 19, 2011) in Disability and Rehabilitation DOI:10.3109/09638288.2010.548896).

Chapter five

This chapter presents the results of studies to evaluate the content and construct validity of the WorkHab FCE.

The first study investigated content validity of the WorkHab and used a cross sectional design to survey health professionals who use FCEs in practice. (See **Manuscript 6 (5.1):** ***"****Content validity of the WorkHab functional capacity evaluation"* submitted for publication in Disability and Rehabilitation).

The second study investigated the construct validity of the manual handling component of the WorkHab FCE. This study used Electromyography (EMG) to study muscle activity in the upper body during a bench to shoulder lift in healthy subjects. (See **Manuscript 7 (5.2):** *"Physiological correlates of functional capacity evaluations: finding the safe maximal lift"* submitted for publication in Archives of Physical Medicine and Rehabilitation).

Chapter six

This chapter provides a general discussion of the psychometric properties of the WorkHab FCE and the implications of the results of this research program to health professionals working in occupational rehabilitation settings. Limitations and future directions for research in this area are also presented.

Chapter 2 Literature Review

Overview: This chapter discusses relevant literature related to functional capacity evaluations, the models and theoretical frameworks informing FCE practice, types of FCEs, best practice and FCE use in clinical practice in the occupational arena. It also discusses evidence of the psychometric properties, including aspects of the reliability and validity of FCE tools with more specific literature discussed in subsequent chapters.

2.1 What is function?

Function is defined in the 'Oxford Dictionary' as 'an activity that is natural to or the purpose of a person' (Soanes and Stevenson 2005). Function is also defined as 'action for which a person is fit' (Hussey et al. 2007 pg 288) and the International Classification of Functioning, Disability and Health (ICF) defines functioning as an umbrella term that encompasses all body functions and structure, activities and participation (World Health Organisation 2001). How function is evaluated is dependent upon the definition of function and each discipline views issues of human function and functional limitation from their own theoretical perspective. Therefore differences exist between disciplines due to their philosophies and traditions. This influences the approach used by disciplines in assessment of function and subsequently the way that an evaluation is completed (Matheson et al. 2001).

The type of theoretical health model used in practice will also influence how function is defined and therefore how function is evaluated. The traditional 'medical model', in simplistic terms, aims to 'fix' impairment from a mechanistic and reductionistic view of human functioning. Joint range of motion or muscle

strength can be classed as function in traditional medical approaches, such as within orthopaedic medical practice or physiotherapy practice. Whereas within a more holistic interpretation of functional testing, the 'whole' person is considered rather than a part in isolation, with the purpose to define work abilities and limitations (Isernhagen 2009).

In occupational therapy, the term function is often linked to the participation of a person in activities of daily living, leisure and productivity and is a term often used synonymously with occupational performance. Occupational performance is defined as the doing of occupation in order to satisfy life needs (Law et al. 2001). According to Law (2001), functional evaluation will consider the whole performance output rather than evaluation of performance components (or components of function) such as strength and range of movement (Law et al. 2001). This concurs with Mathiowetz (1993) who proposed a hierarchy of functional assessment, with assessment of occupational performance (activities of daily living, work, play, leisure) and the performance of actual roles (home-maker, worker, student etc), the observable and productive outcomes, positioned higher up and more important in the functional hierarchy. The actual isolated physical performance components such as strength and range of movement he positioned at the bottom of the hierarchy (Mathiowetz 1993).

Evaluation of the role and occupational performance of an individual rather than evaluation of performance components (such as strength, fine motor skills and problems solving) is endorsed by the American Occupational Therapy Association (Moyers 1999; American Journal of Occupational Therapy 2000). Using this approach difficulties with function are discussed in terms of functional impairment and/or functional limitations. Functional impairment is defined as the

restriction of ability to perform simple observable behaviours that share a common purpose. Others define functional limitation as the restriction or lack of ability in physical or mental functions that hinder an individual's ability to perform tasks or activities that are personal, social or vocational in nature (Matheson et al. 2001; Law 2002).

The ICF provides another perspective on function (World Health Organisation 2001). The ICF aims to provide a unified and standard language to describe health and health related conditions that is common across disciplines (Stamm et al. 2005). Within this classification, functioning is described as dependent upon six inter-related components: disease and disorder; functions and structures; activities or limitations of activities to perform a task; participation or limitations in the involvement of a life situation; environmental factors and personal factors. The ICF describes the situation of each person and different health domains within the context of environmental and personal factors. It considers functioning and disability along with contextual factors and disability (or the impaired interaction between the individual and the environment) and is said to exist when dysfunction occurs at one or more of the following: 1) body or body part; 2) whole person; and 3) whole person in social context (Davis and Madden 2006). However, there remain differences and, at times, confusion of terminology with respect to function, and with respect to functional capacity evaluations (Innes and Straker 1998b; Soer et al. 2008).

2.2 What are Functional Capacity Evaluations?

A functional capacity evaluation is an objective measure of the level of function or ability to perform work tasks and skills of an individual (King et al. 1998;

Chappell et al. 2006). FCEs are batteries of tests that practitioners use to form an evaluation of a person's ability to perform work-related activities. As noted in chapter one, there was 63% consensus from an expert panel on the following definition of a FCE: "A FCE is an evaluation of capacity of activities that is used to make recommendations for participation in work while considering the person's body functions and structures, environmental factors, personal factors and health status" (Soer et al. 2008, pg 394). This definition has similarities to the ICF definition of functioning, with the inclusion of various components related to the individual and the environment; however, there was not absolute consensus on this definition of a FCE.

Most studies demonstrate that practitioners consider a FCE as a valuable tool in work-injury management because it provides information about the occupational performance and impairment of an injured individual. FCEs are designed to provide objective information related to readiness to work following injury or illness (Toeppen-Sprigg 2000; Maher 2006; Mitchell 2008). FCEs are designed to respond to the needs of the worker, health care providers, employers, insurers, social security administrators and the legal community. They aim to answer the question of how injury affects a person's ability to work by providing information related to a person's work capacity, employability and work readiness, return to work abilities, rehabilitation needs, and litigation and compensation issues (King et al. 1998; Schonstein and Kenny 2001; Strong et al. 2004a; Mitchell 2008). To provide this information, FCEs look at the work performance of the individual, by specifically investigating the performance of the physical components of work, and relate this to general or specific job duties and tasks (Innes and Straker 2003a).

FCEs are administered by health professionals and in Australia these are often conducted by occupational therapists and physiotherapists. The FCE may consist of a medical history, physical examination and a battery of work-related performance tests that focus on the physical capacity for performing these in the workplace. The findings from these tests are then interpreted in relation to job requirements and for the purpose of preparing a report (Tramposh 1992; King et al. 1998; Chappell et al. 2006). The specific physical tasks assessed by FCEs in general include: the body postures of sitting, standing, walking; body dexterity including bending, stooping, kneeling, crawling, crouching; limb coordination including climbing, handling, reaching; and strength, which includes lifting, carrying, pushing and pulling, all are assessed using a range of equipment and weights, in a standard and structured approach. The physical tasks of many FCEs are based on the physical domains outlined in the US Department of Labor Dictionary of Occupational Titles (DOT) (U.S.Department of Labor 1965; King et al. 1998). The DOT provides information on the physical demands and work characteristics for many jobs (job factors). There are 20 job factors that express the requirements of the job and the capacities a worker must have to meet or exceed those demands.

Specific FCEs differ because some assessment tools also include job simulation assessment and some include behavioural components such as assessment of effort, symptom exaggeration and fear avoidance. Other FCEs vary in their assessment of safety factors and in the determination of end points for assessment components. Some also include assessment of psychological issues as part of the FCE to determine safe lifting or physical demand

capabilities (Velozo 1993; Gibson and Strong 1996; Innes and Straker 1998b; Tuckwell et al. 2002; Gross 2004).

2.3 Models and theoretical frameworks applied to FCEs

The main theoretical approaches that inform the use of FCEs by practitioners are the biomechanical, physiological, metabolic, psychophysical and kinesiophysical models (Abdel-Moty et al. 1993; Gibson and Strong 1997). The biomechanical model focuses on the relationship of musculoskeletal and neuromuscular systems to function. This focuses on the performance components of a person's range of movement, strength and endurance that underpins their ability to perform work within safe musculoskeletal or neuromuscular limits (McMillan 2006). The physiological model relates to the contribution of the cardiovascular, pulmonary and metabolic systems to work functions. The metabolic approach is similar to the physiological approach and is based on the quantification of physiological measures such as oxygen uptake and physiological stress. The psychophysical model relates to the influence of the cognitive/perceptual systems on functioning. Using the kinesiophysical approach, the evaluator focuses on the person's physiological responses and adaptations to workload. In addition, the evaluator observes physiological responses and thus determines safe maximal capacity. Using the kinesiophysical approach, it is the evaluator who controls the test by monitoring for physiological maximal capacity.

When an evaluator uses the psychophysical approach, control is given to the participant so it is the participant who decides when to terminate the test, and thereby determines maximum function (Abdel-Moty et al. 1993; Gibson and

Strong 1997; Mitchell 2008). An Australian study in 1997 found that the psychophysical approach was cited as the primary model used in FCE practice (Gibson and Strong 1997). However, issues with injured participants determining their own safe limits have been identified, and the kinesiophysical model has been suggested as a model of choice because this utilises observation of movement patterns (biomechanics) and physiological performance to evaluate safe maximum function (Isernhagen 1992). Many FCEs use a combination of models, which facilitates health professionals to gather a range of data, and combined with their clinical reasoning skills, this enables a comprehensive analysis for adequate and appropriate recommendations of functioning.

FCEs tend to lack an identified theoretical framework upon which they are based, especially in relation to occupational therapy practice. Many assessments used by occupational therapists were developed by other disciplines such as physiotherapy, psychology, and medicine. This is despite occupational therapists historically being the primary discipline engaging in functional capacity assessment, and function being a core component of assessment and treatment more generally within occupational therapy practice. As a result, there is some confusion and differences in the terminology used around functional capacity evaluation, partly due to the diversity of disciplinary approaches to 'function' as discussed earlier.

From an occupational therapy perspective, some work-related assessment or measurement tools, such as the Worker Role Interview, have been based on occupational therapy models such as the Model of Human Occupation (Fisher 1999). Other occupational therapy researchers have suggested the use of an

occupational role or occupational performance frame of reference to develop assessment tools, with physical performance components being one aspect to focus upon (Mathiowetz 1993; Law et al. 2001; Gibson and Strong 2003). More recently, occupational therapy authors have suggested the use of the International Classification of Functioning (World Health Organisation 2001) because of its common interdisciplinary language to describe health, health related outcomes and health determinants (Gibson and Strong 2003; Imms 2006; Soer et al. 2008).

The benefit of using the ICF is that it provides a model that considers multiple factors that influence functioning in respect to disease and disorder; functions and structures; activities or activity limitations participation or limitations in the involvement of a life situation; environmental and personal factors (World Health Organisation 2001). FCEs can be considered tools that evaluate participation and functioning or disability in conjunction with the context of work. Using the ICF framework, disability (or impaired interaction) between the individual and the work environment exists when dysfunction occurs at one or more of the three areas identified earlier. In relation to FCEs, assessment will be conducted at the body or body part level where aspects such as posture, movements and strength will be evaluated; at the whole body level where a worker's capacity to complete a range of activities associated with a job are considered; and at the level of the whole person in a social context where a worker's role in the organisation is assessed (Davis and Madden 2006; Imms 2006). Despite agreement on using the ICF as a conceptual framework, and general consensus on the operational definitions in FCEs, which provides for improved communication, a study completed by Soer et al (2008) did not identify better

interpretation of data and patient outcome, or consensus of definitions across all aspects of FCEs.

2.4 Types of Functional Capacity Evaluations

Several authors have attempted to classify FCEs. In their study of current practice of clinicians in relation to work-related assessments, Innes and Straker (2002) identified three types of assessment: workplace assessments and two types of functional capacity evaluations – those for no specific job and those for a specific job. They found that when FCEs were performed on participants with no specific job, the assessments were described as broad, consistent, controlled, formal, generalisable, generic, quantified, reproducible, rigid in structure and standardised, safe and removed from the work environment. FCEs (for a specific job) were described as more conservative, focussed, specific and less standardised (Innes and Straker 2002b).

In 2003, Matheson identified five different types of FCE with each defined by its purpose and capturing a more diverse range of situations. These are listed below:

- Functional goal setting: to determine a person's ability to perform a key task;
- Disability rating: to measure the loss of ability in key functional areas of work to determine an estimate of disability;
- Job Matching: to match the worker's ability to the essential functions of the job;
- Occupation Matching: to match a person's functional capacity to the demands of an occupational group (more comprehensive and demanding

than job matching because all components of an occupation are considered); and

- Work Capacity Evaluation: matching a person's functional capacity to all occupations (no occupational target requires this to be much broader) (Matheson 2003).

There are similarities between the Work Capacity Evaluation category proposed by Matheson (2003) and the FCE (no job) identified by Innes and Straker (2002) in their classification. Both are suggested as broader, more generic types of assessments that are used to determine a person's general ability to work. While the primary focus of many FCEs is on the physical capacities and body functions of an individual, it is recognised that psychosocial, behavioural and environmental factors also impact upon a participant's work ability and it is suggested that these should also be considered when making clinical judgements (Velozo 1993; Gibson et al. 2005; McFadden et al. 2010).

There are a variety of commercially available standardised FCEs and many clinics have developed their own non-standardised and work specific FCEs. A standardised FCE is one that is commercially available, has acceptable measurement/psychometric properties and is conducted using standardised procedures and protocols (Innes and Straker 2003b; Reneman et al. 2006a). Such FCEs are often a packaged collection of tests, equipment and protocols with a standard report writing format, and developers may offer training in how to administer, analyse results and report the findings (Chappell et al. 2006). Conversely, these authors define a non-standardised FCE as a self-designed,

internally adaptable assessment tool that may be specifically designed for a particular industry and may include components of commercially available tools.

In addition, health professionals often make modifications to the FCE battery of tests based on their personal and clinical experience, which represents a departure from any standardised procedures demanded by the tool to maintain its measurement quality. Therefore, such modifications will impact upon the reliability and validity of the tool (James and Mackenzie 2009b; James and Mackenzie 2009a; McFadden et al.2010).

2.5 Best practice and Functional Capacity Evaluation

Chappell et al (2006) identified six key factors for FCE best practice. These are summarised below:

1. The evaluation process should reflect evidence based practice, where the clinician integrates knowledge from research findings, expert consensus, clinical expertise and experiences, and the values, needs and experience of the participant to improve practice;
2. The clinician must be competent to perform the FCE. This involves having appropriate knowledge, training and expertise to conduct the FCE. Specifically knowledge of disease and injury, normal movement patterns, ergonomics and rehabilitation sciences and the implementation of the specific FCE tool are required;
3. There should be clearly identified objectives and referral questions associated with the FCE so that the clinician has direction about the purpose of the FCE. Purposes include using the assessment to determine a

participant's physical capacity for a specific job or more generally for overall capacity to work, or to determine further rehabilitation needs;

4. The participant should be medically stable. The clinician must be aware of the medical status of the participant to ensure that the condition is stable and that the participant does not have any additional conditions that may compromise their safety during the FCE. The participant must have reached a stage in their recovery that is congruent with the goal of the assessment;
5. The evaluator should adhere to the 'gold standard' when selecting a test and should adhere to the core attributes and procedures that make up the 'gold standard'. In the absence of a 'gold standard' the attributes to be considered include: reliability, validity, safety, utility and practicality; and
6. The appropriate test procedures must be followed – to ensure there is no compromise of safety, reliability or validity of the FCE or interpretation of the results (Chappell et al. 2006).

For a FCE to be considered an acceptable measure it needs to have reliability (consistency of measurement) and validity (measurement of what is intended) (King et al. 1998; Mitchell 2008; Isernhagen 2009). Standardisation (a clear set of procedures for administering and scoring) and objectivity (a degree of reliability void of assessor bias and performed within operational definitions) of the assessment tool have also been suggested as necessary for consideration of a well designed tool (King et al. 1998). Other attributes identified as necessary for FCEs are: accuracy, comprehensiveness, and credibility (these relate to validity); consistency, reproducibility and objectivity (these relate to reliability); generalisability; specificity; flexibility, measurability; standardisation and structure. Safety, practicality and utility are also criteria identified for FCEs

(Gibson and Strong 1997; Vert 2002; Innes and Straker 2003a; Gibson and Strong 2005; Matoushek 2008).

The following attributes can be considered more qualitative: practicality, safety, relevance and usefulness; and quantitative: objectivity; dependability; reliability; reproducibility; standardisation; consistency; measurability and structure. All of these attributes are recognised as contributing to ‘best practice’ (Innes and Straker 2002b).

Clinical utility of assessments reflects the degree of confidence that health professionals have about the usefulness of an assessment. Clinical utility is one measure that confirms that an assessment tool is related to the purpose for which it is used and is therefore an important component of an assessment. The following features of an assessment are perceived by evaluators to be evidence of clinical utility: accuracy; comprehensive range of items; credibility of evaluators; flexibility of the instrument; practicality; relevance; clinically useful results; suitability for the assessment purpose; feasibility of applications, adaptability, safety, reliability and validity (Toomey et al. 1995; Gibson and Strong 1997; Gibson and Strong 2002; Innes and Straker 2002b; Simmonds 2002; Vert 2002; Innes and Straker 2003a; Matheson 2003; Barbara and Whiteford 2005; James and Mackenzie 2009a).

The competence and training of the assessor is important and is highlighted as an essential component of FCE best practice (Abdel-Moty et al. 1993; King et al. 1998; Innes and Straker 2002a; Chappell et al. 2006; Isernhagen 2009; James and Mackenzie 2009a). Additionally, assessors need to engage in self-reflection and peer debriefing to achieve impartial feedback. These strategies

have been identified as core competencies for the application of FCEs in practice (Innes and Straker 2002a) and assist in the development of sound clinical reasoning skills by assessors (Chappell et al. 2006).

A client centred approach that meets the needs of the different stakeholders, such as the injured worker and the 'payer', that is also objective and unbiased is identified as important for 'best practice' for FCEs (Law 1998; Law et al. 2001; Chappell et al. 2006). However, within the concept of a client centred approach there is variation that is dependent upon the model of practice used. If a psychophysical approach is used, the client has more autonomy and has more control over their FCE results, whereas in the kinesiophysical approach it is the assessor who, based on observation of movement patterns (biomechanics) and physiological performance, determines safe maximum function levels.

Chappell et al (2006) suggest that multiple methods of data collection should be used in the FCE process, including interview, observation and experiential techniques, with triangulation of data recommended to ensure data is adequate and recommendations are appropriate and accurate. This is because both qualitative findings (such as observation of movement patterns or compensatory techniques) and quantitative findings (such as number of repetitions, frequency, duration) enhance the clinical reasoning processes used by an evaluator during the assessment and in the analysis of the data. Therefore all types of findings contribute to the comprehensiveness of the FCE process (Chappell et al. 2006).

Evidence about the psychometric properties of an assessment tool increase consumer confidence, informs best practice and is essential for the FCE process. Professionals are constantly encouraged to use evidence based

practice when treating or working with clients. Pressure to use evidence based practice comes from within health professional organisations and from those paying for services. Evidence based practice is considered to be the integration of clinical expertise and external clinical evidence (research) to make decisions regarding the care and treatment of individual clients (Law 2002). In the current occupational rehabilitation climate, with a focus on litigation, evidence for the use of specific assessment tools is required to allow appropriate decisions to be made based on accurate and meaningful FCE results.

2.6 The use of Functional Capacity Evaluations in practice – application to the work environment

The use of FCEs in practice is related to the purpose of the assessment and, as mentioned, there are different types of FCEs (Innes and Straker 2002a; Matheson 2003). Therefore the type of FCE and the processes used will vary according to the goals of the evaluation, the approach taken and the measurement instruments employed. Health professionals complete FCEs as an impartial independent clinician for compensation, litigation or disability level determination, or as a health professional involved in the treatment of an injured worker for return to work reasons (Menard and Hoens 1994). FCEs can provide information for return to work, case closure, identification of work restrictions or potential treatments (Matoushek 2008). The following applications of FCE in work-injury management have been identified:

- Determination of work function or work level;
- Return to work and job placement decisions and programs;
- Disability evaluation and ratings;
- Determination of function in non occupational settings;

- Medical and rehabilitation treatment intervention and planning; and
- Case management and case closure (Matoushek 2008).

A FCE also provides an opportunity for the injured worker to be educated in functional behaviours and appropriate body mechanics. A FCE can highlight the need for further medical investigation and preventative measures to avoid injury aggravation (McGuire 1995). As a return to work tool, FCE is often used by evaluators as a first step in defining the physical abilities and functional status of an injured worker, to provide information relevant for the development of appropriate interventions and a return to work plan including work modification and suitable duties (McGuire 1995). The match between FCE activities and work demands has been found for some activities (carrying, pushing, pulling, crouching, kneeling and bending) in a study of the Isernhagen Work Systems (IWS) FCE (Kuijer et al. 2006a). However, there is limited evidence related to the FCE process and the prediction of work ability or injury recurrence (Gross and Battie 2004; Gross et al. 2004; Kuijer et al. 2006a).

FCEs are also used as an assessment of incapacity for negotiation and settlement of a workers compensation or personal injury claim. Information to match the individual's functional capacity to their job requirements is important to provide an accurate picture of the feasibility of their return to work. FCE can therefore be seen as an objective measure to provide information to support re-employment recommendations in workers compensation cases (McGuire 1995).

However, as described above there is limited evidence with regard to the predictive validity of specific FCE tools in relation to return to work. In studies by Gross et al (2004 & 2005), better performance on the IWS FCE was linked with

faster time to suspension of benefits and claim closure (Gross and Battie 2004; Gross and Battie 2005b). Performance on the floor to waist lift appeared to predict outcomes, as well as information from the whole protocol, for those with chronic low back pain (Gross et al. 2004). In another study used to predict sustained recovery with chronic low back injury subjects, Gross and Battie (2004) identified that better performance on the IWS FCE was associated with a higher risk of recurrence of injury (Gross and Battie 2004). A study of workers with upper limb injuries found that greater weight lifted during the FCE was consistent with a faster time to benefit suspension and closure of the claim, although FCE performance was not related to future injury recurrence. Whether job demands were met or exceeded during the FCE did not relate to future return to work or work readiness (Gross and Battie 2006). More recently, in a study of a short form FCE, it was found that results provided information for predicting time to recovery but not injury recurrence (Branton et al. 2010).

These studies suggest FCEs are more able to assist in the prediction of return to work, benefit suspension and claim closure, but not in predicting injury recurrence. The changing nature of the environment and the long term effects of injuries on workers may impact upon the predictive ability of the FCE for injury recurrence. Currently, the limited evidence for predictability is an area for further study, particularly in relation to the range of specific FCE assessment tools available and the different contexts within which they are used.

As discussed, there are diverse purposes for conducting work-related assessments, which include FCEs; however, ten problems associated with FCE use were identified in a study by Innes and Straker (1998). These are:

- Confused definition of terms;
- Confused explanation of the conceptual basis of tools;
- Separation of the individual from their specific work context and the work environment;
- Lack of assessment of psychosocial and cognitive aspects related to work;
- Insufficient evidence of reliability and validity;
- Potential examiner bias;
- Inappropriate use of norm and criterion referenced data;
- Inappropriate statistical manipulation of results;
- Difficulty determining end points of safe functional ability; and
- Difficulty determining the sincerity of effort (Innes and Straker 1998a).

Issues related to terminology and conceptual frameworks used in FCE practice have been discussed earlier in this chapter. The focus on physical aspects of assessment within specific FCEs has been documented in the literature (Gibson and Strong 2003; Gross 2004; Gibson et al. 2005), with suggestions that FCEs should consider behavioural aspects as well as assessment of physical function (Gross 2004).

Consideration of the specific work environment is discussed in the literature and relates to the purpose of the FCE and whether this is in relation to a specific job or a more generic assessment. A FCE for disability cases or for those who have no job to return to, is often a more generic assessment. The ability to alter a FCE depending upon the needs of the client or the work situation has also been identified (King et al. 1998; Innes and Straker 2003a). Factors that have an impact on any adaptations include the reason for referral; any job requirements;

the client's type of injury; the workplace procedures, policies and resources of the FCE assessor; and the assessor's skill and experience (Innes and Straker 2002b; Strong et al. 2004a; Strong et al. 2004b; James et al. 2007). Health professionals have identified flexibility of the assessment tool as a consideration of its clinical utility when using an FCE. However, flexibility will also have an impact upon standardisation and reliability and validity of a tool (Cotton et al. 2006; James et al. 2007). Other issues identified by Innes and Straker (1998) relate to reliability and validity of FCEs, to the approaches used within specific FCEs and to the use of normative data or criterion referenced approaches, which include clinical observations. These features vary between specific tools (Innes and Straker 1998c).

An analysis of the assessments used within work rehabilitation programs in the USA occupational rehabilitation arena found that 94% of respondents conducted job analysis/risk hazard analysis and 91% conducted FCEs. Of the specific FCEs used, 45% used the IWS, 12% the Ergos and 12% the Key system (Jundt and King 1999). In comparison, an Australian study of occupational therapists working in occupational rehabilitation settings found that 96% of respondents conducted workplace assessments, 86% FCEs and 75% job/risk assessments. Of those who conduct FCEs, specific tools used were the WorkHab – 36%, Valpar – 23% and WEST – 18% (Deen et al. 2002). The differences between these findings may be the result of a focus on clinic based work hardening in the US and on return to work as part of a graded return to work program in Australia, and may be affected by the availability and cost of the specific FCE equipment in the respective countries. Cotton et al (2006) found that 75% of rehabilitation providers in NSW conducted their "own design"

FCE rather than a commercially available assessment tool. Of those using commercial tools, the WorkHab was the most commonly used tool (17%), followed by the IWS FCE (10%) (Cotton et al. 2006).

Eleven different FCEs were identified as being utilised in a study of rehabilitation provider practice in Australia, with many health professionals using more than one FCE. The most commonly used FCE was non-standardised (56%, n=43), followed by 52% (n=40) using the WorkHab, and 18% (n=14) using the Valpar (James and Mackenzie 2009b). Health professional practice and FCE usage is further discussed in chapter three.

Despite the widespread use of FCEs in the practice of work-related injury, the scientific evidence to support the psychometric properties of specific tools in measuring performance objectively, and producing reliable and valid results, is limited (King et al. 1998; Innes 2006). The need for research into the reliability and validity of these specific tools has been well documented in the literature, and research in these areas on some of the commercially available assessments has begun to be published more recently (Jay et al. 2000; Ting et al. 2001; Matheson et al. 2002b; Reneman et al. 2002c; Boadella et al. 2003; Durand et al. 2004; Gouttebarge et al. 2004; Reneman et al. 2004; Rustenburg et al. 2004; Reesink et al. 2007; Durand et al. 2008; Lechner et al. 2008; Streibelt et al. 2009).

2.7 The WorkHab Functional Capacity Evaluation

The WorkHab FCE is an Australian standardised assessment used to determine functional capacity in occupational rehabilitation and work environment settings. As previously noted, it is widely used in the Australian

context (Deen et al. 2002; Cotton et al. 2006; James and Mackenzie 2009b). The WorkHab FCE is based on objective, but limited, physiological measures, including heart rate; observations of the client's biomechanics and the client's reported pain and perceived exertion (effort) ratings during various activities (Bradbury and Roberts 1998). The WorkHab FCE includes a pre-screening assessment that includes a general health questionnaire, musculoskeletal evaluation, blood pressure measurement and a three-minute step test to determine heart rate recovery. This is followed by evaluation of: grip strength; push/pull; reaching; standing and sitting; walking and climbing; stooping; kneeling; balance; crouching and squatting; crawling; job simulation tasks; lifting tasks including floor to bench, bench to bench, bench to shoulder and bench to overhead height lifts; carrying tasks including both unilateral and bilateral carrying, and endurance lifting and carrying activities.

The manual handling component of the WorkHab FCE uses a modular box system that allows boxes to be stacked at various heights. The subject is instructed to lift the load box (initially empty) from beginning (e.g. bench) to end height (e.g. shoulder) and return. This is repeated three times before additional weight is added to the load box, until the safe maximum lift is reached.

The WorkHab FCE assessor reviews the heart rate readings as the assessment progresses (objective physiological measurements). The assessor observes the biomechanics of the lift; they record client reported pain, where the client reports their pain levels on a 10 point subjective rating scale, where 0 is no pain and 10 is the worst pain ever experienced. The assessor also records ratings of client perceived exertion (effort). Perceived exertion is rated on the Borg

Modified Rating of Perceived Exertion Scale, where 0 is rated as no effort and is 10 maximal exertion (Borg 1978). The assessor also records the weight lifted.

The WorkHab FCE uses the kinesiophysical approach where the evaluator determines when a client has reached their safe maximal lifting limit, unless a client chooses to cease the activity. The safe maximal lift is based on observation of unsafe lifting techniques such as the use of erratic movements, observation of compensatory lifting techniques, including a change of biomechanics, torque or postural loading, and/or an increase in heart rate exceeding predetermined levels of age-predicted maximal heart rate (Holtgrefe and Glenn 2007). The assessor also rates the safe lifting technique to calculate a manual handling score. This scoring is based on the five principles of safe manual handling. These are:

- Stance – ideal placement of feet at hip/shoulder width or slightly greater and a stable base of support throughout all tasks;
- Posture – ideal maintenance of normal lordosis (neutral spinal curves) throughout the lift;
- Leverage – keep loads close to the body and where possible in the range of centre of gravity;
- Torque – no rotation of the shoulder relative to the pelvis, no spinal twisting particularly of the low back while undertaking work positions; and
- Pacing/ timing – involves the use of smooth and controlled movement patterns at all times with no jerking (Bradbury and Roberts 1998).

Each of these components is rated on a scale of 0-4 with ‘0’ being no adherence and ‘4’ being the highest safety score. The sum of the score for each

component is recorded as the manual handling score for each subject. A higher score indicates that more appropriate manual handling techniques are being used. It is suggested that specific manual handling training is recommended for any client who scores under 70% on this manual handling score (Bradbury and Roberts 1998).

The WorkHab FCE takes about two and a half to three hours to complete and is completed with approval from the client's treating doctor. Subsequently, the evaluator prepares a report that details the client's functional performance during the FCE and discusses any limitations and recommendations in relation to work.

2.8 Evidence of psychometric properties of FCEs

Objective and accurate measurement is a pre-requisite for both the legal and medical stakeholders in the occupational rehabilitation system and provides health professionals with information to demonstrate the achievement of outcomes that indicate the success or effectiveness of services provided. Outcome measurement is becoming increasingly necessary to demonstrate the cost and clinical effectiveness of interventions.

Information obtained from FCEs is widely used to determine musculoskeletal capacity in relation to work tasks and this can have occupational implications for return to work after injury, and legal consequences related to compensation and disability payments (Pransky and Dempsey 2004). However, as previously noted, scientific evidence to support the measurement properties of specific FCEs and practical issues associated with these assessments continue to be a point of discussion.

Best practice is based on high quality evidence and this in turn relates to using a ‘gold standard’ assessment. However, in the case of FCEs, where there is no defined ‘gold standard’, it is necessary to consider the core attributes of an assessment (Chappell et al. 2006). These attributes, as previously discussed, include reliability and validity, which are important pre-requisites for accurate and meaningful measurement (Portney and Watkins 2009). Much literature discusses the need to establish reliability and validity for specific FCEs (King et al. 1998; Innes and Straker 1999a; Innes and Straker 1999b; Pransky and Dempsey 2004; Innes 2006).

Reliability relates to the extent that a measure is consistent and free from error, and this can be linked to the assessment tool or to the clinician performing the assessment. The context within which the reliability of a tool has been developed should be considered by the clinician when selecting an appropriate tool, so that the tool suits the situation in which it will be used in clinical practice (McFadyen and Pratt 1997; Portney and Watkins 2009).

Test re-test reliability (stability of the assessment and consistency of measures between testing occasions), intra-rater reliability (the degree to which one assessor or rater can obtain the same ratings on multiple occasions) and inter-rater reliability (the degree to which two or more assessors, measuring the same group of subjects, obtain the same ratings) need to be established for assessment tools (Portney and Watkins 2009).

Validity relates to the extent that an assessment measures what it is intended to measure. There are several types of validity that need to be demonstrated to have confidence in an assessment tool. Face validity (the extent to which a test

appears to address the construct of interest); content validity (the degree to which the items in an instrument adequately reflect the content domain being measured); construct validity (the degree to which a theoretical construct is measured by an instrument); criterion validity (the degree to which the outcomes of a test correlate with outcomes on a criterion test or external accepted standard); concurrent validity – a form of criterion related validity defined as the degree to which the outcomes of one test correlate with outcomes on a criterion test (or gold standard test) given at the same time; predictive validity – also a form of criterion related validity defined as the ability of an instrument to predict future performance; and responsiveness (the ability of a test to demonstrate change) are all important aspects (Portney and Watkins 2009). Construct validity is often considered the most important of these types of validity because a theoretical rationale is assumed to underpin the assessment. Reliability and validity are commonly expressed as correlation coefficients with results closer to +1.0 showing stronger reliability and validity.

Research studies that investigate aspects of reliability and validity of a selection of commercially obtainable FCEs have been published and several authors have reviewed this literature (King et al. 1998; Innes and Straker 1999a; Innes and Straker 1999b; Jones and Kumar 2003; Gouttebarge et al. 2004; Gross 2004; Innes 2006). While research has often been completed on aspects of a FCE (such as the manual handling component), a comprehensive discussion of the complete assessment is not always provided. Thus validity and reliability of some FCEs are established for only parts of the tool. In other cases, the research has been completed on small sample sizes or with specific groups of people, which limits the generalisability of the results.

Although FCEs are used in work injury management settings, research that aims to determine their measurement properties is often conducted with samples of healthy adults (Ting et al. 2001; Reneman et al. 2004; Soer et al. 2006a; James et al. 2010). Healthy subjects are often chosen for convenience, but this restricts the generalisability of any results to injured client groups. In the case of test-retest reliability studies, healthy participants overcome the possibility of results being affected by changes in injury or recovery status affecting performance between testing sessions compared to when injured workers are studied.

Access to an injured worker population for research can be problematic. In a test re-test study of the Physical Work Performance Evaluation (PWPE) in Australia, an extended time period (17 months) was required to obtain 31 injured worker volunteers because clients were reported as not interested or not prepared to be involved (Tuckwell et al. 2002). Other issues can affect the recruitment of injured workers, such as clients who are concurrently being managed in a litigious workers compensation system and who have reservations about putting their rehabilitation process at risk, and the potential for aggravation or re-injury as a result of participating in any research activities. Other studies both in the Netherlands and in Canada have not identified recruitment to be problematic; however, specialist facilities were used for recruitment, specifically an occupational assessment centre of a university rehabilitation centre, and a rehabilitation facility of a workers compensation board (Gross and Battie 2002; Reneman et al. 2002a). It is possible that any differences in workers compensation systems between countries may affect the willingness of injured workers to participate in research.

Participant recruitment issues in research are well documented and are not limited to FCE research. Some recruitment issues also relate to the involvement of clinicians, either directly as participants or as part of the recruitment process for patients and clients. In this case, factors that have been identified as barriers to recruitment include ethical problems or conflict of interest between the role of practitioner and the role of recruiter, forgetfulness of the recruiting personnel, lack of time, heavy caseloads and time spent in other research activities (Weierbach et al. ; Kadushin 2001; Serxner et al. 2004; Lannin and Cusick 2006).

In the Australian occupational rehabilitation arena, where FCEs are commonly conducted, services are provided on a fee-for-service basis to insurers or employers, and health professionals often work for profit driven companies. These health professionals usually have workload targets to meet and, combined with a lack of time and heavy workloads, this may play a part in the recruitment of these health professionals for research activities. An Australian study of occupational therapists found that those with research qualifications were more effective at recruiting patients; however, their judgements about patient compliance and obtaining informed consent also impacted upon recruitment decisions (Lannin and Cusick 2006). Reminders to health professionals about recruitment and dedicated recruiters have also been identified as factors to improve the recruitment rates of patients for clinical studies (Visanji and Oldham 2001; Lannin and Cusick 2006). The sampling used in research related to FCEs will be subject to these issues of recruitment, and subsequent limitations of a sample need to be considered when evaluating the evidence for any particular FCE.

2.9 A review of published studies related to the reliability and validity of FCEs

To further investigate the reliability and validity of specific FCE tools, a search of literature dated 1995–2010 was conducted using the databases MEDLINE, CINAHL, EMBASE and AMED. Search terms used related to the term functional capacity evaluation and its derivatives, specific FCE tool names and also search terms for reliability, validity, and psychometric properties. By a process of systematically reviewing the literature (Portney and Watkins 2009), 803 papers originally identified were reduced to 290 papers that were further evaluated based on the following inclusion and exclusion criteria.

Inclusion criteria included literature written in English, studies that presented data about FCEs (performance based assessments, physical capacity assessments, work productivity) and about the reliability and validity of the tools in adults. Descriptive studies or those that presented questionnaire or survey instruments, psychological assessments and studies that were not directly work-related were excluded. Following application of the inclusion and exclusion criteria, 56 papers met the criteria for review. Information related to the psychometric properties of FCEs from these papers is presented in Table 1.

Table 1 demonstrates that components of reliability and validity have been published for some of the commercially available FCE tools. The IWS FCE has been extensively studied, with more research completed on aspects of reliability and validity than any of the other FCEs identified. Some studies have compared tools, but without positive findings.

In one study the Ergo-kit and the IWS FCE were compared to see if these could be used interchangeably. This was not possible as there was no concurrent validity in the standard protocols for the upper limb lifting tasks (IJmker et al. 2003). The ERGOS and the Ergo-kit were also studied for concurrent validity; however, differences were found in both the lower and upper lifting results that suggests these tools should not be used interchangeably (Rustenburg et al. 2004). The ERGOs work simulator, however, was determined to provide information on strength and endurance for industrial physical activity that was comparable to a two-week multidisciplinary functional capacity evaluation (Dusik et al. 1993). Ruan et al (2001) studied the Functional Assessment Screening Tool (FAST), a test that measures activity tolerance but is not physically demanding, and the Progressive Isoinertial Lifting Evaluation (PILE) and found that poorer performance on the FAST was associated with poorer performance on the PILE (Ruan et al. 2001). A comparison of the PILE and the Work Well Systems FCE (WWS) reported that these should not be used interchangeably with patients with chronic low back pain (Soer et al. 2006b). The IWS was not correlated with self-report measures (Reneman et al. 2002c) and a comparison of self-report, clinical examination and outcomes of the IWS to assess work-related limitations in patients with chronic low back pain showed large differences in limitations identified (Brouwer et al. 2005).The floor to waist lifting task of the IWS and the recommended weight limit (RWL) of the National Institute for Occupational Health and Safety (NIOSH) guidelines also found different safe lifting limits for patients with chronic low back pain (Kuijer et al. 2006b).

Generally, studies that investigated the prognostic or predictive validity of FCEs had insignificant outcomes, while those that evaluated reliability had more positive outcomes.

Table 1: Summary of findings of functional capacity evaluation psychometric property literature

FCE method	Objective	Population	Procedure	Outcome	Author
AWP	Content Validity	67 administrators of AWP	Questionnaire on how AWP assesses the 3 domains of motor skills, process skills and communication skills	Motor: 89.2% Process: 76.2% Communication: 81.2%	(Sandqvist et al. 2008)
AWP	Construct Validity	364 clients (234 women + 130 men) with work related problems.	AWL assessment instrument	Principle Component Analysis: All items +'ve loadings 2 dimensions – motor + process and communication;	(Sandqvist et al. 2009)
Biomechanical methodology	Inter-rater reliability	30 adults (20 women, 10 men) 5 Raters	Lifting floor to waist level	Kappa ranged 0.56 to 0.82	(Gardener and McKenna 1999)
Blankenship FCE	Sensitivity and Specificity of submaximal effort	49 injured and non injured adults randomised into 2 groups (17 men+ 32 women). 2 raters	Blankenship protocol Repetitive movements Static strength Hand tests Lifting	Inter rater reliability – 81.6% Blankenship FCE validity score = 77.3% (100% effort gp) and 58.3% (50% effort gp).	(Brubaker et al. 2007)
BTE	Criterion validity	20 healthy adults (men)	VO2, Heart rate Endurance time Real and simulated lifting floor to bench	Differences between real vs simulated tasks Endurance :r=.71 VO2: HR correlation = 0.49 to 0.70	(Ting et al. 2001)
BTE Primus	Test-retest reliability	30 healthy men	Static strength and dynamic endurance testing with Primus assessed x2, 7 days apart.	Static ICC: 0.71 to 0.97 Dynamic ICC 0.32 to 0.90	(Lee et al. 2001a)

FCE method	Objective	Population	Procedure	Outcome	Author
BTE Primus + Valpar 19	To determine a work profile	30 carpenters 30 office worker Healthy	Upper limb strength Lifting	Chi square 6.48 (df=2, p<0.05)	(Lee et al. 2001b)
EPIC	Intra rater reliability	531 healthy adults (246 men+285 women) 110 evaluators	ELC protocol 2 testing occasions 1 week apart	ANOVA for gender: Max load p<0.0001 for men Intra rater, ICC: 0.84 to 0.94 Pearson r<0.91	(Matheson 1996)
Ergo –Kit and IWS	Concurrent validity of upper lifting tasks	71 Healthy adults (35 men, 36 women)	Upper lifting strength – Ergo-kit Waist to overhead lift -IWS	Pearson correlation: r= 0.72(p=0.00) mean dif = 6.2kg LOA: -2.4kg to 14.9kg	(IJmker et al. 2003)
Ergo-Kit	Inter-rater reliability Agreement	24 patients (10 men + 14 women) Low back pain in past 3 months 2 Raters	4 Ergo –kit lifting tests assessed on two occasions (3 days apart)	ICC: 0.94-0.97 SE of measurement (agreement) = 1.9 to 8.6kg.	(Gouttebarge et al. 2006)
Ergo-Kit	Construct Validity: Discriminative Convergent	72 Injured construction workers - men	5 Ergo-kit lifting tests + Von Korff questionnaire + Disability risk instrument	VK score and Ergo-kit lift (r=(-.29≤r≤0.5) Ergo-kit lift + disability risk – no significant difference (t=.27, P=.79; t=-.28, P=.78)	(Gouttebarge et al. 2009b)

FCE method	Objective	Population	Procedure	Outcome	Author
Ergo-Kit	Intra & inter-rater reliability	27 (15 men, 12 women) Healthy adults. 2 Raters	7 Ergo-kit tests assessed x3 over 4 days	ICC: >.80 – isometric lift ICC: >0.5-0.8 – dynamic lifts ICC:<.50 – manipulation tests	(Gouttebarge et al. 2005)
Ergo-Kit	Criterion related validity Concurrent validity Predictive validity	72 Injured construction workers	5 Ergo-Kit lifting tests+ Disability risk instrument (IDR) @ 6wks, 6 months and 1yr.	Concurrent validity: r=(-0.15 to 0.04)- isometric r=-0.47 to -0.31 dynamic Predictive validity IDR: r= -0.04 to -0.39 Predictive validity RTW: p≤0.03 (carry + lower lift)	(Gouttebarge et al. 2009a)
ERGOS	Reliability	12 healthy adults (6 men + 6 women)	ERGOS - Upper extremity 3 tests over 4 weeks	Learning effect for sensibility and coordination subsets.	(Boadella et al. 2003)
ERGOS and Ergo- Kit	Concurrent validity – maximum lifting capacity	25 healthy males	Lifting tests- Lower lifting Upper lifting	ERGOS vs Ergo-kit Lower lifing: 22.0kg vs 27.6kg Spearman correlation: 0.50 Upper lifting: 25.5kg vs 35kg. Spearman correlation: 0.66	(Rustenburg et al. 2004)
ERGOS work simulator	Development validity of job specific protocol	20 nurses 10 (heavier dept) 10 (lighter dept)	TRAC (task recording and analysis on computer) + exerted force (force gauge) of onsite observations	New job specific protocol (shorter) for FCE	(Frings-Dresen and Sluiter 2003)

FCE method	Objective	Population	Procedure	Outcome	Author
ERGOS work simulator **+Valpar** **+2 wk FCE**	Concurrent Validity	78 (70 men +8 women) workers compensation claimants	ERGOS test 24 subjects – VALPAR FCE by Rehab therapist	Pearson correlations R= 0.45 to 0.87	(Dusik et al. 1993)
FAST + PILE	Concurrent validity	181 patients (96 men + 92 women) with chronic spinal pain. 17 Healthy (9 women+ 8 men) adults	PILE lifting tests FAST (stoop, kneel, reach, squat, twist) Trunk extension endurance Psychological profiles Gps: Completers (all 5 fast tests completed) vs non completers (>5 fast tests completed)	Compl vs Non compl Trunk: (24.5s vs 52.7s). Pile weight lifted: (22.6% vs 19.1%)	(Ruan et al. 2001)
Functional Range of Motion Assembly Test	Test-retest reliability	51 healthy adults (42 women+9 men	Overhead reach Kneel reach Stoop reach 3-5days apart	Pearson correlations 0.65 to 0.96, ICC 0.42 to0.90	(Matheson et al. 2002b)
GAPP-FCE	Inter rater reliability	*5 healthy adults (pilot study) + 2 raters 5 rehab clients with CLBP (pilot) + 3 raters	GAPP-FCE	*Inter rater ICC= 0.88 Inter rater ICC = 0.72 to 0.8	(Gibson et al. 2005)
	Content validity	5 experts		Strength identified in using 3 models of FCE.	
	Inter rater reliability and predictive validity	14 rehab clients with CLBP + 2 raters		Inter rater ICC = 0.72 to 0.80	

FCE method	Objective	Population	Procedure	Outcome	Author
GAPP-FCE	Test-retest reliability	48 healthy adults	GAPP-FCE – 12 core components 2 sessions 7-14 days apart.	ICC physical level of work = 0.86 and 0.93. ICC : Range for individual components = 0.15 to 0.94	(Gibson et al. 2010)
GAPP-FCE	Content validity using expert panel	5 expert s	Questionnaire on FCE inclusion, utility etc.	Good support for aspects of content validity and technical quality.	(Gibson and Strong 2002)
IWS	Ecological validity	24 healthy adults (12 men+ 12 women)	IWS static endurance tests: Overhead work, Crouching Kneeling Manipulation of environment.	Difference between conditions: Overhead work – slight difference – (MANOVA P=0.128) Crouching – no sig difference (P=0.895) Kneeling – (P=0.034)	(Reneman et al. 2001)
IWS	Inter and intra rater reliability	4 (men) Healthy adults 5 Raters	Video analysis of lifts- IWS	Inter-rater: 87-96% agreement Intra rater: 93-97% agreement	(Reneman et al. 2002b)
IWS	Test-retest reliability	50 patients (39 men + 11 women) with non specific CLBP	IWS lifting tests- Lifting low Lifting overhead Carrying Testing completed 2 days apart.	ICC: Lifting low: 0.87 Lifting OH: 0.87 Carrying: 0.77	(Reneman et al. 2002a)

FCE method	Objective	Population	Procedure	Outcome	Author
IWS	Test-retest reliability	26 Healthy adults	IWS- 28 tests completed in 1 day. Material handling Criterion +Ceiling tests (2-3 weeks apart)	Material handling : ICC: 0.68 to 0.98 LOA: 4.8 to 21.5kg Criterion/ceiling test: Kappa= <0.60, % agreement = <80%. ICC: range 0.58 to 0.93	(Reneman et al. 2004)
IWS	Concurrent validity	92 patients (60 men + 32 women)with non specific CLBP	IWS - Floor to waist lift RWL of NIOSH lifting guideline	Mean different FCE and RWL = 15.0kg	(Kuijer et al. 2006b)
IWS	Inter-rater reliability Test-retest reliability	28 adults with LBP 5 raters	IWS – floor to waist, waist to crown, horizontal lifting, carry 2 testing occasions 2-4 days apart.	Test-retest: ICC 0.78:0.94 Inter rater: ICC 0.90+	(Gross and Battie 2002)
IWS	Inter rater & intra rater reliability	3 men on workers compensation benefits 12 Raters	IWS – floor to waist Carry Waist to crown	Kappa 0.68 (light medium, heavy) Kappa 0.81 (light, heavy) % agreement =< 90%	(Isernhagen et al. 1999)

FCE method	Objective	Population	Procedure	Outcome	Author
IWS	Predictive Relationship of FCE and RTW	650 adults with functional limitations	IWS protocol Lifting Grip force	Ch sq, ANOVA Multiple regression RTW performed better than no RTW, $p<0.05$ Regression: 80.3% (RTW) 56%(no RTW).	(Matheson et al. 2002a)
IWS	Construct Validity relationship of FCE + clinical +psychosocial factors	170 workers compensation claimants	IWS protocol Floor to waist lift Failed tasks + questionnaires	Correlation (r 0.2 to -0.7) Regression: adjusted R2:0.02 to 0.2	(Gross and Battie 2005a)
IWS	Construct Validity + PDI + VAS	321 adults with LBP	IWS protocol PDI + VAS	Pearson correlation IWS:PDI r= -0.44 to 0.52 IWS: VAS r= 0.79	(Gross and Battie 2003)
IWS	Predictive validity – matching to work demands	18 workers (11 men + 7 women) with CLBP	IWS protocol + questionnaire + workplace assessment	7 FCE activities matched with work – carry, push, pull, crouch, kneel, static forward bend, dynamic bend + rotating. Lifting indirectly matched with WPA data.	(Kuijer et al. 2006a)
IWS	Prognostic value with patients with CLBP	150 WC claimants with CLBP	IWS FCE + failed tasks	Time to claim closure Low association of IWS and recovery.	(Gross et al. 2004)

FCE method	Objective	Population	Procedure	Outcome	Author
IWS	Prognostic value with patients with CLBP	226 WC claimants with CLBP	IWS, failed tests floor to waist lift + days to benefit suspension	Logistic regression. Pts 6% less likely to experience future recurrent for each failed test (OR:0.94).	(Gross and Battie 2004)
IWS	Correlation between FCE, self report + clinical examination	92 Patients with CLBP	IWS FCE + Clinical examination + Self report	FAL (functional ability list). Little agreement found Kvalues:- 0.05 to 0.32.	(Brouwer et al. 2005)
IWS	Predictive validity for upper limb disorders	336 WC claimants with upper extremity disorders	IWS + days to suspension of benefit.	Higher weight lifted on waist to overhead lift modestly associated with faster benefit suspension (r2 Cox regression 2-5%)	(Gross and Battie 2006)
IWS	Predictive validity	130 WC claimants with CLBP	IWS	Fewer failed tests (HRR 0.94) and higher floor to waist lift (HR 1.38) associated with faster RTW	(Gross and Battie 2005b)
IWS	Test-retest reliability	30 patients with CLBP	IWS FCE 2 session held 2 wks apart.	Kappa, % agreement, ICC ICC ranged: 0.39 to 0.96 15 /19 tests showed acceptable agreement 11/18 tests showed acceptable reliability.	(Brouwer et al. 2003)
IWS +Roland + Oswestry + Quebec disability questionnaires	Concurrent validity of self report and performance based approach to assessment	64 Patients (54 men + 10 women) with non specific CLBP	Self report- Roland then 1-2 wks later Quebec + Oswestry questionnaires + FCE	Correlations: FCE with Roland, -0.2; Oswestry,- 0.52; Quebec, 0.50 Ro/Os/ Qu (range 0.50 – 0.74 – all p=<0.01)	(Reneman et al. 2002c)

FCE method	Objective	Population	Procedure	Outcome	Author
Job Fit	Test-retest Reliability Inter / intra rater reliability	28 healthy men 2 raters	Pre employment FCE Lifting	ICC. Test-retest: 0.56 to 0.88 Inter rater: 0.89 to 0.96 Intra rater: 0.86 to 1.0	(Legge and Burgess-Limerick 2007)
LIDO workset	Test-retest reliability	19 subjects with motor & sensory neuropathy, motor dystrophy + 11 healthy subjects	LIDO: a. Knob turn b. Linear motion c. Lever arm (1 week apart)	ICC – total work 0.87 to 0.91 0.85 to 0. 97 0.72 to 0.90 ICC - peak torque 0.88 to 0.94 0.85 to 0.96 0.67 to 0.91	(Kilmer et al. 2000)
PILE	Inter- rater reliability Test-retest reliability	22 nursing aides/ assistants 2 raters.	PILE lifting tests (2) Standing balance Isometric endurance Two testing occasions. (5 + 16 days apart)	Inter-rater: ICC Balance : 0.99 to 1.0 Pile lift lumbar: 1.0 Pile lift cervic.: 1.0 Test-retest: ICC Balance: 0.72 to 0.83 Pile lift lumbar: 0.69 to 0.71 Pile lift cervic. 0.88 to 0.96	(Horneij et al. 2002)

FCE method	Objective	Population	Procedure	Outcome	Author
PILE	Test-retest reliability	31 patients (17 women + 14 men) with LBP	PILE lifting tests (2 days apart)	ICC: 0.91 Error of measurement: ±4.5kg	(Lygren et al. 2005)
PILE	Test-retest reliability Agreement	53 rehabilitation patients (30 Gp A no experience; 23 Gp B – with experience	a.Walking – 5 min/ 50ft b.Sit to stand c.Loaded forward reach d.Stair climbing e.PILE lifting tests (5-7 days apart)	a. ICC:0.89/ 0.76 b. ICC: 0.91 c.ICC: 0.74 d ICC: 0.96 d. ICC: 0.92	(Smeets et al. 2006)
PILE	Sensitivity and specificity of determining effort	90 rehabilitation clients (44 into 60% effort gp, 46 into 100% effort gp).	PILE lifting protocol floor to bench lift Handgrip	Specificity = 84% Sensitivity = 65%	(Lemstra et al. 2004)
PILE+ WWS	Concurrent validity	53 patients (32 men + 21 women) with CLBP	Lifting tests – PILE Lifting tests – WWS	Pearson correlation = 0.75 (p=<0.01) Mean difference = 6.0kg on WWS compared to PILE (p=<0.01).	(Soer et al. 2006b)
PWPE	Predictive validity	30 compensation patients with musculoskeletal dysfunction	PWPE protocol of 36 tasks	RTW: Kappa 0.74 RTW modified: Kappa 0.69 No RTW: Kappa 0.70	(Lechner et al. 2008)
PWPE	Inter rater reliability Concurrent criterion related validity	50 adults with musculoskeletal disorders 2 raters	PWPE protocol	Kappa Inter rater :0.54 to 0.83 Spearman rho 0.41 to 0.55 (p=0.002 to p=0.001)	(Lechner et al. 1994)

FCE method	Objective	Population	Procedure	Outcome	Author
PWPE	Test-retest reliability	24 clients with stable physical injuries	9 of 21 tasks of PWPE: a. lifting b. carry c. pushing d. sitting e. standing f. kneeling g. stair climbing h. squatting (rep) i. walking	PWPE: Kappa/% agreement a. 0.77/ 87.5% b. 0.75 / 87.5% c. 0.75 / 87.5% d. 0.38/ 66.7% e. 0.60 / 79% f. 0.70 / 83% g. 0.19/ 78% h. 0.60 / 83% i. 0.37 / 66.7%	(Tuckwell et al. 2002)
PWPE	Relationship between work function and LBP	100 industrial metal workers 2 gps – with and without past/ present LBP 2 Evaluators	PWPE dynamic strength + questionnaires	T tests, logistic regression. floor to waist, p= 0.02 Waist to eye, p=0.03 Bilat carry, p=0.007 Unilat carry, p=0.005 Push, p=0.1 Pull, p=0.056 Weights lifted lower in LBP gp ($p<0.05$).	(Ratzon et al. 2007)
PWPE	Responsiveness Comparison of 2 groups	27 Adults with non specific LBP 30 Healthy adults	PWPE 2 tests 6 wks apart	Wilcoxon Mann-Whitney: p value 0.06 (overall).	(Durand et al. 2008)
PWPE	Inter rater reliability	40 adults with LBP 2 raters	PWPE	Kappa 0.76	(Durand et al. 2004)

FCE method	Objective	Population	Procedure	Outcome	Author
Short form FCE (IWS)	Development and Predictive validity	183 claimants with LBP + 132 claimants with LBP	IWS _ floor to waist lift Crouching Standing	Days to suspension of benefits 5 items predictive α0.05	(Gross et al. 2006)
Short form FCE (IWS)	Predictive validity	147 workers with musculoskeletal injuries tested on Short form 199 workers tested on IWS	IWS +PDI + VAS	FCE job demand levels more likely to have claim closed (HRR 5.52) & benefits suspended (HRR 5.45) over the year follow up.	(Branton et al. 2010)
WRULD FCE	Test-retest reliability	33 healthy adults (14 men + 19 women).	8 tests including: Lifting Grip strength Pinch strength Wrist extension strength (10 days apart)	ICC ranged 0.73 to 0.97	(Soer et al. 2006a)

AWP = Assessment of Work Performance
BTE = Baltimore Therapeutic Evaluation
FAST = Functional Assessment Screening Test
GAPP-FCE = Gibson Approach to Functional Capacity Evaluation
IWS = Isernhagen Work Systems
PDI = Pain Disability Index
PILE = Progressive Isoinertial Lifting Evaluation
PWPE – Physical Work Performance Evaluation
WWS = Work-well Systems Functional Capacity Evaluation
WRULD = Work Related Upper Limb Disorder
VAS = Visual Analogue Scale

2.10 Chapter Summary

For any test or measurement used in any setting, scientific evidence should be available to contribute to evidence based practice. Reliability and validity should be demonstrated. This has been achieved in part for some commercially available FCE tools (King et al. 1998; Innes and Straker 1999a; Innes and Straker 1999b; Gouttebarge et al. 2004; Gross 2004; Innes 2006). The IWS FCE has been extensively studied for reliability with a variety of populations and for aspects of validity. As Table 1 indicates some of the psychometric properties for the IWS are stronger than others. The Ergo- Kit, the PWPE and the GAPP-FCE tools have also had a number of studies investigate reliability with encouraging results in different populations. The results from validity research are less convincing, which suggests that further study is needed, particularly in relation to the predictive validity of FCEs.

No articles related to the WorkHab FCE other than those presented as part of this thesis were identified in the literature search. These articles were not included in Table 1, but are fully discussed in relation to reliability and validity in subsequent chapters of this thesis.

This literature review identifies the lack of scientific evidence to support the measurement properties of specific FCEs and the absence of evidence for the WorkHab FCE. To investigate the usage of FCEs by health professionals in the Australian occupational rehabilitation arena, qualitative and quantitative methodology was employed to study current practice and to determine what FCEs were being used within this context. This was completed to evaluate the attitudes, perceptions and behaviours in relation to FCE use, the factors that

influence the clinical judgements in relation to FCE and the factors that influence the selection of any FCE tool, as well as the investigation of the level of usage of the WorkHab FCE in this setting. This was completed with a view to provide justification for further study of the reliability and validity of the WorkHab FCE. Chapter three discusses the results of the studies that investigated the current practices of FCE use in the Australian context.

Chapter 3 Current Practice with Functional Capacity Evaluations

Overview: This chapter discusses two studies (reported in three manuscripts) that investigated how FCEs are currently used in practice in the Australian occupational rehabilitation arena. The attitudes, beliefs, perceptions and experiences of health professionals in the use of FCEs, factors that influence the selection of any FCE tool and the level of usage of the WorkHab FCE were examined. These studies were completed with a view to provide justification for further study of the reliability and validity of the WorkHab FCE.

Functional Capacity Evaluations are used extensively in occupational and vocational rehabilitation to determine the capacity of injured workers to return to work. They are also used to screen potential employees for specific jobs, to assess physical rehabilitation needs for rehabilitation planning, to monitor progress, and to determine a person's functional capacity for compensation or litigation purposes (King et al. 1998; Strong et al. 2004a; Innes 2006).

Current practices and current beliefs of therapists in relation to workplace assessments and functional capacity evaluations in Australia were investigated by Innes, and it was identified that assessments were grouped into workplace assessments and two types of functional capacity evaluations: those for no specific job (FCE (no job)) and those for a specific job (FCE (job)) (Innes and Straker 2002b). This study indicated that therapists believed themselves to be acting as the main assessment tool in a FCE, and their credibility was therefore central to the credibility of the assessment (Innes and Straker 2003b). The study also found strategies used by therapists in FCEs could be described as

quantitative based approaches in the case of FCEs for clients without a specific job, and a combination of qualitative and quantitative approaches were used in FCEs for those with a specific job to return to.

FCEs are used for common purposes internationally and, despite differences in practice settings and legislation, the practice issues facing health professionals in relation to FCEs are similar (Innes and Straker 2002b; Pransky and Dempsey 2004; Strong et al. 2004a).

This chapter discusses three published manuscripts and describes studies that aimed to investigate the attitudes, behaviours, perceptions and practices of Australian health professionals in relation to the use of FCEs, to investigate health professionals' views on the clinical utility of FCEs and to investigate the level of usage of the WorkHab FCE.

To describe, understand and interpret experiences from the perspective of the health professionals who conduct FCEs as part of their work within occupational rehabilitation, a phenomenological, qualitative study design was utilised to explore the attitudes and practices to FCE use (Deen et al. 2002; Bowling 2004). In-depth, semi structured interviews were conducted with five health professionals. The results are discussed in **Manuscript 1 (3.1):** *"Health professionals' attitudes and practices in relation to functional capacity evaluations"* published in 2007 in Work: A Journal of Prevention and Rehabilitation, 29(2), 81-88. It was found FCEs are applied differently according to the reason for referral and client goal, the health professional's workplace procedures, policies and resources and the health professional's skill and experience. Results also indicated that health professionals adapted the FCE to

suit their requirements. This study identified further options for research including a larger survey based study to include a bigger sample and allow a more in-depth understanding of FCE usage in the Australian context.

As a result of this qualitative study, a survey was developed to examine a larger sample of health professionals and to allow a more in-depth understanding of health professionals' attitudes and practices, and the implications of these findings on the reliability and validity of current practice in relation to FCEs. A quantitative cross-sectional study design was used to survey health professionals who conduct FCEs and who worked for rehabilitation providers in NSW, Australia. The results of this survey are discussed in **Manuscript 2 (3.2):** *"Health professionals' perceptions and practices in relation to functional capacity evaluations: Results of a quantitative survey"* published in 2009, in the Journal of Occupational Rehabilitation, 19(2) 203-211 and in **Manuscript 3 (3.3):** "*The clinical utility of functional capacity evaluations: the opinion of health professionals working within occupational rehabilitation"* published in 2009, in Work: A Journal of Prevention and Rehabilitation, 33(3) 231-239.

The survey study identified the most commonly used FCE to be a non-standardised tool followed by the WorkHab and the Valpar. It identified that health professionals used parts of FCEs and chose to adapt the FCE to suit the client, injury type and job, rather than use standardised measures. However, important criteria were identified as standardisation, reliability, and validity plus task characteristics and flexibility within the FCE, despite the impact of flexibility and adaptability on the psychometric properties of a tool. Health professionals reported that a FCE assisted in predicting return to work and manual versus sedentary duties and felt FCE practice was a specialist area that required

specific training. This study identified that further research to explore the use of evidence in FCE practice and to provide evidence of the reliability and validity of the tools in use is needed.

This research has also been presented as a poster at the 14th Congress of the World Federation of Occupational Therapists in Sydney, New South Wales, Australia in July 2006 (Appendix 2) and as an oral poster at the OT Australia 23rd National Conference in Melbourne, Victoria in September 2008 (Appendix 2).

3.1 Manuscript: Health Professionals' Attitudes and Practices in relation to Functional Capacity Evaluations

Citation: James, C., Mackenzie, L. and Higginbotham, N. (2007). "Health Professionals' Attitudes and Practices in relation to Functional Capacity Evaluations." Work 29 (2):81-88.

ABSTRACT

Functional Capacity Evaluations (FCEs) are part of practice in work injury prevention and rehabilitation, and are designed to define an individual's functional abilities or limitations in the context of safe, productive work tasks.

Qualitative research methodology was used to investigate the attitudes and behaviours of health professionals in relation to FCE use. The study aimed to identify why health professionals chose a particular FCE, and to identify what factors influence health professionals' clinical judgements when providing results and recommendations for the individual being assessed. Five health professionals from the Hunter Region of New South Wales, Australia participated in semi-structured, individual interviews using a phenomenological approach. Following inductive analysis of the data, four themes reflecting participants' attitudes and behaviours of FCE use emerged: i) referrals and expectations, including why and when the assessment is completed ii) outcomes, - what the results aim to provide iii) workplace / practice / usage issues and iv) skills of the assessor.

The results indicate the need for further research on the clinical utility of FCEs. A large scale quantitative study would allow results to be generalised to a wider community of FCE users.

KEYWORDS: Functional Capacity Evaluation, Occupational Rehabilitation.

3.1.1 INTRODUCTION AND LITERATURE REVIEW

Functional Capacity Evaluations (FCEs) are part of practice in work injury prevention and rehabilitation, with the aim of defining an individual's functional abilities or limitations in relation to work tasks (King et al. 1998). They are commonly used with individuals who have suffered work related musculo-skeletal injuries, to assist in decision making about return to work, entitlements and rehabilitation (Strong et al. 2004a).

There are many different FCEs available commercially and many clinicians/rehabilitation providers have developed their own non- standardised, work specific FCEs. All FCEs attempt to measure functional performance objectively. However, there are limited published studies appraising the reliability, validity and utility of the assessments, to establish if this objectivity is achieved (King et al. 1998; Innes and Straker 1999a; Innes and Straker 1999b; Innes and Straker 2003; Strong et al. 2004a). Many of the studies that have been completed do not relate to tools developed in the Australian context (Innes and Straker 1999a; Matheson et al. 2002; Boadella et al. 2003; Strong et al. 2004a). Strong et al (2004) identified limitations of current research and difficulties of extrapolating information from a single point in time in the assessment process (Strong et al. 2004b). They also found that FCEs were conducted with limited contextual information, with variations in guidelines or practice standards and that practices were influenced by referral source and market demands (Strong et al. 2004a). Pransky and Dempsey (2004) suggest the dynamic nature of a job and of capacity, and of the differences between tasks completed by employees of the same job description present issues for FCEs. It is suggested a job analysis is required to identify the specific tasks within a job and absence of formal job evaluation constitutes a threat to validity (Innes and Straker 2003; Pransky and Dempsey 2004).

Within Australia, Innes and Straker (2002) studied the current practice of therapists in relation to work assessment and found that generally assessments were grouped into work assessments, FCE (no job) and FCE (job) (Innes and Straker 2002). A variety of factors affecting the type, purpose and characteristics of the assessment and their influences and constraints were described, Deen et al (2002) surveyed Occupational Therapists in Australian work practice and found 96% conducted Workplace assessments, 86% Functional capacity evaluations and 75% Job/risk assessments. Of those conducting FCEs, specific tools being used were, WorkHab- 36%, Valpar – 23% and West – 18% (Deen et al. 2002). However, Innes and Straker (1999) found none of these tools had adequate documentation for validity. There are some more recent studies that investigate components of specific assessment tools in relation to validity

and reliability (Deen et al. 2002; Matheson et al. 2002; Boadella et al. 2003; Durand et al. 2004), however little is known about therapists' attitudes and practices in relation to the different FCE tools. Literature regarding attitudes to practice only illustrated articles related to clients or related groups rather than assessment tools.

The current beliefs of therapists in Australia in relation to workplace and functional capacity evaluations were also studied by Innes and Straker (2003). In this study it was found that therapists believed they, as therapists, were the assessment tool, and were central to the credibility of the assessment (Innes and Straker 2003). They also found many of the strategies used by therapists in FCEs were similar to those used in qualitative research, such as using multiple data sources and methods of data collection, collecting information until no new data is gained, triangulating data sources, and member checking to confirm results from the assessments.

This study aims to build upon the existing studies of FCEs to investigate the attitudes and behaviours of Australian health professionals in relation to Functional Capacity Evaluation (FCE) use, to identify why health professionals chose a particular FCE and to identify what factors influence health professionals' clinical judgements when providing results and recommendations for the individual being assessed. It is a precursor to a more in-depth quantitative study measuring which FCEs are in use.

3.1.2 METHODOLOGY

Research Design

A phenomenological, qualitative study design was utilised to explore the attitudes and practices of health professionals who use Functional Capacity Evaluations as part of their work within occupational rehabilitation.

This approach seeks to describe, understand and interpret experiences from the perspective of those experiencing the phenomenon (Deen et al. 2002; Bowling 2004).

Participants

Four occupational therapists and one physiotherapist participated in the study. They had been qualified between 1 year and 6 years, and worked for a range of public and private rehabilitation providers. They had been conducting FCEs for varying lengths of time ranging from 6 months to 4 years. All participants, except the Physiotherapist, had worked solely in the area of occupational rehabilitation (See Table 1).

Table 1: Details of Participants

Participant	Gender	Profession	Years of Experience	Time conducting FCEs
Participant 1	**Male**	**OT**	**4 years**	**3.5 years**
Participant 2	**Female**	**OT**	**3 years**	**3 years**
Participant 3	**Female**	**OT**	**1 year**	**1 year**
Participant 4	**Female**	**PT**	**6 years**	**1 year**
Participant 5	**Male**	**OT**	**1 year**	**6 months**

Procedure

Recruitment of health professionals (eg. physiotherapists, occupational therapists, occupational health nurses) who conduct FCEs and work for WorkCover (NSW) accredited rehabilitation providers in the Hunter region of NSW, took place following ethics approval from the University Human Research Ethics Committee. Contact details of providers in the Hunter region (16 providers), were obtained from the Yellow pages phone directory. A letter was sent to the manager of each provider asking them to distribute information letters and consent forms to therapists or nurses who conduct FCEs for the organisation. Participants returned consent forms to the researcher in pre-paid envelopes, and were then contacted to arrange a mutually convenient time and place for the interview.

Data Collection

Data were collected by in-depth, one to one interviews using a semi-structured interview schedule (Appendix A). This schedule was developed by the researcher from existing literature about FCEs; the use of the assessment; and following consultation with a senior academic experienced in qualitative research methodology (Carpenter and Hammell 2000). The schedule provided a framework of topic areas to be addressed in the interviews and ensured similar issues were explored with each participant (Patton 2002). The use of open ended questions encouraged descriptive responses to be given according to each participant's own narrative style and facilitated opportunities for the researcher to probe further into participant experiences (Suto 2000). Interviews ranged in length from 60 - 90 minutes. Each interview was audio-taped, to increase accuracy of data, and was transcribed verbatim.

Data Analysis

Transcripts were coded inductively, whereby phrases, sentences or words were coded according to the topic or issue being discussed (Deen et al. 2002). These codes were

then grouped into categories using the constant comparison method (Morse and Field 1995). This method involves comparing and contrasting new information with previously obtained data. Next, the data assigned to the various categories was analysed, and patterns, similarities and relationships were grouped into larger categories or themes.

To increase data credibility reflexivity was used. This is a process of self examination whereby reflection on bias, theoretical predispositions and perspectives, and how this had influenced data collection and analysis took place using a personal diary to record thoughts, feelings and ideas (Barry et al. 1999; Deen et al. 2002). Member checking was also conducted by contacting some participants by phone to discuss themes derived from the data and the researcher's supervisor also reviewed and provided feedback on the analysis.

Triangulation was used to enhance rigour throughout the interview development stage, data collection and analysis. This is the use of multiple methods to cross check the validity and offers deeper insight into the relationship between inquiry and the phenomenon under study (Patton 2002). Triangulation was accomplished using existing literature, reflexive analysis as the interviews progressed and through member checking.

3.1.3 RESULTS AND DISCUSSION

Four key themes emerged from the data relating to health professionals' attitudes and practices in relation to functional capacity evaluations. These were; i) referrals and expectations, ii) outcomes, iii) workplace / practice / usage issues and iv) skills of the assessor.

Referrals and expectations

Several factors emerged from the interviews, relating to referrals and the expectations of the referrer, the employer, the insurer and the doctor.

All participants commented that the reason for referral or clients' goal, affected the FCE completed. If a client had a job to return to, all participants stated that tasks related to the specific job would be included in the FCE, and that if the client did not have a job, a more general FCE would be completed.

Referral reasons can be expanded upon, as Participant 2 states: *'a lot of them are for vocational retraining…..to clarify the job they are looking at is indeed suited to their functional ability.'* Participant 1 also added *'the legal status of the client makes a difference to the FCE.'* This was specifically in relation to Section 40 Assessments

which are assessments completed under Section40A of the NSW Workers Compensation Act 1987, of a partially incapacitated injured worker's ability to earn in some suitable employment . With regard to Section 40 Assessments a more general FCE was conducted as the goal is to look at the capacity of the client to earn rather than a specific job goal.

Participant 3 stated *'Section 40 FCEs (i.e. with no job to go back to), should be exactly the same for everyone - it should be quite big and cover everything, whereas a rehab FCE should be set up depending upon exactly what duties, what job they have and what injury.'* However, in relation to a Section 40 assessment being exactly the same for everyone, this was not the case, even within this small sample where a range of FCEs, not one standardised FCE, were being used for all assessments.

Two distinct forms of FCE - no job and job, depending upon the employment status of the injured worker and the potential for returning to the pre-injury workplace have been identified (King et al. 1998; Innes and Straker 2002). An FCE (no job) is more comprehensive and assesses generic work skills and physical demands, whereas the FCE (job) has a more job specific focus and includes job simulation tasks. However, as a result, standardisation of the specific tasks used was precluded because of customisation for job simulation (Innes and Straker 2002). In relation to FCE (job) when return to work is the major focus, it has been suggested a job analysis should be performed to determine the tasks required for the job. The results of the FCE can then be compared with the physical requirements of the job (King et al. 1998; Pransky and Dempsey 2004; Strong et al. 2004a; Strong et al. 2004b).

Insurers, employers, solicitors and a range of health professionals referred for an FCE, and this along with the reason for referral impacted upon the type of assessment that was performed. Participant 2 stated *'they (the insurer) are looking for a stronger opinion on whether the person can or cannot do something.'* This is consistent with previous findings indicating that legislation and related regulations, and the expectations of referrers impacted upon the assessment and when it was performed (Innes and Straker 2002).

In relation to the views of other professionals on the FCE, Participant 4 commented that *'Doctors are responsive to the FCE, as it gives them something concrete to put on a medical certificate, and 'it can be used to identify functional abilities when there are discrepancies and differences in the goals of rehab, between the Doctors and the employee.'* Participant 3 also commented that Rehabilitation Counsellors view the FCE positively as it gives them an indication of the clients' functional abilities and limitations.

Strong et al (2004) studied the users' perceptions of the FCE reports and found that FCEs were being described as useful information tools

Outcomes

The outcome of the FCE was discussed as being related to the goal of rehabilitation and, specifically related to whether the client had a job to return to or not, and the legal status of the client. The ability of the FCE to predict outcomes was discussed as was the limitations of the assessment and the recommendations made as a result of the assessment.

Several participants commented about the ability of the FCE to predict a clients' abilities:

'The FCE is good at predicting whether a person can go back to suitable duties, but not always a predictor of returning to normal duties.' (Participant 1).

'It clarifies a client's abilities and limitations …. is a starting point for rehab.' (Participant 3). Participant 2 stated *'the FCE is able to predict if someone is suitable for sedentary or manual work.'*

The limitations of the FCE in relation to predicting outcomes was also commented upon, Participant 5 stated *'the FCE is able to determine if a person is not able to return to pre-injury duties as this is beyond their functional capacity, however the assessment is only an accurate indication of what the client can and can't do at the time of testing.'*

Participant 3 continues: *'the FCE is not there to predict vocational outcomes for those without a job … but makes sure a person can do a proposed outcome (job type) and prevents putting them in a situation where they won't cope or where there is an increased chance of injury or aggravation.'*

Participant 2 stated: *'the FCE we use is able to indicate if a client needs counselling for fear avoidance behaviours or cardiovascular conditioning.'* The FCE was discussed by this group as providing information which allows recommendations in relation to work, strengthening or other services that may assist the clients' rehabilitation.

All participants commented that the recommendations made in relation to the FCE related to consistencies and inconsistencies of performance observed during the assessment, as was also found by Strong et al (2004).

It is interesting to note that the purposes documented in the literature relating to why FCEs are conducted discuss the need to match the worker and the work duties and to identify the individuals' physical abilities and limitations for employment (King et al.

1998). However, any FCE will only give a picture of the time of the assessment, and the ability to predict return to work or injury recurrence has not been proven (Pransky and Dempsey 2004; Strong et al. 2004a). The reliability and validity of the FCE is also questioned if FCEs are adapted to suit each individual injured workers' situation.

Workplace/Practice/Usage Issues

The therapists' workplace had an impact upon FCE use. All participants stated the FCE they used was the result of what was available at their workplace, and what the referrer requested. The general feeling of participants was that this also related to economic issues of what was cost effective for their employer and the payer. Participant 4 commented: *'initial set up cost for the FCE and the equipment needed is a factor that my employer would consider.'* Strong et al also found referrers wanted value for money spent (Strong et al. 2004b).

The therapists interviewed identified a range of FCE assessments they currently use and some that they had used with previous employers. Some therapists discussed adapting the assessment to suit the clients' injury and job, and others discussed using parts of an assessment rather than the whole.

Usage was also discussed as being related to what the referrer wanted. Participant 2 stated: *'large employers want a standardised FCE.'* However, Participant 1 contradicted this comment saying *'employers request the FCE to be job specific so requiring the adaptation of components.'* This participant later added *'the ability to adapt components of the FCE is desirable to suit the individual and the injury.'* The other factor that was commented upon was the issue of the assessment being from a clinical or a functional viewpoint, which also relates to the issue of making the assessment job or work focused. Participant 3 stated: *' look at kneeling, if someone can't kneel it doesn't necessarily mean they can't work at ground level adopting some other posture.'*

Linking to this issue of being standardised or not, Participant 1 commented that: *'A workplace assessment is more realistic than an assessment in the clinical environment'* and by nature of the workplace assessment this is non standardised. However, Participant 5 talked of using the clients' workplace for some components of the FCE so making it very job specific and therefore combined the FCE and the Workplace assessment.

King, Tuckwell & Barrett (1998) support the use of workplace information within the FCE [12], however, Innes and Straker (2003) identified modifications to standardised

FCEs caused concern regarding the medico-legal implications, however not to the assessment process.

Policies within the workplace also impacted upon the therapists. Most participants stated it was a workplace policy not to conduct an FCE on clients they were case managing. Participant 2 stated: '*doing a FCE on someone I know that would affect my objectivity.'* Participant 1 added: 'it *may influence how you do the FCE, how you react and you may have some preconceived ideas about the assessment.'*

In contrast, Participant 5 stated: *'I prefer to do the assessment on a client known to me, as I have a better knowledge of what a person can do and therefore the risk of aggravating the injury is reduced.'*

Skills of the assessor

The skill of the therapist was discussed in relation to FCE use and this is consistent with the findings of Innes and Straker (2002) who found therapists believed that they were the assessment tool and the quality of the assessor (therapist) was related to the credibility of the assessment.

All participants agreed training in the use of FCEs was essential, however the type of training varied. Several of the participants had completed formal training and accreditation in the use of a specific FCE, however Participant 4, who had completed a formal training, commented: *'observation of others conducting the FCE was also useful as a learning tool'.* Participant 2 commented when they started work, observing another therapist complete the FCE was the training however, they went on to comment that the training procedure was now more comprehensive. On the job training was also discussed by Participant 3, as the FCE conducted was a non standardised assessment, with competency based training being employed in conjunction with a coach or experienced therapist.

Observation was discussed by all participants as an essential skill in conducting FCEs; to be able to observe how tasks were completed, the behaviours of the client, physical signs and specifically in relation to safety. In relation to this, a thorough understanding and knowledge of body mechanics and anatomy was seen as being essential (Participant 4).

An ability to gain rapport with the client was discussed by participants 1 & 4, to assist put the client at ease and understand the purpose of the FCE. Participant 4 commented:*' the client doesn't have a choice and often thinks it is just a process.'* Participant 4 went on to comment that listening to the stories of the client, the client

narrative, was also important, *'you can learn a lot from what they are telling you.'* Innes and Straker (2003) also identified that establishing communication, rapport and trust with the injured worker and employer was a strategy employed by therapists in conducting FCEs.

Experience was discussed by several participants- Participant 4 commented this was one of the most significant skills and was linked to confidence in their own abilities. Participant 5 commented that experienced therapists needed to be open minded and that a broad range of experience rather than a lot of experience in a narrow field also made a difference. Innes and Straker (2003) found therapist knowledge and experience contributed to the trustworthiness or consistency of results and was critical to establishing credibility in medico-legal settings. Strong et al (2004b) found reasons for choosing a particular FCE provider related to the experience, professionalism, knowledge and use of clinical reasoning of the provider. Strong et al (2004a) also identified that FCEs rely on the training and experience of the provider.

Participants 4 and 5 also commented upon the different skills of therapists specifically discussing the different approaches of occupational therapists and physiotherapists. This raised the issue of reliability of providers conducting the same assessment or writing the same report. Innes and Straker (2003) found within some organisations that inter-rater reliability was attempted to be achieved using training, multiple data sources, triangulation of results and consistent report formats. However, they identified there is limited research to indicate the reliability of specific assessment tools (Innes and Straker 1999a).

Limitations of the study

A limitation of this study was the small sample size, with all participants working in one regional area in NSW. Some of the participants were known to the researcher which may have implications of bias. The conclusions therefore, should be considered in light of this. The five participants provided rich data of their attitudes and practices in relation to FCE usage, and further study including a larger quantitative study would allow a more in-depth understanding of therapist's attitudes and practices and greater generalisation of the findings.

3.1.4 CONCLUSION

This study produced rich descriptive data from a small sample of therapists about their attitudes and practices in relation to FCEs.

It was found FCEs are applied differently according to the reason for referral and client goal, the therapists' workplace procedures, policies and resources and the therapist's skill and experience. There was a mix of standardised and non standardised assessments used however therapists discussed adapting the FCE, irrespective of standardisation, for specific purposes and to meet the goal of the assessment. This goal varied according to the reason for referral, client's job requirements and the client's injury type.

Personal skills and experience of the health professional was raised as an important consideration and concurs with previous research findings.

Despite some FCEs being standardised tools, and therefore requiring certain procedures to produce reliable and valid results, therapists in this study adapted the assessment to suit their requirements. This has implications for the reliability and validity of the assessment tool, however as previous research has indicated there is limited research on the validity and reliability of these tools.

From this study further options to explore include: i). building on these results with a larger survey based quantitative study to allow a more in-depth understanding of therapist's attitudes and practices and ii). further investigating the implications of these findings on the reliability and validity of current practice in relation to FCEs.

FCEs are being used widely in the areas of disability management and occupational / vocational rehabilitation. Consideration needs to be given to the practices, the reliability and validity, and of the outcomes, to ensure best practice is achieved.

ACKNOWLEDGEMENTS

We wish to acknowledge and thank the participants for kindly giving their time and sharing their valuable experiences.

REFERENCES

"NSW Workers Compensation Act (1987)." from http:www.legislation.nsw.gov.au/fullhtml/inforce/act+70+1987+FIRST+0+N?#pt.3-div.2-sec.40a.

Barry, C., Britten, N., Barber, N., Bradley, C.& Stevenson, F. (1999). "Using Reflexivity to Optimize Teamwork in Qualitative Research." Qualitative Health Research **9**(1): 26-44.

Boadella, J., Sluiter, J.& Frings-Dresen, M. (2003). "Reliability of Upper Extremity Tests Measured by the Ergos Work Simulator: A Pilot Study." Journal of Occupational Rehabilitation **13**(4): 219-232.

Bowling, A. (2004). Research Methods in Health; Investigating Health and Health Services. London, McGraw Hill.

Carpenter, C.& Hammell, K. (2000). Evaluating Qualitative Research. Using Qualitative Research: A Practical Introduction for Occupational and Physical Therapists. K. Hammell, Carpenter, C.& Dyck, I. Edinburgh, Churchill Livingstone**:** 107-119.

Deen, M., Gibson, L.& Strong, J. (2002). "A Survey of Occupational Therapy in Australian Work Practice." Work **19**(3): 219-230.

Durand, M., Loisel, P., Poitras, S., Mercier, R., Stock, S.& Lemaire, J. (2004). "The Interrater Reliability of a Functional Capacity Evaluation: The Physical Work Performance Evaluation." Journal of Occupational Rehabilitation **14**(2): 119-129.

Innes, E.& Straker, L. (1999a). "Reliability of Work-Related Assessments." Work **13**(2): 107-124.

Innes, E.& Straker, L. (1999b). "Validity of Work-Related Assessments." Work **13**(2): 125-152.

Innes, E.& Straker, L. (2002). "Workplace Assessments and Functional Capacity Evaluations: Current Practices of Therapists in Australia." Work **18**(1): 51-66.

Innes, E.& Straker, L. (2003). "Workplace Assessments and Functional Capacity Evaluations: Current Beliefs of Therapists in Australia." Work **20**(3): 225-236.

King, P., Tuckwell, N.& Barrett, T. (1998). "A Critical Review of Functional Capacity Evaluations." Physical Therapy **78**(8): 852.

Matheson, L., Rogers, L., Kaskutas, V.& Dakos, M. (2002). "Reliability and Reactivity of Three New Functional Assessment Measures." Work **18**(1): 41-50.

Morse, J.& Field, P. A. (1995). Nursing Research: The Application of Qualitative Approaches. London, Chapman Hall.

Patton, M. (2002). Qualitative Evaluation and Research Methods. Newbury Park, CA, Sage Publications.

Pransky, G.& Dempsey, P. (2004). "Practical Aspects of Functional Capactiy Evaluations." Journal of Occupational Rehabilitation **14**(3): 217-229.

Strong, S., Baptiste, S., Clarke, J.& Costa, M. (2004a). "Use of Functional Capacity Evaluations in Workplaces and the Compensation System: A Report on Workers' and Users' Perceptions." Work **23**(1): 67-77.

Strong, S., Baptiste, S., Cole, D., Clarke, J., Costa, M., Shannon, H., Reardon, R.& Sinclair, S. (2004b). "Functional Assessment of Injured Workers: A Profile of Assessor Practices." The Canadian Journal of Occupational Therapy **71**(1): (13-23).

Suto, M. (2000). Issues Related to Data Collection. Using Qualitative Research: A Practical Introduction for Occupational and Physical Therapists. K. Hammell, Carpenter, C.&Dyck, I. Edinburgh, :Churchill Livingstone: 25-46.

APPENDIX A

Interview Schedule

(Introduction: thanks for agreeing to participate; ask permission to tape record and explain right to stop recording at any time, erase part or all on request; go over consent statement; any questions).

What is your background to conducting FCEs?

Discipline, time working in OH, reasons for conducting FCEs.

What type of FCEs do you conduct? And can you explain why you chose to use this particular FCE.

Key, West, Ergos, Blankenship, WorkHab, Isernhagen, non standardised etc.

Client type, referral reason etc.

Social factors/ work factors.

What do you feel about conducting this particular type of FCE?

Referral reasons - FCE - job / no job etc.

Opinion of others to FCE.

Do you complete the FCE in its entirety or do you complete selected sections only and can you explain this?

Type of injury, type of job, referral reason, information for RTW.

What influences the recommendations you make at the conclusion of the FCE?

Reason for referral, expected outcomes - job/ no job.

Clinical reasoning, narratives, client history, legal status, relationship with client.

What are your thoughts about the FCE and predicting outcomes?

+/- conclusive, starting point, predicting RTW/ retraining.

Do you have any other experiences of this topic that you would like to share?

3.2 Manuscript: Health Professionals' Perceptions and Practices in relation to Functional Capacity Evaluations – Results of a Quantitative Survey

Citation: James, C. and Mackenzie, L. (2009). "Health Professional's Perceptions and Practices in Relation to Functional Capacity Evaluations: Results of a Quantitative Survey." Journal of Occupational Rehabilitation 19(2):203-211.

ABSTRACT

Introduction

This study investigated the perceptions and practices of Australian health professionals in relation to the use of Functional Capacity Evaluations (FCEs)

Methods

A quantitative cross- sectional study design was used to survey health professionals who conduct FCEs and who were working for one of 219 rehabilitation providers in NSW, Australia. Seventy seven returned surveys were eligible for inclusion.

Results:

11 different FCEs were being utilised with many health professionals using more than one FCE. The most commonly used FCE was non-standardised (56%, n=43) followed by 52% (n=40) using the WorkHab, and 18% (n=14) using Valpar.

Both non-standardised and standardised assessments were being used by 90% (n=69) of respondents.

Health professionals reported using all or parts of the FCE, and indicated identical FCEs are not always conducted, with adaptation of the FCE, due to client injury (82%, n=62) and job (80%, n=43) occurring. 60% of respondents had no choice in the type of FCE they conducted, and of the 40% with a choice, this was not influenced by other stakeholders in the process. Accreditation and training, characteristics of assessment tasks, standardisation, reliability, cost, length and flexibility were all identified as factors affecting the selection of an FCE.

Conclusions:

This study demonstrated that health professionals in NSW Australia are not routinely using standardised tools for FCEs. Health professional perceptions suggest accreditation, training and the characteristics of the FCE were important factors in FCE

selection. In practice, participants tended to use parts of an FCE rather than the whole FCE. Adaptation of FCEs was common, due to client injury and specific job requirements.

KEYWORDS: Functional Capacity Evaluation, Occupational Rehabilitation.

3.2.1 INTRODUCTION

Within the practice areas of workplace injury management, occupational rehabilitation and work injury prevention, functional capacity evaluations (FCEs) are used to define an individual's functional abilities or limitations in the context of safe, productive work tasks (King et al. 1998). Functional capacity evaluations are commonly used with individuals who have suffered work related musculo-skeletal injuries as part of the rehabilitation process to remain or return to work. FCEs are designed to measure function, and therefore they can be used for a range of injury and disease types, both work and non work related. FCEs are used for work fitness determinations and to facilitate return to work (Gross et al. 2007) and have been identified as providing complementary information for insurance physicians assessing physical work ability (Wind et al. 2008). FCEs are also used within the medico-legal arena (Reneman et al. 2006).

There are a variety of FCEs available commercially and many clinics have developed their own non- standardised, and work specific FCEs. All FCEs attempt to measure functional performance objectively, to produce reliable and valid results, however, studies to establish if this is achieved are limited (King et al. 1998; Innes 2006). Research in these areas, specifically on the reliability and validity of some of the commercially available assessments, is now beginning to be published (Jay et al. 2000; Ting et al. 2001; Matheson et al. 2002; Boadella et al. 2003; Durand et al. 2004; Gouttebarge et al. 2004; Reneman et al. 2004; Reesink et al. 2007; Durand et al. 2008). A Standardised FCE can be defined as one that is commercially available, has acceptable measurement/ psychometric properties and is conducted using standardised procedures and protocols (Innes and Straker 2003b; Reneman et al. 2006). A non-standardised FCE can be defined as a self designed, internally adaptable assessment tool which may be specifically designed for a particular industry. Commercially available FCEs have the common goal of identifying functional abilities however different tools use slightly different models and approaches. The main models in use are biomechanical, physiological, psychophysical and kinesiophysical (Gibson and Strong 2002). Many FCEs are based on the physical domains outlined in the Dictionary of Occupational titles (King et al. 1998) and comprise a range of tasks that include these domains. Some tools include job simulation assessment; some include assessment of effort, symptom exaggeration, fear avoidance and other psychological issues as part of the FCE to determine safe lifting or physical demand capabilities.

Strong et al (2004) identified limitations of current research and difficulties of extrapolating information from a single point in time (Strong et al. 2004b). In a

Canadian study, FCEs were conducted with variations in guidelines or practice standards, with limited contextual information, and practices were influenced by referral source and market demands (Strong et al. 2004a).

Each country has its own process and procedures in relation to occupational rehabilitation (defined as: a managed process that involves appropriate, adequate and timely services based on assessed needs, aiming to maintain an injured worker in, or return them to, suitable employment (WorkCover(Victoria) 2006), and as such differences in what and how assessments are being used is to be expected. As occupational rehabilitation falls under the workers compensation umbrella in many jurisdictions, legislation may influence requirements of this process, which will therefore vary across states and countries. FCEs are however, used for common purposes across the globe and commercially available standardised assessments are used across many different countries (Reneman et al. 2006; WorkHab International 2008). There have been differences reported in the results of standardised FCEs between countries perhaps being the result of differences in patients or context (Reneman et al. 2006), however, the practice issues facing the health professional in relation to FCEs are similar (Innes and Straker 2002; Pransky and Dempsey 2004; Strong et al. 2004a).

Therapists have identified flexibility of the assessment tool as a consideration when using an FCE, and this has an impact upon standardisation of a tool and it's reliability and validity (Cotton et al. 2006; James et al. 2007). Several authors (King et al. 1998; Innes and Straker 2003a) discuss the issue of generic versus specialist (or job specific) FCE testing, and the ability to alter the test depending upon the needs of the client, work and situation. Factors that have been identified as having an impact upon this adaptation include the reason for referral; client's job requirements; client's injury type; the therapists' workplace procedures, policies and resources; and the therapist's skill and experience (Innes and Straker 2002; Strong et al. 2004a; Strong et al. 2004b; James et al. 2007).

When looking at what specific assessments are being used within the occupational rehabilitation arena, Jundt and King (1999) completed a study of work rehabilitation programs in the USA, and found 94% of respondents conducted job analysis/risk hazard analysis and 91% conducted functional capacity evaluations. Of the FCEs being used, 45% used the Isernhagen work system, 12% the Ergos and 12% the Key system (Jundt and King 1999). Deen et al (2002) looked specifically at occupational therapists (OT's), working in occupational rehabilitation settings in Australia and found 96% conducted workplace assessments, 86% functional capacity evaluations and 75% job/risk assessments. Of those conducting FCEs, specific tools being used were,

WorkHab- 36%, Valpar – 23% and West – 18% (Deen et al. 2002). Cotton et al (2006) found 75% of rehabilitation providers in NSW conducted an “own design” FCE rather than a commercially available assessment tool and of those using commercial tools, WorkHab was the most commonly used (17%), followed by Isernhagen (10%) (Cotton et al. 2006).

Despite previous studies investigating FCEs and what assessments are being used, there is less data available about how FCEs are applied in practice.

This research aimed to build on existing research to identify the perceptions and describe the practice of Australian health professionals in relation to the use of FCEs.

3.2.2 METHOD

Following ethics approval, a cross sectional study design was used to survey health professionals, who conduct FCEs, and who were working for rehabilitation providers in NSW. Only those health professionals who conduct FCEs were included in the study.

Survey

Items within the survey were developed from a qualitative study (James et al. 2007) which investigated the perceptions and practices of health professionals in relation to FCE usage, and using literature to identify assessment tools currently in use. The survey was piloted with twelve health professionals. The pilot survey was modified to improve the internal reliability of the survey subscales (Higginbotham et al. 2001), and sent to participants in the mail.

The survey included 60 questions within six sections: Demographic data of the participants; Type of FCE used; FCE choice; FCE usage; Perceived consequences of using an FCE and Perceptions of FCEs (see Appendix 1). The survey included questions seeking responses both of a categorical nature (related to the background of the health professional and the type of FCE used). It included items about standardised FCEs, defined as those that are commercially available, have acceptable measurement/ psychometric properties and are conducted using standardised procedures and protocols (Innes and Straker 2003b; Reneman et al. 2006); and non standardised FCEs, defined as self designed, internally adaptable assessment tools. Ordinal data relating to choice, usage, perceived consequences and perceptions of FCEs was also collected. These questions used direct estimation using likert scales. Raters expressed an opinion based on agreement with a series of statements about the use of FCEs (Portney and Watkins 2000). Completed surveys were returned to the researcher by mail.

Participants

Health professionals who use FCEs and were working for WorkCover (NSW) accredited rehabilitation providers were invited to participate. Under the WorkCover (NSW) guidelines, Occupational therapists and Physiotherapists are approved to conduct FCEs. Other health professionals such as Exercise Physiologists and Occupational Health Nurses need WorkCover (NSW) approval to conduct these assessments under this regulatory system. In NSW, Australia, this consisted of 219 WorkCover (NSW) accredited rehabilitation providers, who may employ several health professionals to conduct FCEs.

Procedure

The survey was sent to WorkCover (NSW) accredited rehabilitation providers. Accredited rehabilitation providers are organisations accredited by Workcover (NSW) to offer specialist services to help injured workers to return to work. Rehabilitation providers employ different health professionals to assess the needs of the injured worker and the workplace requirements, to then develop a rehabilitation plan to assist the injured worker to return to work (WorkCover (N.S.W.) 2002). The Managers of each rehabilitation provider were asked to distribute information about participating in the study to any health professionals who conducted FCEs within their organisation. Completed surveys were returned anonymously to the researcher in the mail.

Data analysis

Survey responses as either numerical, scaled items, categorical or nominal data, was entered into STATA [v8.0] (Statacorp 2003). Data was checked after entering and missing variables were checked for accuracy prior to analysis. Descriptive analysis including mean values, confidence intervals and estimates of proportions were calculated.

3.2.3 RESULTS

Participants:

Eligible surveys were received from 77 participants working for 65 different rehabilitation providers, this being a response rate of 30% from the 219 rehabilitation providers invited to participate. Of those who replied, 82% (n=63) were occupational therapists, 13% (n=10) physiotherapists and 5% (n=4) exercise physiologists. This response is generally representative of the ratio of the different professionals working for rehabilitation providers in NSW.

The mean number of years of professional experience was 10.9 years (range:-22yrs) and the mean number of years of FCE experience was 5.3 years (range: 1-16yrs).

Table I provides a description of the sample and shows the number of professionals involved in the study, years of experience by profession and years of FCE experience by profession.

Table I: Descriptive information on sample (n=77)

Professionals	Number		Mean Yrs of experience		Mean FCE Yrs experience	
Occupational Therapists	82%	(n=63)	10.4	(SD:7.2)	5.6	(SD:4)
Physiotherapists	13%	(n=10)	16.6	(SD:6.7)	5.1	(SD:2.9)
Exercise Physiologists	5%	(n= 4)	5.3	(SD:2.7)	2.0	(SD:0)

Practice issues:

FCEs in use in NSW.

The results indicated 10 standardised FCEs were being used throughout NSW. Additionally health professionals described using non-standardised FCEs in practice. Health professionals reported all the assessments they were currently using, and therefore could indicate more than one assessment. The most frequently used FCE was a non-standardised assessment (56% n=43). Both non-standardised and standardised FCEs were being used by 46% (n=35) of health professionals and 91% (n=69) were using standardised FCEs. Of the standardised assessments being used the most frequently used was the WorkHab (53% n=40), followed by the Valpar assessment (18% n=14).

Non standardised assessments only, were used by 10.5% (n=8) of health professionals. 44% (n=34) used only a standardised assessment, and the remaining health professionals reported using both a non-standardised and standardised assessment – 46% (n=35). 26% of all respondents (n=20) were using more than one type of standardised assessment. Of these, 13%(n=10) were using two different standardised assessments, 12% (n=9) were using three different standardised assessments and one (1.3%) health professional reported using five different standardised assessments. Table II outlines FCE usage by profession.

Table II: Usage by FCE type and Profession (n=77).

FCE	No's OT's (N=63)		No's PT's (N=10)		No's ExPh (N=4)		Total Number (N=77)	
Non Standardised	**33**	**(52%)**	**6**	**(60%)**	**4**	**(100%)**	**43**	**(56%)**
Non Standardised +Standardised	**29**	**(46%)**	**4**	**(40%)**	**2**	**(50%)**	**35**	**(46%)**
Standardised	**59**	**(94%)**	**8**	**(80%)**	**2**	**(50%)**	**69**	**(91%)**
Standardised FCE by type:								
WorkHab	**35**	**(56%)**	**4**	**(40%)**	**1**	**(25%)**	**40**	**(53%)**
Valpar	**14**	**(22.5%)**	**0**		**0**		**14**	**(18%)**
Isernhagen	**10**	**(16%)**	**3**	**(30%)**	**0**		**13**	**(17%)**
Pile	**8**	**(12.9%)**	**0**		**1**	**(25%)**	**9**	**(12%)**
West	**9**	**(14.5%)**	**0**		**0**		**9**	**(12%)**
Keys	**6**	**(9.7%)**	**1**	**(10%)**	**0**		**7**	**(9%)**
Blankenship	**2**	**(3.2%)**	**2**	**(20%)**	**0**		**4**	**(5%)**
PWPE	**4**	**(6.5%)**	**0**		**0**		**4**	**(5%)**
Ergos	**0**		**1**	**(10%)**	**0**		**1**	**(1.3%)**
Workability	**1**	**(1.6%)**	**0**		**0**		**1**	**(1.3%)**

* Health professionals could select more than one.

The seven most commonly used FCEs in NSW were: Non-standardised assessments, WorkHab, Valpar, Isernhagen, West, Pile and Keys in descending order of popularity.

Patterns of FCE Utilisation

Health professionals identified that at times they used all of the components of the assessment and at times they only used parts of the assessment – for both non-standardised and standardised assessments. The proportions of health professionals using part or all of the components of the seven most popular assessments is outlined in Figure I and 2. Health professionals could indicate pattern of use for more than one FCE.

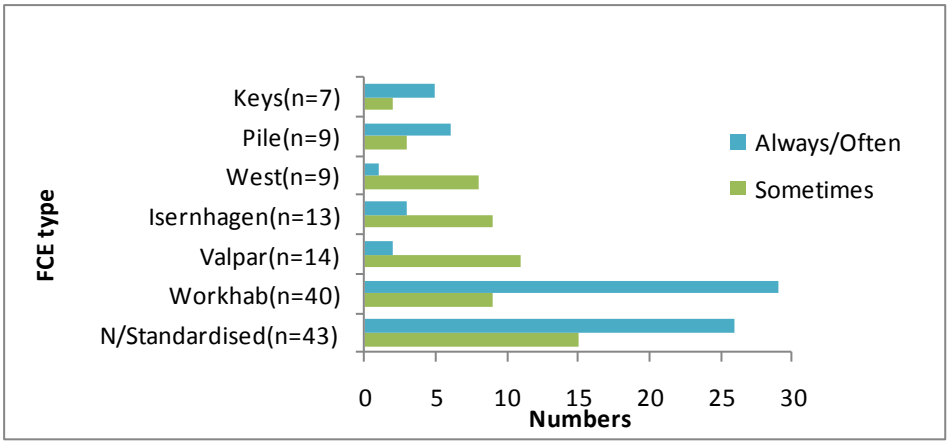


Figure 1: Utilisation of All Components of Standardised FCE (n=71*)

Health professionals could report on more than one FCE.

*6 missing variables.

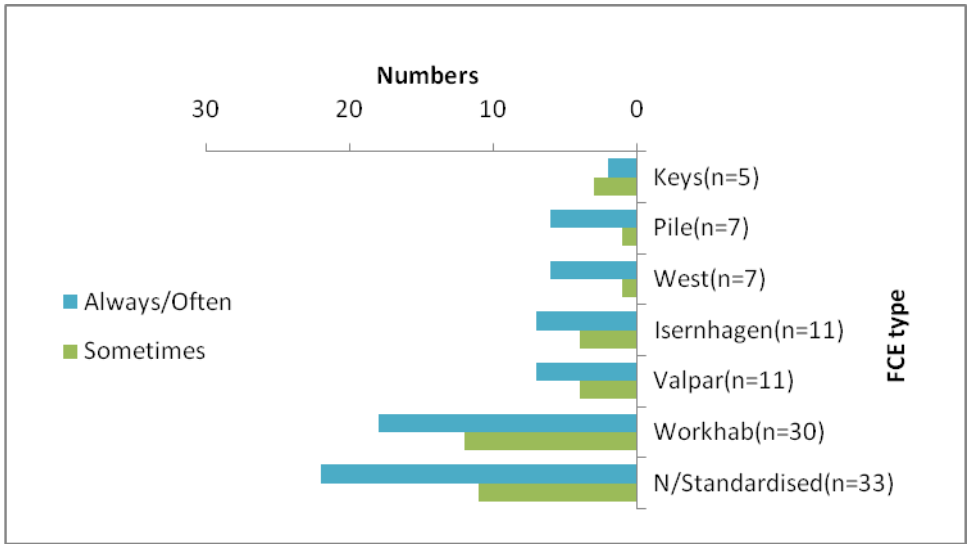


Figure 2: Utilisation of PARTS of FCE (n=60)

Health professionals could report on more than one FCE.

Forty percent (40% n=31) of health professionals indicated they were unlikely to use identical FCEs with all clients, whereas 36% (n=28) indicated they would. Of those respondents (n=43) who used a non-standardised assessment 34% (n=15) indicated they were unlikely to use identical FCEs and the same amount (34% n=15) would. Of those health professionals (n=69) using standardised assessments, 40% (n=28) indicated they were unlikely to use identical FCEs and 36% (n=25) indicated they would use identical assessments. Health professionals indicated they adapted the FCE conducted in response to the clients' injury (82% n=63), and 80% (n=43) indicated they would adapt the FCE in response to the clients' job.

Frequency of use

Figure 3 shows the frequency of use of the seven most popular standardised assessments identified. 56% (n=43) of health professionals indicated that they used a non-standardised assessment. Of these 68% (n=28) used this less than twice per month and 31% (n=13) indicated they used this more than three times a month.

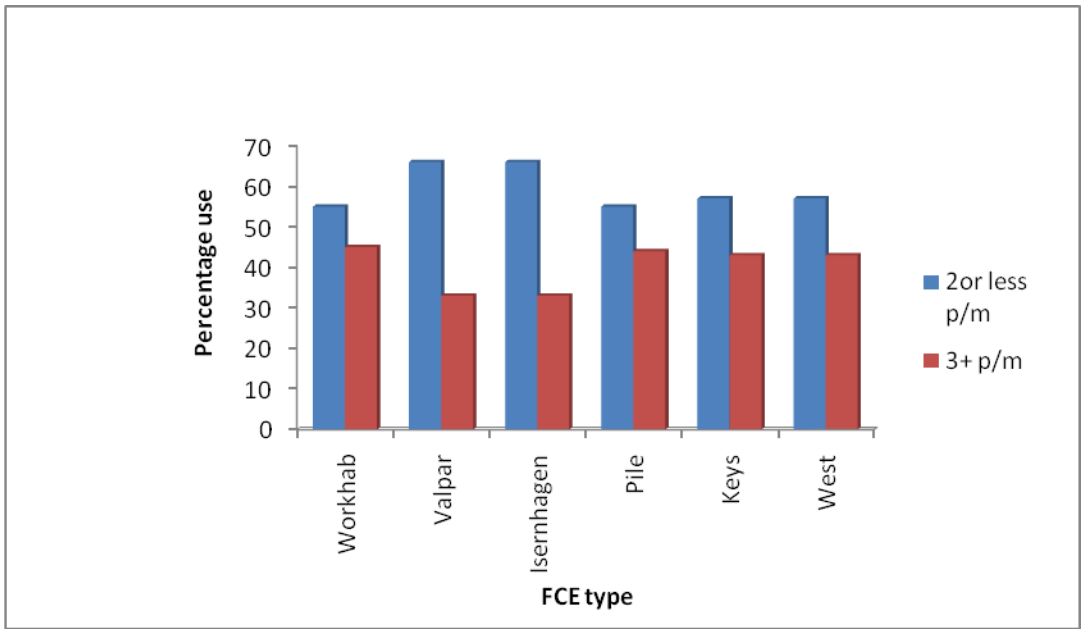


Figure 3: Frequency of use of most popular Standardised FCEs

Health professionals could report on more than one FCE.

* 1 missing variable.

FCEs and Workplace assessments

Out of the total participant group, 33% (n=25) of health professionals indicated for each client, that they frequently conducted an FCE (either standardised or non-standardised) along with a workplace assessment. A further 47% (n=36) of health professionals indicated they often conducted Workplace assessments instead of FCEs for individual clients.

Perceptions in regard to FCEs

Factors affecting selection of FCE

Regardless of preferences in use of FCEs in NSW, Australia, the importance of different criteria in the selection of an FCE identified accreditation and training to conduct the assessment and the characteristics of the assessment tasks as highly important, and standardisation, reliability, cost, length and effort as moderately important. This is outlined in Table III.

Table III: Mean score of importance of different qualities/ criteria in choice of FCE (n=77)

Criteria/ Quality	Mean Score (0-5)	
Availability	**3.9**	**(SD:1.3)**
Flexibility	**3.8**	**(SD:0.8)**
Standardisation	**3.6**	**(SD:1.1)**
Reliability/Validity	**3.6**	**(SD:1.0)**
Task Characteristics	**4.0**	**(SD:0.7)**
Accreditation/training	**4.2**	**(SD:0.8)**
Length	**3.4**	**(SD:0.9)**
Cost	**3.5**	**(SD:0.9)**
Effort	**3.0**	**(SD:1.1)**

40% (n=31) of Health Professionals indicated they had a choice in the type of FCE that was conducted. 60% (n=46) had no choice, suggesting that only one assessment was available to them.

Of those health professionals with a choice of FCE type, 64% indicated they were the only person choosing the FCE. Choice of FCE used was not influenced by the: referrer, client, doctor, the health professionals' employer or the clients' employer for the majority of health professionals. Figure 4 shows the frequency of different influences on the choice of FCE used, for those health professionals that did have a choice.

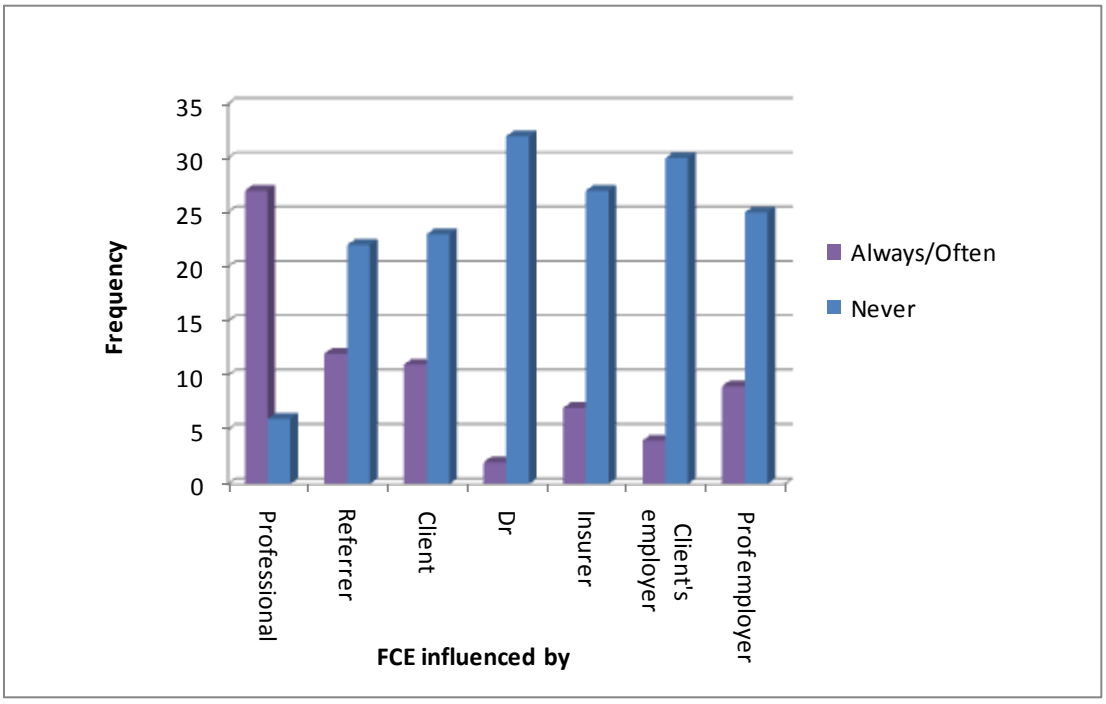


Figure 4: Influences for choices of FCE

3.2.4 DISCUSSION

Practice issues

FCEs in use in NSW

Despite the wide range of FCEs available, most health professionals continue to use a non-standardised FCE, as identified in previous research (Cotton et al. 2006). The use of non-standardised assessments raises questions in regards to what and how aspects of function are being assessed and the consistency of these assessments. The expectation that health professionals are using evidence to inform practice, can therefore be questioned as it raises the issue that evidence relating to the properties of many FCEs, especially non-standardised tools, is lacking. Amongst those that used a standardised FCE, most used the WorkHab (52%) and, as was identified by Cotton et al, the two most commonly used are the least expensive to purchase (Cotton et al. 2006). WorkHab was also identified as the most commonly used FCE in research by Deen (Deen et al. 2002). WorkHab is an Australian developed and produced tool which may impact upon its popularity. The other commercially available tools identified in this study have all been developed internationally, and are identified as less commonly used in Australian studies, including this one (Deen et al. 2002; Cotton et al. 2006).

Many health professionals had a choice in the FCE they used, with 44% indicating they used both a non-standardised and a standardised FCE however when they chose to use one over the other was not identified. Use of more than one type of standardised tool was also found in 26% of health professionals however the reasons they chose a specific standardised assessment for any particular client was not clarified. Of the health professionals who had a choice in the FCE used, 64% indicated they were the only person choosing the FCE, suggesting health professionals with a choice had the most control over which FCE they used. Having the ability to choose a specific FCE also allows increased professional judgement on the part of the health professional. Clinical reasoning skills to determine the most appropriate FCE tool, based on the model of FCE, the client injury, client job and purpose of the assessment may impact upon this choice, however it may also be the result of personal preferences and perceptions relating to FCEs as discussed later.

Availability at the workplace was rated highly by health professionals in this study, which concurs with the study by Strong et al (Strong et al. 2004b). Availability is also a factor in relation to choice of FCE type. In this study, 60% of health professionals did not have a choice in the type of FCE that was conducted, this links with previous research conducted in NSW Australia, where health professionals were found to use

only one FCE tool, however it was not indicated in that research whether the health professionals had a choice and chose to only use one FCE or whether these health professionals did not have a choice (Cotton et al. 2006).

Choice may be limited by the availability of FCE tools with a particular rehabilitation provider and economic factors may influence this, as a result of the cost of purchase of the assessment tool, and additional costs of training and accreditation for the health professional to use the tool.

Patterns of FCE utilisation

Health professionals identified that they were using parts of, rather than the whole assessment in many instances. Depending upon the actual assessment tool, using part of the assessment can impact upon the standardisation of the tool and therefore may impact upon the reliability and validity of the tool.

Of the commercially available standardised FCEs the health professionals using the Keys assessment were the most likely to use all components of the tool. It is difficult to know with a non-standardised FCE what is the whole or a part of the assessment, however the responses from participants indicated that they used parts of the tool, suggesting the assessment comprises a variety of components.

Health professionals adapted FCEs to suit specific requirements, especially in relation to the clients' injury type or job. This concurs with other authors (King et al. 1998; Innes and Straker 2003a) who identified the issue of generic versus specialist (or job specific) FCE testing, and the ability to alter the test depending upon the needs of the client, work and situation. The use of, or lack of use of identical FCEs was identified in this study, with adaptation of an assessment for a specific client commonly being reported. This has been identified in previous research (King et al. 1998; Innes and Straker 2003a) and also relates to the issue of the clinical utility of the assessment tool (Toomey et al. 1995).

For those health professionals who said they would conduct identical assessments, this could suggest a lack of alternatives or a more standardised approach. This has implications in relation to standardised protocols being used with FCEs regardless of client presentation. Similar percentages of users (using both non-standardised and standardised FCEs) indicated they conducted identical FCEs, which does not suggest using a standardised FCE results in more standardised protocols and procedures being used, as might be expected.

Frequency of Use and Workplace Assessments

FCEs were reported to be used several times a month however, health professionals also reported conducting a workplace assessment in place of or in conjunction with an FCE. Under the Workcover (NSW) occupational rehabilitation guidelines a workplace assessment is required for those clients returning to the workplace (WorkCover (N.S.W.) 2002) and it would therefore be expected most clients would undergo a workplace assessment, whereas a functional assessment is not mandatory and therefore only completed on some clients. This may be associated with the type of client involved in the rehabilitation process, and whether they have a job to return to or not. Previous research has suggested a workplace assessment is desirable to measure ability for a specific job (Pransky and Dempsey 2004), and identified three types of work related assessment – workplace assessments that are specific to the work place and work tasks, functional capacity evaluations for those with a job and thirdly for those without a job (Innes and Straker 2002). Further research could investigate the relationship between those injured workers who undergo a FCE and a workplace assessment or only a work related assessment to determine if there is a link between an FCE and those clients who have a job to return to or not.

Perceptions in relation to FCEs

Health Professionals in this study rated accreditation and training to conduct an FCE highly. This suggests that FCEs are perceived as a specialised area of practice that requires the health professional to undertake specific training. This is the case with some standardised FCEs that require health professionals to: undergo training; obtain accreditation; and in some instances have ongoing requirements to maintain accreditation. All health professionals responded that accreditation/ training was important; for those health professionals who use a non-standardised assessment tool, for which there is no accreditation, it can be presumed that training was considered important and is required. WorkCover (NSW) recognises minimum qualifications of professionals to conduct any FCEs (standardised or non standardised). For occupational therapists or physiotherapists, this is a minimum of 3 months occupational rehabilitation experience. Other health professionals are required to undergo an assessment process, to be eligible to conduct FCEs under the WorkCover (NSW) system (WorkCover (N.S.W.) 2002).

Health professionals in this study did not identify reliability and validity of assessment tools as an area of concern nor did they identify standardisation as particularly important, which could be directly related to the fact that many health professionals were using non-standardised FCEs. Research based appropriate assessment tools are

generally considered essential to ensure credible practice, however as Clemson and Fitzgerald (Clemson and Fitzgerald 1998) found issues of reliability and validity were not clearly understood by therapists which may account for this not being rated highly in this study. It is interesting to note that specific FCEs are not being requested by the referring agencies. This suggests the issues of the measurement properties including the reliability and validity of an assessment tool is not an important consideration for these referrers nor for the health professionals in their own choice of an assessment tool.

It is important to note, that much FCE research, particularly about the reliability and validity of specific assessment tools, is only just beginning to become available for some of the commercially available assessment tools. Therefore health professionals need to be reviewing and updating their knowledge to assist in providing the best possible evidence based care for clients. With a large percentage of health professionals using non- standardised assessments, the issues relating to the measurement properties of the tools are of more concern. Further research related to specific FCEs is needed, to provide evidence of the measurement properties and reliability and validity in relation to their usage. This is particularly important in an age when health professionals are being encouraged to justify treatments and services with evidence, both by the health professions as a whole and by those paying for services (Muir Gray 1997).

Task characteristics and flexibility within the FCE, were rated as important criteria by health professionals, in the choice of the FCE and these relate to the adaptation of the FCE. Flexibility within the assessment can threaten the standardised application of the assessment tool, which may also account for standardisation not being rated highly in this study.

Cost of the assessment, length and effort to conduct the FCE were all rated as moderately important in this research. Within the WorkCover (NSW) occupational rehabilitation system, (Strong et al. 2004b) FCEs are charged to insurers and employers based on the time taken to complete, and costs are discussed and approved by the referrer prior to the service being provided in this environment.

Effort required to conduct the assessment was rated least important when looking at the different qualities of the FCE, suggesting the health professional is motivated to put in more effort if needs are being met. These needs could be related to those of the client and his or her rehabilitation, or related to the needs of the health professional in fulfilling their work requirements and managing their case load.

The issues of time, effort, length, flexibility and task characteristic will also inform the health professional in regard to the choice of FCE used (where choice is available). Further research could investigate the reasons health professionals choose one FCE over another and investigate the issue of effort, motivation and the health professional in relation to the occupational rehabilitation process.

3.2.5 CONCLUSIONS

This study identified that despite different FCE assessment tools being used in NSW Australia, a non-standardised tool was preferred by most participants. Many health professionals adapted FCEs by only using parts of them, or changing items in response to the client injury and job. Standardisation, reliability and validity were identified as important criteria when participants selected FCEs. However, task characteristics and flexibility within the FCE were also identified as important criteria; leading to increased adaptation of FCEs to suit individual clients, thus impacting on the standardisation, reliability and validity of the FCEs.

Despite the limitations of this study being a small sample and the sample being from one state within Australia, the results provide a profile of usual practice in relation to FCE usage within this environment.

Further research to investigate what components of an FCE are being included or not, or under what circumstances this occurs in practice is recommended.

ACKNOWLEDGEMENTS

The authors would like to thank all health professionals who gave their time to participate in this study.

REFERENCES

Boadella, J., Sluiter, J.& Frings-Dresen, M. (2003). "Reliability of Upper Extremity Tests Measured by the Ergos Work Simulator: A Pilot Study." Journal of Occupational Rehabilitation **13**(4): 219-232.

Clemson, L.& Fitzgerald, M. (1998). "Understanding Assessment Concepts within the Occupational Therapy Context." Occupational Therapy International **5**(1): 18-34.

Cotton, A., Schonstein, E.& Adams, R. (2006). "Use of Functional Capacity Evaluations by Rehabilitation Providers in NSW." Work **26**(3): 287-295.

Deen, M., Gibson, L.& Strong, J. (2002). "A Survey of Occupational Therapy in Australian Work Practice." Work **19**(3): 219-230.

Durand, M., Brassard, B., Hong, Q., Lemaire, J.& Loisel, P. (2008). "Responsiveness of the Physical Work Performance Evaluation, a Functional Capacity Evaluation, in Patients with Low Back Pain." Journal of Occupational Rehabilitation **18**(1): 58-67.

Durand, M., Loisel, P., Poitras, S., Mercier, R., Stock, S.& Lemaire, J. (2004). "The Interrater Reliability of a Functional Capacity Evaluation: The Physical Work Performance Evaluation." Journal of Occupational Rehabilitation **14**(2): 119-129.

Gibson, L.& Strong, J. (2002). "Expert Review of an Approach to Functional Capacity Evaluation." Work **19**(3): 231-242.

Gouttebarge, V., Wind, H., Kuijer, P.& Frings-Dresen, M. (2004). "Reliability and Validity of Functional Capacity Evalutation Methods: A Systematic Review with Reference to Blankenship, Ergos Work Simulator, Ergo-Kit and Isernhagen Work System." International Archives of Occupational Environmental Health **77**(8): 527-537.

Gross, D., Battie, M.& Asante, A. (2007). "Evaluation of a Short Form Functional Capacity Evaluation: Less Maybe Best." Journal of Occupational Rehabilitation **17**(3): 422-435.

Higginbotham, N., Albrecht, G.& Connor, L. (2001). Health Social Science: A Transdisciplinary and Complexity Perspective. Melbourne, Oxford Press.

Innes, E. (2006). "Reliability and Validity of Functional Capacity Evaluations: An Update." International Journal of Disability Management Research **1**(1): 135-148.

Innes, E.& Straker, L. (2002). "Workplace Assessments and Functional Capacity Evaluations: Current Practices of Therapists in Australia." Work **18**(1): 51-66.

Innes, E.& Straker, L. (2003a). "Attributes of Excellence in Work Related Assessments." Work **20**(1): 63-76.

Innes, E.& Straker, L. (2003b). "Workplace Assessments and Functional Capacity Evaluations: Current Beliefs of Therapists in Australia." Work **20**(3): 225-236.

James, C., MacKenzie, L.& Higginbotham, N. (2007). "Health Professionals' Attitudes and Practices in Relation to Functional Capacity Evaluations." Work **29**(2): 81-88.

Jay, M., Lamb, J., Watson, R., Young, I., Fearon, F., Alday, J.& Tindall, A. (2000). "Sensitivity and Specificity of the Indicators of Sincere Effort of the EPIC Lift Capacity Test on a Previously Injured Population." Spine **25**(11): 1405-12.

Jundt, J.& King, P. (1999). "Work Rehabilitation Programs: A 1997 Survey." Work **12**(2): 139-144.

King, P., Tuckwell, N.& Barrett, T.(1998). "A Critical Review of Functional Capacity Evaluations." Physical Therapy **78**(8): 852.

Matheson, L., Rogers, L., Kaskutas, V.& Dakos, M. (2002). "Reliability and Reactivity of Three New Functional Assessment Measures." Work **18**(1): 41-50.

Muir Gray, J. (1997). Evidence-Based Healthcare: How to Make Health Policy and Management Decisions. London, Churchill Livingstone.

Portney, L.& Watkins, M. (2000). Foundations of Clinical Research: Applications to Practice. New Jersey, Prentice Hall Health.

Pransky, G.& Dempsey, P. (2004). "Practical Aspects of Functional Capacity Evaluations." Journal of Occupational Rehabilitation **14**(3): 217-229.

Reesink, D., Jorritsma, W.& Reneman, M. (2007). "Basis for a Functional Capacity Evaluation Methodology for Patients with Work-Related Neck Disorders." Journal of Occupational Rehabilitation **17**(3): 436-449.

Reneman, M., Brouwer, S., Meinema, A., Dijkstra, P., Geertzen, J.& Groothoff, J. (2004). "Test-Retest Reliability of the Isernhagen Work Systems Functional Capacity Evaluation in Healthy Adults." Journal of Occupational Rehabilitation **14**(4): 295-305.

Reneman, M., Kool, J., Oesch, P., Geertzen, J., Battie, M.& Gross, D. (2006). "Material Handling Performance of Patients with Chronic Low Back Pain During Functional Capacity Evaluation: A Comparison between Three Countries." Disability and Rehabilitation **28**(18): 1143-1149.

Statacorp (2003). Statastatistical Software: Release 8.0. College Station, TX, Stata Corporation.

Strong, S., Baptiste, S., Clarke, J.& Costa, M. (2004a). "Use of Functional Capacity Evaluations in Workplaces and the Compensation System: A Report on Workers' and Users' Perceptions." Work **23**(1): 67-77.

Strong, S., Baptiste, S., Cole, D., Clarke, J., Costa, M., Shannon, H., Reardon, R.& Sinclair, S. (2004b). "Functional Assessment of Injured Workers: A Profile of Assessor Practices." The Canadian Journal of Occupational Therapy **71**(1): (13-23).

Ting, W., Wessel, J., Brintnell, S., Maikala, R.& Bhambhani, Y. (2001). "Validity of the Baltimore Therapeutic Equipment Work Simulator in the Measurement of Lifting Endurance in Healthy Men." American Journal of Occupational Therapy **55**(2): 184-90.

Toomey, M., Nicolson, D.& Carswell, A. (1995). "The Clinical Utility of the Canadian Occupational Performance Measure." Canadian Journal of Occupational Therapy **62**(5): 242-249.

Wind, H., Gouttegbarge, V., Kuijer, P., Sluiter, J.& Frings-Dresen, M. (2008). "Complementary Value of Functional Capacity Evaluation for Physicians in Assessing the Physical Work Ability of Workers with Musculoskeletal Disorders." International Archives of Occupational and Environmental Health **82** (4): 435-443.

WorkCover (NSW). (2002). Rehabilitation Providers. Standards and Conditions of Accreditation. WorkCover. Sydney, WorkCover NSW.

WorkCover(Victoria) (2006). Occupational Rehabilitation Service Standards, WorkCover Victoria.

WorkHab International. (2008). "About Workhab FCE Systems." Retrieved 0202, 2009, from http://www.workhab.com/.

3.3 Manuscript: The Clinical Utility of Functional Capacity Evaluations: the Opinion of Health Professionals working within Occupational Rehabilitation

Citation: James, C. and Mackenzie, L. (2009)."The Clinical Utility of Functional Capacity Evaluations: the opinion of Health Professionals working within Occupational Rehabilitation." Work 33(3):231-239

ABSTRACT:

Functional Capacity Evaluations (FCEs) are used within the occupational rehabilitation arena with the aim of assessing an individual's functional abilities in relation to work tasks. Therapists' use a variety of different FCEs, both standardised and non standardised. This study aimed to investigate therapists' views on the clinical utility of FCEs in general and to identify if these differed between professional groups.

A cross sectional study design was used. Health professionals who conduct FCEs and who worked for WorkCover accredited rehabilitation providers in NSW were surveyed.

Surveys were returned from 79 participants working for 65 different rehabilitation providers. Of those who replied, 82 % (n=63) were occupational therapists, 13% (n=10) physiotherapists and 5% (n=5) exercise physiologists. The mean years of professional experience was 10.9 years and the mean years of FCE experience was 5.3 years.

Data were analyzed using STATA [v8.0] and the clinical utility of FCEs was considered relating to: usefulness and relevance; adaptability and flexibility; therapist perceived requirements and issues in practice. No differences were found related to the clinical utility of FCEs between professional groups or years of professional experience. The results suggest consistency and similarities in how FCEs are currently used in practice across NSW (Australia). Limitations of this study and areas for further research are suggested.

KEYWORDS: Functional Capacity Evaluation, Clinical Utility, Occupational Rehabilitation.

3.3.1 INTRODUCTION AND LITERATURE REVIEW

Functional Capacity Evaluations (FCEs) are an integral part of work injury prevention and occupational rehabilitation. FCEs are designed to define the functional abilities and limitations of an individual in the context of safe, productive work tasks (King et al. 1998). The overall aim of conducting assessments, including FCEs, in occupational rehabilitation is to ensure that the worker's capacity and abilities match those of the job, and the work environment. FCEs are commonly used with individuals who have suffered work related injuries, particularly musculo-skeletal injuries.

Therapists working in occupational rehabilitation use a variety of FCEs. Some are standardised and commercially available. Others have been developed by specific clinics or therapists related to local needs and are therefore non-standardised. These FCEs are often specific to certain work environments or situations. All FCEs attempt to measure functional performance objectively. There is, however, a shortage of research confirming the reliability and validity of assessments to evaluate the rehabilitation needs of workers and the assessment of their work capacity (King et al. 1998; Innes and Straker 1999a; Innes and Straker 1999b; Ekbladh et al. 2004; Gouttebarge et al. 2004). However, applicable research for some of the commercially available assessment tools is beginning to appear in the literature (Ting et al. 2001; Matheson et al. 2002; Durand et al. 2004; Pransky and Dempsey 2004; Reneman et al. 2004). This research is an important contribution to evidence based practice for therapists using FCEs.

Ensuring client safety when conducting FCEs has been identified as a critical issue for consideration (Innes and Straker 2003a; Gibson and Strong 2005). Innes and Straker (2003a) suggest that therapists using FCEs would undertake a decision making process, that considers safety first, then determines the constructs of dependability (quantitative attributes and the concept of reliability) and utility (qualitative attributes and the concept of validity) demonstrated by the FCE used.

Clinical utility reflects the degree of conviction therapists have about the usefulness of an assessment (Toomey et al. 1995). Clinical utility confirms that the tool is related to the purpose for which it is used (Barbara and Whiteford 2005). Table 1 outlines the important features that relate to the clinical utility of a tool.

Table 1: Features related to the clinical utility of a tool.

	Features related to clinical utility
Accuracy	**The evaluation provides accurate and correct information about the client's functional ability (Soanes and Stevenson 2005).**
Comprehensiveness	**The evaluation covers a wide range of aspects of function.**
Credibility	**Therapists using the evaluation require professional skills, knowledge, background and experience (Innes and Straker 2003a).**
Flexibility	**Flexibility of the instrument (Toomey et al. 1995).**
Practicality	**Ease of administration and interpretation of the assessment. Related to the direct and indirect costs of the evaluation procedure (Gibson and Strong 1997; Simmonds 2002).**
Relevance	**The evaluation is relevant to the client and the assessment situation (Innes and Straker 2003a).**
Usefulness	**The test procedure must meet the needs of the client, referrer and payer (Innes and Straker 2003a).**
Suitability	**The evaluation meets its intended purpose (Gibson and Strong 1997).**
Feasibility	**Potential to complete the evaluation and for information provided to be implemented with the client and/or the workplace (Innes and Straker 2003b; Soanes and Stevenson 2005).**
Value	**Providing valuable information to the client and the therapist (Barbara and Whiteford 2005).**
Adaptability	**The assessment is adaptable to various disability types and situations (Gibson and Strong 2002)**

Clinical utility can be considered at an instrument level, at an organizational level and at an individual level (Wind et al. 2006). Psychometric properties and usage issues of the specific tool are investigated when examining the instrument (Law 2002). At an organizational level, clinical utility relates to the application of the instrument to health policy, procedures or processes and issues relating to employees. At an individual level clinical utility relates to the usefulness, purpose and the provision of relevant information about a worker being assessed, and what this information is being used for (Wind et al. 2006).

The clinical utility of an assessment tool provides important information about the usefulness of the tool and the ease with which a therapist can conduct the assessment. Qualitative studies have investigated the clinical utility of specific assessment tools by interviewing users of the assessment tools (Toomey et al. 1995; Barbara and Whiteford 2005). Other qualitative studies have explored how assessments are perceived more generally (Innes and Straker 2003b; Innes and Straker 2003a; Wind et al. 2006).

This study explored the clinical utility of FCEs using a questionnaire distributed to health professionals that investigated their attitudes, practices and perceptions about their use of FCEs. The study addressed the research question:

What are therapists' views on the clinical utility of FCEs and do these perceptions differ between professional groups?

3.3.2 METHOD

A cross sectional study design was used. Following ethics approval, a pilot survey was developed based on findings from an earlier qualitative study (James et al. 2007) that investigated the attitudes and practices of health professionals in relation to FCE usage. The pilot survey also included some qualitative questions relating to the format and structure of the questionnaire. Twelve health professionals completed the pilot survey. Data analysis of the pilot survey resulted in the removal of some items to improve the internal reliability of the survey subscales (Higginbotham et al. 2001).

The final survey consisted of 60 items divided into six sections: Demographic data; Type of FCE used; FCE choice; FCE usage; Perceived consequences of using an FCE and Perceptions of FCEs. Some items sought categorical responses related to the background of the health professional and the type of FCE used. Other items elicited ordinal responses using Likert scales relating to the subscales of FCE choice, FCE usage, perceived consequences of conducting FCEs and perceptions of FCEs. A direct estimation method of a 5, 6 or 7 point Likert scale was used for each sub-scale to measure rater attitudes to a series of statements relating to the use of FCEs (Portney and Watkins 2000).

The final survey was mailed to all 219 WorkCover accredited rehabilitation providers in NSW. Accredited rehabilitation providers are organisations accredited by WorkCover (NSW) to offer specialist services to help injured workers to return to work. Rehabilitation providers employ different health professionals to assess the needs of the injured worker and the workplace requirements. Rehabilitation plans are then developed to assist the injured worker to return to work (WorkCover (N.S.W.) 2002). The Managers of each rehabilitation provider were asked to distribute information about participating in the study to any health professionals who conducted FCEs. Only those health professionals who conducted FCEs were included in the study.

Data analysis

Surveys were coded and data entered into STATA [v8.0] (Statacorp 2003) for descriptive analysis. Cronbach's alpha was used to measure the internal consistency of

the scales and to evaluate the homogeneity of the items for each subscale identified in the survey. A Cronbach's alpha of between .70 and .90 was considered to indicate sufficient internal consistency and indicates the items within the scale are measuring the same construct (Depoy and Gitlin 1998; Portney and Watkins 2000). Analysis of variance (ANOVA) was used to determine any differences between the professional groups on continuous scores from scale items.

3.3.3 RESULTS

Participants

Surveys were returned from 79 participants working for 65 different rehabilitation providers; a response rate of 29.7% from the 219 rehabilitation providers invited to participate. Of those who replied, 82% (n=63) were occupational therapists, 13% (n=10) physiotherapists and 5%(n=5) exercise physiologists. This response is generally representative of the ratio of the different professionals working for rehabilitation providers in NSW. The mean number of years of professional experience was 10.9 years (range: 1yr-29yrs). The mean number of years of FCE experience was 5.3 years (range: 1yr-16yrs).

The results are presented for the four subscales included in the survey instrument: usefulness and relevance; adaptability and flexibility; therapist perceived requirements and issues in practice.

Usefulness and relevance of FCEs

Participants indicated that they obtained relevant and useful clinical information from the FCE used in their workplace and used results to i).predict return to work and pre-injury duties, ii).define functional abilities and limitations, iii).differentiate between manual and sedentary work, iv).help improve the outcomes of retraining, and v).improve the understanding of the case manager about further rehabilitation needs. The Items for this scale were correlated with each other with an internal consistency of 0.82 (Cronbach's alpha).

Participants rated all items related to the usefulness and relevance of FCEs as important (see Figure 1).

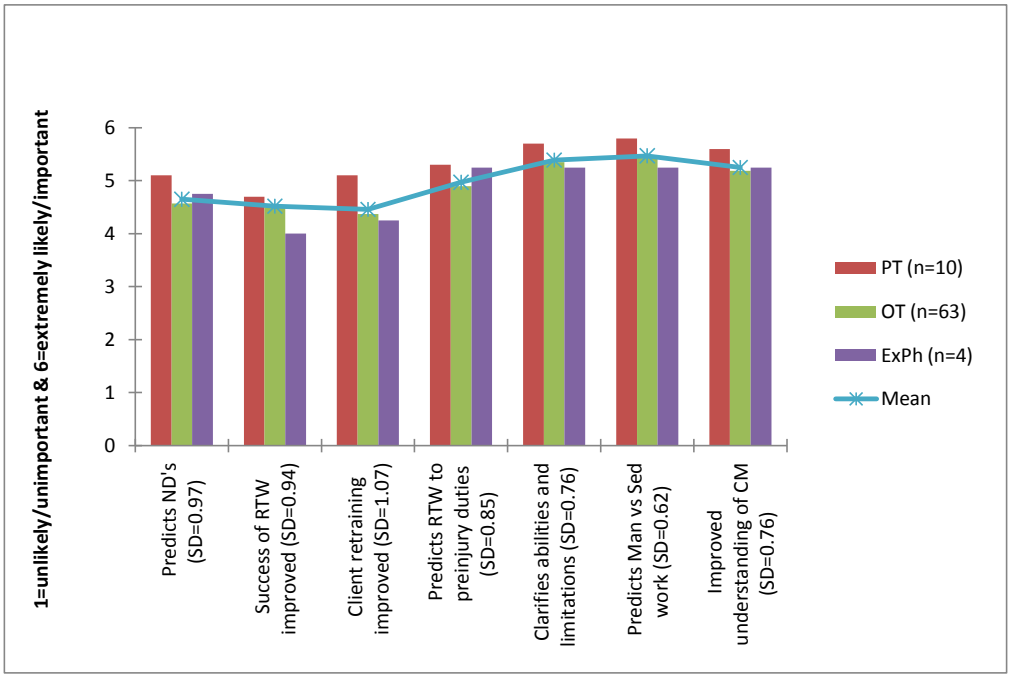


Figure 1: Usefulness and Relevance of FCEs (n=77)

ND= Normal duties; RTW = Return to Work; Man = Manual; Sed = Sedentary; CM = Case Manager

The mean value for the ratings about the usefulness and relevance subscale was 5.04 (95%CI: 4.86-5.23) out of a maximum score of six (most important) for all professionals. Physiotherapists had a slightly higher mean of 5.38 (95%CI: 5.017-5.75), Occupational Therapists a mean of 5.0, (95%CI: 4.78-5.2) and Exercise Physiologists a mean of 4.94, (95%CI: 3.64-6.23). No significant difference was detected between groups (p=1.17)

Adaptability and Flexibility of FCEs

The adaptability and flexibility subscale included items that investigated the clinical utility issues of adapting FCEs and the flexibility of FCEs in practice. Participants indicated the adaptability and flexibility of FCEs was important across all items. Respondents indicated that they adapted FCEs to suit differences in job requirements, injuries or as a result of information provided in medical reports. The perceived importance of the flexibility of FCEs and the characteristics of the assessment tasks within FCEs was also considered important. These results are presented in Figure 2. Items were correlated with each other indicating adequate internal consistency (Cronbach's alpha = 0.7).

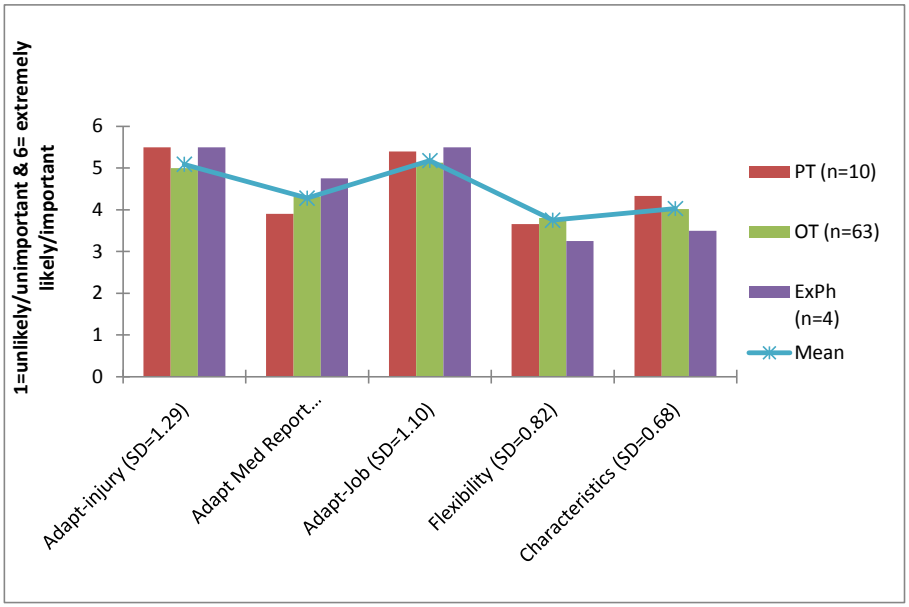


Figure 2: Adaptability and Flexibility of FCEs (n=77)

The overall mean value for the adaptability and flexibility subscale was 4.47 (95%CIi 4.22-4.71), out of a maximum of six (most important). Physiotherapists reported a mean of 4.56 (95%CI: 3.93-5.19), Occupational therapists a mean of 4.45 (95%CI: 4.17-4.72) and Exercise Physiologists a mean of 4.5 (95%CI: 2.9-6.09). No significant differences were detected between groups (p=0.892) on the overall adaptability and flexibility scale.

Therapist perceived requirements

This group of items included requirements for therapists to be able to use FCEs effectively in practice. The most important items were reported as having an adequate knowledge of anatomy and biomechanics, being highly skilled, competent and having further training in the use of FCE assessment tools. Accreditation and a personal knowledge of research literature were rated less important on this subscale. Items on this subscale were correlated with other indicating adequate internal consistency (Cronbach's alpha= 0.74).

Figure 3 outlines the mean scores given by participants for each of the items relating to therapist perceived requirements to administer FCEs.

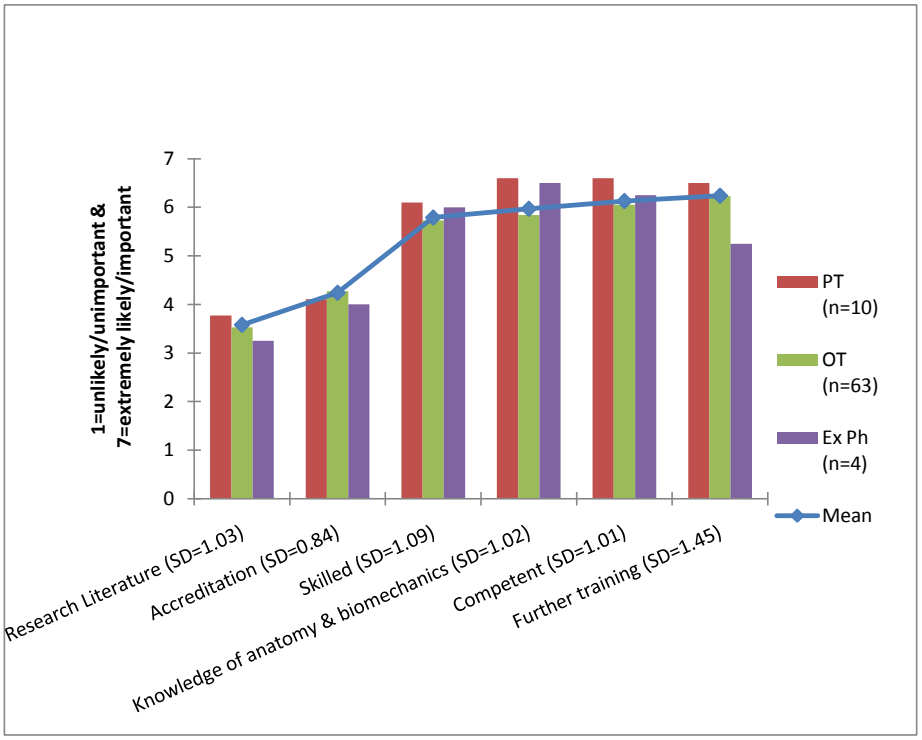


Figure 3: Therapist perceived requirements to administer FCEs (n=77)

The mean value for the therapist perceived requirements scale was 5.32 (95%CI: 5.07-5.57) out of a maximum of six (most important). Physiotherapists reported a mean of 5.61 (95%CI 4.93-6.3), Occupational therapists a mean of 5.29 (95%CI: 5.09-5.48) and Exercise Physiologists a mean of 5.2 (95%CI: 4.23-6.19). No significant difference between groups was detected (p=0.258).

Issues in Practice

Issues related to using FCEs in practice were presented individually in the survey (not in a scale) with issues of availability of the FCE at the workplace rating most importantly. The length of the assessment and the cost were rated less importantly. Figure 4 outlines the mean scores for the different professional groups and as an overall mean for each item.

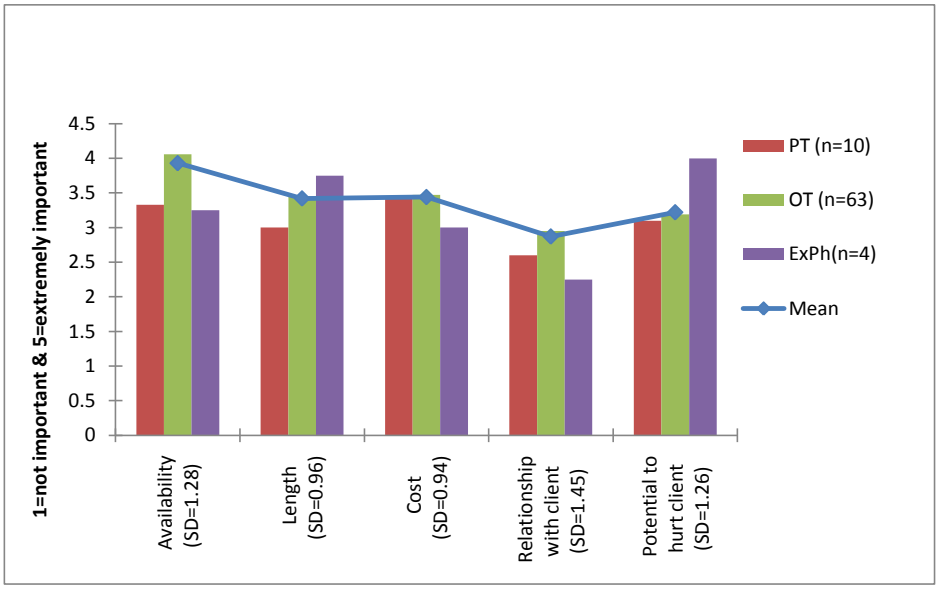


Figure 4: Issues in practice administering FCEs (n=77)

3.3.4 DISCUSSION

This study was conducted in an attempt to understand the overall clinical utility of FCEs and to determine if professional background or level of experience was associated with variations in perceptions.

Differences were anticipated between professional groups because of the diversity of undergraduate theoretical approaches to each professional role. However, the survey results did not confirm this expectation. All professional groups indicated a tendency to adapt FCEs based on a clients' job and injuries and as a result of information provided in medical reports, rather than adhering consciously to professionally defined models of practice. This is suggestive that participants applied pragmatic clinical reasoning when using FCEs. Pragmatic clinical reasoning is defined by (Hagedorn 2001) as 'the evaluation of whether an action is feasible, and whether the context and resources in a given situation facilitate an intervention or make it inadvisable….also takes account of a therapist's knowledge, skill, and interests and wider organizational, socio-cultural and political considerations'. Flexibility of FCEs was also identified as important for all the professionals groups surveyed. These factors impact upon the reliability and validity of an assessment tool, however in this study, it can be suggested that therapists are adapting FCEs in practice despite the risks to the reliability and validity of the findings.

Usefulness and relevance of FCEs

Ratings about the usefulness of FCEs in the management of clients were uniform across all respondents. This suggested some consistency about how FCEs are currently being used in practice. Clarification of clients' abilities and limitations was rated as the most important feature of FCEs in providing information that clarifies client function. FCE usage to predict client ability for manual versus sedentary work, for return to work or for return to work completing pre-injury duties was also rated highly. Although studies examining the predictive benefits of FCEs are scarce and inconclusive, therapists believe that FCEs assist with predicting a client's management (Gross and Battie 2004; Gross et al. 2004). Therapists indicated the FCE also assists case managers (representing professional diversity and varied backgrounds) to have improved understanding of the need for continued rehabilitation involvement, to maximize the potential of successful rehabilitation outcomes.

The study results regarding the usefulness of FCEs are consistent with Wind et al. (2006) who found 'return to work' case managers also perceived FCEs to be useful in the management of clients. Interestingly in his study the claims experts, who worked for the insurers, did not view FCEs to be as useful. Wind (2006) speculates this could be attributed to the difference in organizational context between the two groups (insurers and rehabilitation case managers). In contrast, this study surveyed health professionals working within rehabilitation provider organizations, and they also indicated an FCE was useful for case managers. In the rehabilitation provider context, a case manager would usually have a health related background and therefore it can be speculated that they would have a more comprehensive understanding of the purpose of an FCE and the information it can provide. Further study to investigate the usefulness of FCEs for claims experts (who may not come from a health background), within the Australian context, is needed for additional comparison and to explore their influence on rehabilitation outcomes. Despite the fact that FCEs were viewed as useful by health professionals, further evidence for the reliability and validity of specific FCE tools is needed to support this perception.

Adaptability and flexibility of FCEs

Participants agreed that it was highly likely that an FCE would be adapted according to a client's injury type or job. Flexibility within the assessment and the characteristics of the FCE itself were identified as important components of clinical utility. However, this approach can threaten the standardised application of assessment tools. In this study, the importance of a tool being standardised was rated less importantly than its flexibility. This suggests therapists' value the ability to adapt an FCE more than the

standardised properties of the tool. The reliability and validity of the assessment tool may be compromised in practice if therapists adapt FCEs to suit individual clients.

Several authors (King et al. 1998; Innes and Straker 2003a) discuss the issue of generic versus specialist (or job specific) FCE testing, and the ability to alter the test depending upon the needs of the client, work and situation. Innes and Straker (2003a) found that FCEs conducted for those with no job were more generic and the most quantitative, whereas work assessments for someone returning to a job were the most specific and qualitative. They concluded that FCEs for specific jobs consisted of a balance between the two. (Strong et al. 2004) suggested a continuum of assessment protocol from a fixed protocol through to a flexible protocol based on clinical reasoning. They defined flexible protocol delivery as one where the therapist plans an individualized assessment based on referrer needs and client injury and defined a fixed protocol as one that requires low clinical understanding and is directed by tools and technology. In this study, FCEs were considered and no distinction was made between FCEs for those clients with a job or without a job. However, the finding that adapting the FCE in accordance with the clients' injury and job were rated as highly important indicates agreement with previous studies on this issue (King et al. 1998; Innes and Straker 2003b; Innes and Straker 2003a). This is consistent with the flexible protocol delivery continuum suggested by (Strong et al. 2004).

The characteristics of the FCE itself, was another aspect that was rated as highly important. King et al (1998) suggested there is not one appropriate test for any one client or assessment situation and that the evaluator needs to select the most appropriate for any given situation. The variety of tasks that are assessed within the FCE, the ability to adapt these to suit the individual and the flexibility within the assessment to alter procedures and processes are all characteristics that could be considered important.

Therapist perceived requirements

Therapists strongly agreed that further training to conduct FCEs was necessary; however, the importance of accreditation was rated as less important. This suggests that FCEs are perceived as a specialized area of practice that requires therapists to undertake specific training. This is the case with some standardised FCEs that require therapists to undergo training, obtain accreditation and in some instances have ongoing requirements to maintain accreditation. Low scores were given for the item 'undergraduate education provided me with adequate skills to perform the FCE' suggesting that therapists felt their undergraduate education did not provide adequate

skills in this area. This concurs with the requirement for postgraduate training and accreditation in the use of specific FCE tools. Furthermore, WorkCover (NSW) recognises minimum qualifications of professionals to conduct FCEs. For occupational therapists or physiotherapists, this is a minimum of 3 months occupational rehabilitation experience. Other health professionals are required to undergo an assessment process to be eligible to conduct FCEs under the WorkCover (NSW) system (WorkCover (N.S.W.) 2002).

Therapists identified that they needed to be highly competent, highly skilled and have an adequate knowledge of anatomy and biomechanics in relation to FCEs. This relates to the FCE being viewed as a specialized area of practice. Personal knowledge of published literature was identified as only moderately important, notwithstanding the current imperatives of evidence based practice. Professionals are constantly being encouraged, by the health professions as a whole and by those paying for services, to use evidence based practice when treating or working with clients. Evidence based practice is considered to be the integration of clinical expertise and external clinical evidence (research) when making decisions regarding the care and treatment of individual clients (Law 2002).

If there is limited external clinical evidence (research) relating to a particular service then health professionals will rely predominantly on their clinical expertise when making decisions (Law 2002). Strong et al found assessors of FCEs, in their Canadian study ranged from consumers of research and evidence based practice to those basing their decisions on historical practices (Strong et al. 2004). FCE research is only just beginning to become available for some of the commercially available assessment tools and as such it is important for therapists to be reviewing and updating their knowledge to assist in providing the best possible, evidence based care for clients. Despite this, health professionals, in this study, did not rate this as being an important aspect. Further research could explore the use of evidence and FCEs.

Issues in practice

The important issue identified by therapists conducting FCEs, was the availability of specific FCEs at the workplace. Availability of resources and opportunities was also identified by Strong et al as an issue affecting assessor practice (Strong et al. 2004). Economic reasons may inhibit providers having a variety of assessments for therapists to choose as a result of the cost of purchase of the assessment tool, and any additional costs of training and accreditation for each therapist. Most therapists used only one FCE tool in previous research conducted in NSW, Australia (Cotton et al. 2006). Other

less highly rated aspects identified by therapists were the effort required to conduct the FCE, the cost, and the time it took to complete the assessment. In practice, FCEs are charged to insurers and employers based on the time taken to complete, and costs are discussed and approved by the referrer prior to the service being provided. This was not rated as an important part of the FCE process for therapists. Relationship to the client was not rated as important, despite this being raised as a consideration in previous research (Strong et al. 2004). The potential to hurt a client was rated as more important by the exercise physiologists than the physiotherapists or occupational therapists. This may be related to the exercise physiologists having less experience, knowledge and skills in regards to injury and health due to differences in undergraduate education.

3.3.5 CONCLUSION

This study looked at the perceptions of Health Professionals, who conduct FCEs and who worked for WorkCover accredited rehabilitation providers in NSW, about the clinical utility of FCEs.

The study did not identify any differences in opinion related to the clinical utility of FCEs between professional groups or experience. Rather, results suggest a level of consistency in how FCEs are currently being used in practice across NSW (Australia) and with attitudes towards their clinical utility.

Therapists reported the FCE assisted in predicting RTW and manual versus sedentary duties despite limited research on the predictive validity being available. Therapists valued the flexibility of FCEs and chose to adapt them to suit the client, injury type and job, rather than use standardised measures, as has been found in previous studies. Therapists felt FCE practice was a specialist area of practice and training was needed to be able to conduct these assessments.

Personal knowledge of published literature was only rated as moderately important suggesting therapists are basing decisions on historical information and practices rather than evidence based practice and current research.

Limitations of this research include a small sample and the sample being from one state within Australia – results cannot therefore be generalised beyond this group. Additional follow up with rehabilitation providers may have assisted to increase the response rate of this survey.

Further research could explore the issue of use of evidence with FCEs and to provide further evidence of the reliability and validity of the tools in use.

ACKNOWLEDGEMENTS

The authors would like to thank all therapists who gave their time to participate in this study.

REFERENCES

Barbara, A.& Whiteford, G. (2005). "Clinical Utility of the Handicap Assessment and Resource Tool: An Investigation of Its Use with the Aged People in Hospital." Australian Occupational Therapy Journal **52**(1): 17-25.

Cotton, A., Schonstein, E.& Adams, R. (2006). "Use of Functional Capacity Evaluations by Rehabilitation Providers in NSW." Work **26**(3): 287-295.

Depoy, E.& Gitlin, L. (1998). Introduction to Research. Understanding and Applying Multiple Strategies. St Louis, Mosby.

Durand, M., Loisel, P., Poitras, S., Mercier, R., Stock, S.& Lemaire, J. (2004). "The Interrater Reliability of a Functional Capacity Evaluation: The Physical Work Performance Evaluation." Journal of Occupational Rehabilitation **14**(2): 119-129.

Ekbladh, E., Haglund, L.& Thorell, L. (2004). "The Worker Role Interview - Preliminary Data on the Predictive Validity of Return to Work of Clients after an Insurance Medicine Investigation." Journal of Occupational Rehabilitation **14**(2): 131-141.

Gibson, L.& Strong, J. (1997). "A Review of Functional Capacity Evaluation Practice." Work **9**(1) 3-11.

Gibson, L.& Strong, J. (2002). "Expert Review of an Approach to Functional Capacity Evaluation." Work **19**(3): 231-242.

Gibson, L.& Strong, J. (2005). "Safety Issues in Functional Capacity Evaluation: Findings from a Trial of a New Approach for Evaluating Clients with Chronic Back Pain." Journal of Occupational Rehabilitation **15**(2): 237-251.

Gouttebarge, V., Wind, H., Kuijer, P.& Frings-Dresen, M. (2004). "Reliability and Validity of Functional Capacity Evaluation Methods: A Systematic Review with Reference to Blankenship, Ergos Work Simulator, Ergo-Kit and Isernhagen Work System." International Archives of Occupational and Environmental Health **77**(8): 527-537.

Gross, D.& Battie, M. (2004). "The Prognostic Value of Functional Capacity Evaluation in Patients with Chronic Low Back Pain: Part 2: Sustained Recovery." Spine **29**(8): 920-4.

Gross, D., Battie, M.& Cassidy, J. (2004). "The Prognostic Value of Functional Capacity Evaluation in Patients with Chronic Low Back Pain: Part 1: Timely Return to Work." Spine **29**(8): 914-9.

Hagedorn, R. (2001). Foundations for Practice in Occupational Therapy. London, Churchill Livingstone.

Higginbotham, N., Albrecht, G.& Connor, L. (2001). Health Social Science: A Transdisciplinary and Complexity Perspective. Melbourne, Oxford Press.

Innes, E.& Straker, L. (1999a). "Reliability of Work-Related Assessments." Work **13**(2): 107-124.

Innes, E.& Straker, L. (1999b). "Validity of Work-Related Assessments." Work **13**(2): 125-152.

Innes, E.& Straker, L. (2003a). "Attributes of Excellence in Work Related Assessments." Work **20**(1): 63-76.

Innes, E.& Straker, L. (2003b). "Workplace Assessments and Functional Capacity Evaluations: Current Beliefs of Therapists in Australia." Work **20**(3): 225-236.

James, C., MacKenzie, L.& Higginbotham, N. (2007). "Health Professionals' Attitudes and Practices in Relation to Functional Capacity Evaluations." Work **29**(2): 81-88.

King, P., Tuckwell, N.& Barrett, T. (1998). "A Critical Review of Functional Capacity Evaluations." Physical Therapy **78**(8): 852-866.

Law, M. (2002). Evidence Based Rehabilitation: A Guide to Practice. Thorofare, Slack.

Matheson, L., Rogers, L., Kaskutas, V.& Dakos, M. (2002). "Reliability and Reactivity of Three New Functional Assessment Measures." Work **18**(1): 41-50.

Portney, L.& Watkins, M. (2000). Foundations of Clinical Research: Applications to Practice. New Jersey, Prentice Hall Health.

Pransky, G.& Dempsey, P. (2004). "Practical Aspects of Functional Capactiy Evaluations." Journal of Occupational Rehabilitation **14**(3): 217-229.

Reneman, M., Brouwer, S., Meinema, A., Dijkstra, P., Geertzen, J.& Groothoff, J. (2004). "Test-Retest Reliability of the Isernhagen Work Systems Functional Capacity Evaluation in Healthy Adults." Journal of Occupational Rehabilitation **14**(4): 295-305.

Simmonds, M. (2002). "Physical Function in Patients with Cancer: Psychometric Characteristics and Clinical Usefulness of a Physical Performance Test Battery." Journal of Pain and Symptom Management **24**(4): 404-414.

Soanes, C.& Stevenson, A. E., Eds. (2005). Oxford Dictionary of English. Oxford, Oxford University Press.

Statacorp (2003). Statastatistical Software: Release 8.0. College Station, TX, Stata Corporation.

Strong, S., Baptiste, S., Cole, D., Clarke, J., Costa, M., Shannon, H., Reardon, R.& Sinclair, S. (2004). "Functional Assessment of Injured Workers: A Profile of

Assessor Practices." The Canadian Journal of Occupational Therapy **71**(1): (13-23).

Ting, W., Wessel, J., Brintnell, S., Maikala, R.& Bhambhani, Y. (2001). "Validity of the Baltimore Therapeutic Equipment Work Simulator in the Measurement of Lifting Endurance in Healthy Men." American Journal of Occupational Therapy **55**(2): 184-90.

Toomey, M., Nicolson, D.& Carswell, A. (1995). "The Clinical Utility of the Canadian Occupational Performance Measure." Canadian Journal of Occupational Therapy **62**(5): 242-249.

Wind, H., Gouttebarge, V., Kuijer, P., Sluiter, J.& Frings-Dresen, M. (2006). "The Utility of Functional Capacity Evaluation: The Opinion of Physicians and Other Experts in the Field of Return to Work and Disability Claims." International Archives of Occupational and Environmental Health **79**(6): 528-534.

WorkCover (NSW). (2002). Rehabilitation Providers. Standards and Conditions of Accreditation. WorkCover. Sydney, WorkCover NSW.

3.4 Chapter Summary

The studies presented in this chapter identified that there is consistency in relation to how FCEs are currently used in practice. FCEs are applied differently according to: the reason for referral; client goal; the therapist's workplace procedures, policies and resources; and the therapist's skill and experience. The most important feature was clarification of a client's abilities and limitations. Other important aspects were the usage of FCEs to predict manual versus sedentary work abilities, and for identification of return to work capabilities, including pre-injury, modified or new duties. FCEs are perceived as a specialised area of practice with the personal skills and experience of the health professional identified as significant attributes, and training in the use of specific FCEs recommended. Further investigation of the effect of health professional experience, both general professional experience and FCE specific experience would provide additional information of these identified, important attributes and of any consequence of this experience to FCE practice. As a result of the lack of research related to the WorkHab FCE, it is unknown if health professionals believe this specific instrument reflects the content domain being measured or whether they believe it has the ability to predict work abilities or return to work capabilities. Assessors are trained and accredited in the use of the WorkHab FCE, which supports the findings that this is a specialised area of practice and training in specific tools is required.

A personal knowledge of published literature was identified as only moderately important by health professionals, which raises concerns about the level of evidence based practice within FCE use. A mixture of standardised, and non-

standardised assessments have been shown to be used in practice; however, non standardised tools were most commonly used. This may relate to workplace resources where the availability of FCE tools will impact upon a health professional's practice. Workplace resources may also be related to economic factors such as the cost of FCE tools plus the additional costs of training. Of the standardised tools in use, the WorkHab FCE was the most extensively used, which concurs with previous research (Deen et al. 2002; Cotton et al. 2006) and supports the need for research into the WorkHab FCE. However, it is acknowledged that any research needs to be made accessible to health professionals as well as being relevant to their practice because this will assist in the use of evidence based practice. Health professionals indicated a tendency to adapt a FCE based on the client's job, injuries, medical information and client goals. This has implications for the reliability and validity of the assessment tool, however, as previous research has indicated there is limited research on the validity and reliability of many commercially available FCEs. The attitudes of health professionals to FCEs in general and specifically to adapting the tool to suit the situation may, in part, be related to the lack of published information of the psychometric properties of specific tools. This further supports the need to undertake research to investigate the reliability and validity of the WorkHab FCE and to publish the results as evidence to inform practice.

It is clearly established that the WorkHab FCE is commonly used in Australian occupational rehabilitation settings, specifically in NSW. However, there is a lack of evidence of the reliability and validity of this tool (Innes and Straker 1999a; Innes and Straker 1999b; Innes 2006). The WorkHab FCE is comprised

of a range of functional tasks and chapters four and five extend the research presented in this chapter by investigating the psychometric properties of reliability and validity of the WorkHab FCE. It was not within the scope of this thesis to investigate reliability across all components of the WorkHab FCE and therefore the manual handling (lifting) component was the focus for these studies (chapter four). The manual handling component of the WorkHab FCE, as with other specific FCE tools, is a significant part of the evaluation and provides valuable information that can be applied to the client's work environment. As such, this was the focus for these studies. Likewise, the validity studies were limited to investigation of content and construct validity. A global approach was taken for the content validity study to include all aspects of the WorkHab FCE, but the focus on the lifting component of the WorkHab FCE continued for the construct validity study (chapter five).

The following chapter discusses the test-retest, inter- and intra-rater reliability of the manual handling component of the WorkHab FCE.

Chapter 4 The WorkHab Functional Capacity Evaluation: Reliability

Overview: This chapter discusses reliability as it relates to FCEs and presents the findings of two studies that investigated the reliability of the manual handling components of the WorkHab FCE. These studies are presented in two manuscripts. The first discusses test-retest reliability with healthy subjects and the second reports on inter-rater and intra-rater reliability of the manual handling component of the WorkHab FCE.

Reliability is one component of accurate measurement and is the pre-requisite of a psychometrically sound measurement tool because of the contribution reliability makes to provide dependable, accurate and meaningful data. Reliability is noted as the extent to which a measurement is consistent and produces the same results on more than one occasion when measuring a set of behaviours that remain constant (Schneider et al. 2007; Portney and Watkins 2009). Therefore health professionals can be confident in the data collected and the conclusions drawn from a reliable tool. Assessments need to be reproducible or dependable to enable consistent results to be obtained under the same conditions and with the same assessor.

The usefulness of an assessment tool depends upon the extent that health professionals and those who request (and pay for) services can rely upon data as being accurate and meaningful (Portney and Watkins 2009). Without evidence of psychometric properties, which includes reliability (and validity), the confidence in a tool can be questioned. There are several types of reliability that

an assessment needs to demonstrate: test-retest reliability, intra-rater reliability, inter-rater reliability and internal consistency.

Test-retest reliability relates to the stability of an instrument or measure and is based on repeated administrations of the test to the same individual over a period of time (Bowling 2004; Portney and Watkins 2009). Test-retest reliability is used to establish that an assessment is consistent in measuring a variable assuming that the variable stays constant. To determine this, one sample is tested on two separate occasions with the same assessment tool and assessors (Innes and Straker 1999a). With regard to FCEs, the ideal time interval between testing occasions can be difficult to establish, especially with an injured population who may experience some aggravation from the initial FCE and therefore need time to recover from any such aggravation. Therefore, if the time between testing is too long, then the injured worker's condition may have changed. There are ethical considerations in relation to subjecting an injured population to two such assessments within a short space of time because this may interfere with their recovery and rehabilitation, particularly if they experience an aggravation to their injury. Such an injury will compound recovery and potentially slow the pace of rehabilitation. Test-retest reliability for FCEs is therefore often reported on healthy individuals (Reneman et al. 2004; Soer et al. 2006a; Legge and Burgess-Limerick 2007) or on subjects with stable physical injuries (Tuckwell et al. 2002).

In clinical practice, FCE tools require the assessor to apply operational criteria to observations and rate performance in such a way that the assessor is reliable and consistent in their evaluation of performance. Inter-rater reliability is the extent to which two or more assessors measuring the same group of subjects

obtain the same results (Bowling 2004; Portney and Watkins 2009). Inter-rater reliability is best assessed when assessors measure and rate a response at the same time; however, the use of one or more raters in a 'real life' situation may not always be possible (Portney and Watkins 2009). In the case of a FCE, space, procedures and information used to determine results (such as observation, physiological measures and client comment) may impact upon the ability of more than one rater to be involved. Two raters simultaneously measuring and rating a response has been used to assess reliability in some FCEs (Gardener and McKenna 1999; Gross and Battie 2002). Video recording of subjects performing activities provides an opportunity for multiple observers to rate the same performance with the same information and has also been used to assess inter-rater reliability in FCEs (Smith 1994; Isernhagen et al. 1999; Reneman et al. 2002b; Legge and Burgess-Limerick 2007).

Intra-rater reliability is defined as the degree to which one assessor or rater can obtain the same ratings on multiple occasions of measuring the same variable when their own ratings are compared. Internal consistency is another form of reliability that assesses the degree to which a set of items in an instrument measure the same trait and involves looking at the correlation among items in a scale (Portney and Watkins 2009).

The kappa statistic and the intra-class correlation coefficient methods are commonly promoted as the method of choice for reliability research. The kappa statistic is the preferred method for nominal data and the intra-class correlation coefficient is also recommended as it reflects both degree of correspondence and agreement among ratings and can be used to assess reliability with two or more ratings. The interpretation of the results will depend upon the method

used and various authors caution on the interpretation. It is noted that a low correlation can sometimes be difficult to interpret because it may reflect an actual change rather than poor reliability of an instrument. It is suggested that results need to be considered in accordance with how the data will be used and the degree of precision required for any particular assessment (Innes and Straker 1999a; Bowling 2004; Legge and Burgess-Limerick 2007; Portney and Watkins 2009). Cronbach's alpha is commonly used to assess internal consistency as this can be used on scale items that are dichotomous or ordinal (Portney and Watkins 2009).

Evidence based practice is encouraged when treating or working with clients in any setting. Reliable assessment tools provide evidence for best practice and also confidence in the accuracy and meaningfulness of results and findings. The WorkHab FCE, despite being widely used in the Australian occupational rehabilitation arena, specifically in NSW, has limited published evidence of psychometric properties to inform practice (Innes and Straker 1999b; Innes and Straker 1999a; Cotton et al. 2006; Legge and Burgess-Limerick 2007; James and Mackenzie 2009b).

This chapter discusses reliability of the manual handling component of the WorkHab FCE specifically investigating the test-retest reliability in a study with healthy participants, and inter-rater and intra-rater reliability of the WorkHab FCE manual handling component and internal consistency of the WorkHab FCE manual handling scoring system in a study with injured participants.

Test-retest reliability was studied with a convenience sample of 25 healthy adults who completed the manual handling component of the WorkHab FCE on

two occasions, a week apart. Healthy subjects were selected for convenience and due to difficulties with regards to the recruitment of an injured population and potential ethical issues as previously discussed. The results are discussed in **Manuscript 4 (4.1):** *"Test- retest reliability of the manual handling component of the WorkHab FCE in healthy adults"* published in 2010 in Disability and Rehabilitation, 32(22): 1893-9. This research has also been presented as an oral poster at the 15th Congress of the World Federation of Occupational Therapists in Chile in May 2010 (Appendix 2).The study found substantial levels of test-retest reliability for the lifting components of the WorkHab FCE in healthy adults.

A cross sectional study was conducted to investigate intra-rater and inter-rater reliability. A DVD was produced that contained footage of the manual handling components of the WorkHab FCE conducted with four injured workers. Health professional raters (n=17) who were trained and accredited in use of the WorkHab FCE scored these components and 14 of these raters re-evaluated them after approximately two weeks. Findings are discussed in **Manuscript 5 (4.2):** *"Inter and intra-rater reliability of the manual handling component of the WorkHab functional capacity evaluation"* published early online (January 19, 2011) in Disability and Rehabilitation, DOI:10.3109/09638288.2010.548896. This research has also been accepted for presentation at the Occupational Therapy Australia, 24th National Conference to be held 29 June – 1 July 2011 at the Gold Coast, Australia. This study found substantial levels of inter-rater and intra-rater reliability for the manual handling components of the WorkHab FCE.

4.1 Manuscript: Test - Retest reliability of the Manual Handling component of the WorkHab Functional Capacity Evaluation in healthy adults

Citation: James, C., Mackenzie, L. and Capra, M. (2010).“Test - Retest reliability of the Manual Handling component of the WorkHab Functional Capacity Evaluation in healthy adults.” Disability and Rehabilitation 32(22):1893-9.

ABSTRACT

Purpose

The WorkHab Functional Capacity Evaluation (FCE) is one of many FCEs currently available and is widely used in the Australian workplace injury management and occupational rehabilitation arena. This study investigated the test-retest reliability of manual handling tasks within the WorkHab Functional Capacity Evaluation (FCE) in healthy adults.

Method

A convenience sample of 25 healthy subjects, consisting of 19 women and 6 men with a mean age of 29 years (SD: 12.0) participated in this study. Two FCE sessions were held a week apart and subjects completed a floor to bench, bench to shoulder and bench to bench lift. Analysis of the outcomes of the FCE included; descriptive analysis; Intra Class Correlations (ICC); kappa; percentage agreement; and 95% limits of agreement where appropriate.

Results

The ICC’s for the three lifts show an excellent reliability (0.90 – 0.92), and a moderate reliability for the manual handling score (0.74). Further analysis of the components of the manual handling score found the percentage agreement was high for all components ranging from 72-92%, however the kappa scores suggested poor to moderate reliability (range -0.06 to 0.52). Internal consistency of the manual handling score was good (Cronbach’s Alpha = 0.92) indicating this is a reliable scale.

Conclusions

The ratings for the lifting components identified substantial levels of test-retest reliability for the lifting components of the WorkHab FCE in healthy adults.

KEYWORDS: Functional Capacity Evaluation, Occupational Rehabilitation, WorkHab Functional Capacity Evaluation, Test-retest reliability.

4.1.1 INTRODUCTION

Functional capacity evaluations (FCEs) are an integral part of work injury prevention and occupational rehabilitation. FCEs are designed to provide a comprehensive performance based assessment, to define the functional abilities and limitations of an individual in the context of safe, productive work tasks and are used for work fitness determinations and to facilitate return to work (King et al. 1998; Lee et al. 2001; Gouttebarge et al. 2006; Gross et al. 2007). Functional capacity evaluations are commonly used with individuals, who have suffered work related musculo-skeletal injuries, as part of the rehabilitation process, and can be used for a range of injury and disease types. FCEs have been broadly categorised into three groups; for those with and without a specific job; and a job or work capacity evaluation (Innes and Straker 2003; Jones and Kumar 2003). There are a variety of FCEs commercially available (Jay et al. 2000; Ting et al. 2001; Matheson et al. 2002; Boadella et al. 2003; Durand et al. 2004; Gouttebarge et al. 2004; Reneman et al. 2004; Reesink et al. 2007; Durand et al. 2008).The WorkHab functional capacity evaluation is a popular assessment in the Australian occupational rehabilitation arena (Deen et al. 2002; Cotton et al. 2006; James and Mackenzie 2009). It is used in any of the three categories of FCE mentioned above, however, there is limited published literature on it's psychometric properties (Innes and Straker 1999a; Legge and Burgess-Limerick 2007). The usefulness of an assessment tool depends upon the extent that health professionals and those requesting (and paying for) services can rely upon data as being accurate and meaningful (Portney and Watkins 2009). Without evidence of psychometric properties including reliability and validity the confidence in a tool can be questioned. Evidence based practice is encouraged when treating or working with clients, however in relation to the WorkHab FCE there is limited published evidence available to inform practice, despite it's wide spread use (Innes and Straker 1999b; Innes and Straker 1999a; Cotton et al. 2006; Legge and Burgess-Limerick 2007; James and Mackenzie 2009).

Portney and Watkins (2009) identify reliability of an assessment as the first prerequisite when considering measurement and a precursor to determining validity. Reliability is the extent to which a measure is consistent, free from error, and demonstrates the reproducibility or dependability of the assessment over time (Portney and Watkins 2009). Test-retest reliability indicates stability of the assessment and determines the consistency of measures from one testing occasion to another, on the assumption that the behaviour being scored does not change over time (Portney and Watkins 2009).

Test-retest reliability can be influenced by: testing effects (practice); rater bias; and the test-retest interval or time between the two testings.

Healthy adult populations have been used by various researchers in determining measurement properties of FCEs (Ting et al. 2001; Reneman et al. 2004; Soer et al. 2006) and are often chosen for convenience. For use in test-retest reliability studies, healthy participants overcome the possibility of results being affected by changes in performance between testing sessions if injured workers are used. Access to an injured worker population for research is not always easy (Tuckwell et al. 2002) and may be difficult when injured workers are concurrently being managed in a litigious workers compensation system.

This study aimed to evaluate the test-retest reliability for the floor to bench, bench to shoulder and bench to bench lifts of the WorkHab FCE in healthy adults. The study also aimed to investigate the test-retest reliability of the manual handling score given as part of the WorkHab FCE and to investigate the internal consistency of this scoring system. This research is an important contribution to evidence based practice for therapists using the WorkHab FCE in practice.

4.1.2 METHOD

Subjects

A convenience sample of 25 healthy adult volunteers recruited from a University staff and student population participated in this study. This sample size gave a power of 90%. The study sample consisted of 19 women and 6 men ranging in age from 19 years to 54 years, with a mean age of 29 years (SD :12.0). The mean weight was 66.4kg (SD: 10.4) and the mean height was 167.9cm (SD: 8.4).

WorkHab FCE

The WorkHab FCE is based on objective physiological measures, observations of biomechanics, reported pain and ratings of client perceived exertion (effort).The procedures for the manual handling component of the WorkHab FCE uses a modular box system, which allows boxes to be stacked at various heights. Each lift is fully explained and demonstrated to the subject prior to commencement. Boxes are set at an appropriate height, and the subject is instructed to lift the load box (initially empty) from beginning (e.g.bench) to end height (e.g. shoulder) and return. This is repeated three times before additional weight is added to the load box. The assessment uses a protocol of increasing load at each height until the safe maximum lifting limit is reached. Baseline heart rate is taken initially and then heart rate readings are taken after each

three lift set. During the WorkHab FCE, if a manual handling technique is observed to be poor, the assessor should educate the subject before proceeding (Bradbury and Roberts 1998). Termination of the assessment can occur as a result of any of the following: i) the subject choosing to cease the activity; ii) if the subjects' heart rate, as determined by the heart rate monitor worn throughout the assessment, exceeds predetermined levels of age predicted maximal heart rate (Holtgrefe and Glenn 2007); or iii) the assessor terminates the activity if lifting becomes unsafe. The assessor records the results of the weights lifted in kilograms after each component. The assessor also calculates the Manual handling score, out of 20. This is calculated from the manual handling components of stance, posture, leverage, torque and pacing using the scale of 0-4 with '0' being no adherence and '4' being the highest score. The higher the score, the more appropriate the manual handling technique used during the lifting component of the FCE.

Procedures

Ethics was obtained from the University Human Research Ethics Committee, following which subjects were recruited using posters to advertise the study and via an email sent from the School of Health Sciences office at the University inviting participation. Interested persons contacted the researcher directly to discuss the study, receive an information statement and subsequently arrange an assessment time.

Prior to commencing the FCE, each subject gave informed consent and signed a consent form. Each subject completed a pre-assessment screening, including: completion of a questionnaire to determine medical status; a musculoskeletal evaluation and blood pressure check, to determine any medical risks and to screen for current injuries -this was not used as a predictor of performance. Subjects also completed a 3 minute step test (aerobic fitness test) to determine heart rate recovery times prior to undertaking the manual handling component of the assessment.

For this study the height of the lift was relative to the subject's waist (for the bench components) and shoulder (for the shoulder component).The manual handling component / safe maximal lifting limits were determined for a floor to bench lift, a bench to shoulder lift and a bench to bench lift. Two testing sessions were held a week apart with the time of day being kept constant where possible. The FCE was conducted by one Occupational Therapist who was trained and accredited as a WorkHab assessor, with ten years experience in conducting FCEs. The subjects were asked to perform to their maximum abilities. Subjects were not provided with information on the results of

session 1 until after session 2 was completed. Following session 1, the assessor was not given access to these results again until the conclusion of session 2.

Data analysis

All data was entered into SPSS (version16.0) for analysis. Descriptive analysis, one way random Intraclass Correlation Coefficients (ICC's), 95% confidence intervals, limits of agreement (Bland and Altman 1999), paired sample t-test, kappa (weighted for ordinal data) and percentage agreement were calculated where appropriate. A ratio between the limits of agreement and the mean score was also calculated using the following formula (1.96 x standard deviation of mean difference) / mean session 1 and 2 X100%. Percentage agreement, as a measure of agreement, can be used to determine reliability and kappa is a change corrected measure of agreement considering both the proportion of observed agreements and the proportion expected by chance (Portney and Watkins 2000). The internal consistency of the components of the FCE manual handling scoring system were calculated using Cronbach's alpha.

An ICC of 0.90 or more was considered a measure of excellent reliability, an ICC of 0.75 – 0.90 was considered good and an ICC of less than 0.75 was considered moderate to poor (Innes and Straker 1999a; Portney and Watkins 2009). A kappa score of more than 0.80 represents excellent agreement, above 0.60 represents substantial levels of agreement; from 0.40 – 0.60 represents moderate agreement and below 0.40 poor to fair agreement (Portney and Watkins 2009). A Cronbach's alpha of between 0.70 and 0.90 was considered to indicate sufficient internal consistency and indicates the items within the scale are measuring the same construct and can be considered reliable (Depoy and Gitlin 1998; Portney and Watkins 2000).

4.1.3 RESULTS

The means, standard deviations, limits of agreement, 95% confidence intervals, and ICC's for the lifts and manual handling score are presented in table 1.

Table 1: Results of means, differences, standard deviations, 95% confidence intervals and ICC's for lifts.

Lift	Mean 1[a]	SD 1	Mean 2[b]	SD[c] 2	Mean dif	SD of mean dif	95% CI[d] of difference	Limits of agreement	Ratio of LoA & mean (%)[e]	ICC[f]	95% CI of ICC	Interpretation of ICC
Floor to Bench	17.0	3.9	16.8	3.6	0.18	2.1	0.68 to 1.04	-4.0 to 5.2	24%	0.92	0.82 to 0.96	Excellent
Bench to Shoulder	13.3	3.3	13.0	2.6	0.31	1.7	-0.41 to 1.03	-3.0 to 4.0	22.6%	0.90	0.78 to 0.96	Excellent
Bench to Bench	16.8	4.7	16.8	3.8	-0.02	2.5	-1.07 to 1.04	-3.7 to 2.2	21%	0.91	0.79 to 0.96	Excellent
Manual Handling score	15.7	2.04	16.7	1.3	-1.02*	1.3	-1.55 to -0.49	-	-	0.74	0.42 to 0.88	Moderate

[a]: Mean 1 = group mean in first session (kg).

[b] Mean 2 = group mean in second session (kg).

[c] SD = Standard deviation.

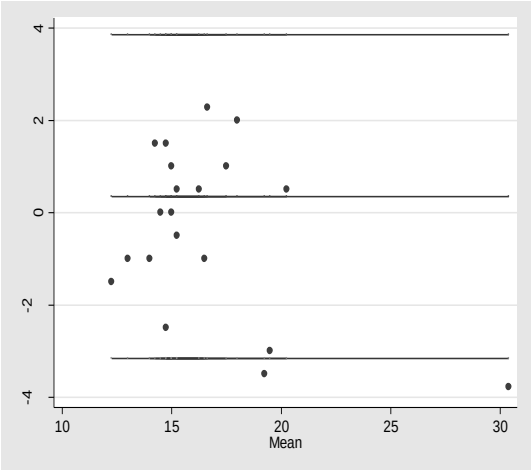
[d] 95% CI = 95% Confidence interval.

[e] ratio between limits of agreement and mean score x100%

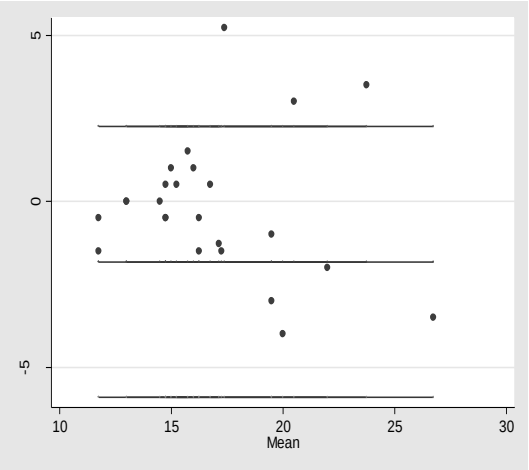
[f] ICC = Intraclass Correlation Coefficient (one way random).

*Significant (two tailed) at $p<0.05$

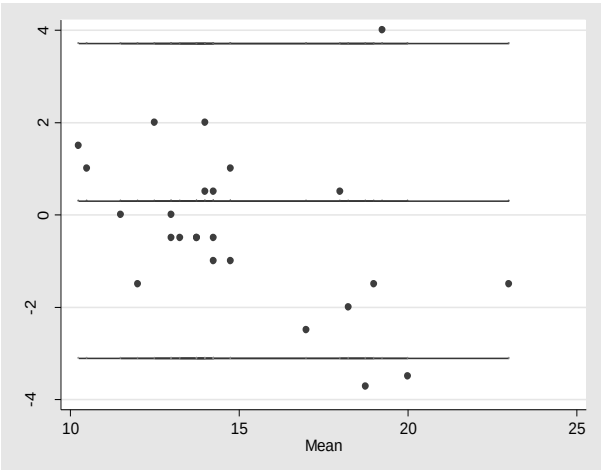
The ICC's for the three lifts show an excellent reliability (0.90 – 0.92), with the CI lower bounds being considered highly reliable. The limits of agreement of the three lifts ranged between ± 24% of the mean score. This equates to ± 3 to 4kg variation in weight lifted. Figure 1 shows the distributions of the limits of agreement, where it can be seen that the majority of lifts were within ± 2.5kg difference, however lower levels of agreement are seen with higher mean loads.



Bench to bench lift



Floor to bench lift



Bench to shoulder lift

Figure 1: Limits of Agreement: Bench to bench, Floor to bench and Bench to shoulder lifts.

A moderate reliability for the manual handling score (0.74) was identified. Further analysis, using a paired sample t-test, identified that manual handling scores in session two were rated significantly higher than the manual handling scores in session one. To more closely examine the components of the manual handling score (stance, posture, torque, leverage and pace), kappa and percentage agreements were calculated. The percentage agreement between the initial test and retest for these manual handling

components was high in most cases however the kappa results suggested more chance agreement. These results are presented in table 2.

To determine the internal consistency of the manual handling score the Cronbach's alpha was calculated for each lift indicating sufficient internal consistency, and that the items within the scale are measuring the same construct, in this instance manual handling technique. Together, the Cronbach's alpha for all manual handling was 0.92 indicating high internal consistency and therefore reliability of this rating scale.

Table 2: Percentage agreement and Cronbach's alpha for the manual handling score components.

Lift	Manual Handling component	Percentage agreement between test & retest	Kappa	Cronbach's alpha.	Cronbach's alpha overall
Floor to Bench	Stance	87%	Moderate (0.47)		
	Posture	92%	Moderate (0.52)		
	Leverage	80%	Poor (0.32)	0.86	
	Torque	84%	Poor (0.25)		
	Pace	80%	Poor (0.31)		
					0.92
Bench to Shoulder	Stance	79%	Poor (0.30)		
	Posture	85%	Poor (0.24)	0.85	
	Leverage	80%	Poor (0.30)		
	Torque	81%	Poor (0.38)		
	Pace	72%	Poor (-0.06)		
Bench to Bench	Stance	85%	Moderate (0.51)		
	Posture	85%	Moderate (0.41)		
	Leverage	75%	Poor (0.33)	0.82	
	Torque	84%	Poor (0.35)		
	Pace	83%	Poor (0.17)		

4.1.4 DISCUSSION

Literature has highlighted the importance of therapists using reliable and valid assessment tools to identify abilities and limitations of individuals (Innes and Straker 1999a). Test-retest reliability is used to determine the consistency of a measure from one testing occasion to another (Portney and Watkins 2009). Test-retest reliability for the lifting components of the WorkHab FCE was expressed by Intraclass Correlation Coefficients (ICC's) which is a measure of between subject variance and within subject

variance and is an accepted measure of reliability in relation to the discriminative capacity of a test. To avoid bias, the test-retest time intervals must be far enough apart to avoid fatigue and close enough to avoid changes in performance. In this study a one week interval was used, with testing occurring at a similar time of day in both instances wherever possible to reduce bias.

For all three lifts evaluated in this study, the test-retest reliability was high (ICC's: 0.90-0.92). 95% limits of agreement were calculated as a descriptive measure of agreement with results being considered from a clinical rather than a statistical interpretation. Clinically a variation of ± 3 to 4kg in weight lifted maybe appropriate in some situations where heavier loads are lifted however, when looking at the distribution of the limits of agreement (Figure 1), the majority of lifts were within ± 2.5kg difference, which is clinically more acceptable. The results of any FCE need to be interpreted and applied to the specific tasks, job and workplace for the individual, using clinical judgement skills. Therefore, when considering the use of FCE to determine functional ability for return to work this variation can be regarded as acceptable and it implies that this is an acceptable clinical measure of agreement when looking at physical demands related to work tasks. This suggests that the administration procedures for the WorkHab FCE are dependable and the average performance by the subjects was relatively stable over the study period. Similar types of lifting tasks evaluated in other FCEs such as the Ergo-kit, Isernhagen and Physical Work Performance Evaluation, have reported substantial or acceptable test-retest reliability (Reneman et al. 2002; Tuckwell et al. 2002; Reneman et al. 2004; Gouttebarge et al. 2006; Legge and Burgess-Limerick 2007).

Whilst the manual handling scores indicated only a moderate level of reliability, there was a significant increase in scores from the first to the second assessment, which may be due to an improvement in the manual handling technique in the retest. The testing effect of practice with this sample suggested that the education provided and practiced in the initial test was learned and applied in the retest situation. Bloom suggests that timing is crucial in mastering learning, with weaker students needing more time to reach proficiency than more able students. In our study, a university population was used, therefore, these individuals, by nature of the environment sampled, are more able, suggesting the learning effect from the initial test was transferred to the retest resulting in an improvement in manual handling technique (Bloom 1976).

The percentage agreement results for the manual handling score components need to be considered in conjunction with the kappa results when determining test-retest

reliability, taking into account any learning effects. The kappa scores were all in the poor to moderate range, although the percentage agreement was good. Both these calculations have limitations as percentage agreement does not take into account chance whereas the kappa does, and in the case of lack of variation in the spread of data across cells, large discrepancies may exist between the kappa score and the percentage of agreement (Innes and Straker 1999a), as was the case in this study. This discrepancy can result from most agreement being limited to only one of the possible rating choices which means only one decision can make the difference between a poor and excellent reliability kappa score (Brouwer et al. 2005).

The other consideration in relation to the manual handling score is that from a clinical perspective, an improvement in this score is positive, as the individual has learned and is utilising appropriate lifting techniques. In practice, the administration of the WorkHab FCE requires feedback to the client to ensure appropriate lifting techniques are learned. Therefore, adequate test-retest reliability for these manual handling components can be assured from these results when applied to the practice context, given the overall results.

Results indicated that the WorkHab FCE manual handling scoring scale demonstrated sufficient internal consistency, where all the items within the scale for each lift measured the same construct of manual handling. This provides further evidence for clinicians about the reliability of the manual handling scoring scale.

One limitation of this study is that subjects were healthy individuals with no manual handling restrictions, which may have produced more positive reliability than if subjects were from an injured population. As the WorkHab FCE is typically conducted with injured workers, this needs to be taken into account, however, the WorkHab FCE is also used as a pre-employment functional capacity evaluation (Legge and Burgess-Limerick 2007) and in this instance would be assessing healthy individuals, so the study results can be confidently applied with this population. This study used only one assessor who was a trained, experienced WorkHab FCE assessor, and data collected during the initial test was not reviewed prior to the retest occasion to reduce bias. The small sample may also be considered a limitation, although sample size calculations indicated that this number provided us with adequate power to make generalizations across a healthy population. Further studies on an injured population would overcome the limitations of this study.

4.1.5 CONCLUSION

The primary purpose of this research was to investigate the test –retest reliability for the floor to bench, bench to shoulder and bench to bench lift during the WorkHab FCE with healthy adults. The results of the lifting found substantial levels of test-retest reliability with this group. It can therefore be concluded, that these lifts in the WorkHab FCE are reliable in healthy adults. The manual handling scoring scale, as part of the WorkHab FCE was also investigated and had good internal consistency. However, the results relating to test-retest reliability showed moderate reliability as a result of an improvement in manual handling technique at the retest assessment. Further research is recommended to establish other forms of reliability and validity of this assessment tool, using a range of client samples. The results from this study contribute to the growing evidence of FCEs in practice and the importance of reliability and validity in work related assessments.

ACKNOWLEDGEMENTS

The authors would like to thank the individuals who volunteered to participate in this study and WorkHab Australia who donated the equipment to complete the study. This study was made possible with the support of a University of Newcastle grant (No: GO187308).

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

Bland, J.& Altman, D. (1999). "Measuring Agreement in Method Comparison Studies." Statistical Methods in Medical Research **8**(2): 135-160.

Bloom, B. (1976). Human Characteristics and School Learning. New York, McGraw-Hill.

Boadella, J., Sluiter, J.& Frings-Dresen, M. (2003). "Reliability of Upper Extremity Tests Measured by the Ergos Work Simulator: A Pilot Study." Journal of Occupational Rehabilitation **13**(4): 219-232.

Bradbury, S.& Roberts, D. (1998). Workhab Functional Capacity Evaluation Procedural Manual. WorkHab Australia.

Brouwer, S., Dijkstra, P., Stewart, R., Goeken, L., Groothoff, J.& Geertzen, J. (2005). "Comparing Self Report, Clinical Examination and Functional Testing in the Assessment of Work-Related Limitations in Patients with Chronic Low Back Pain." Disability and Rehabilitation **27**(17): 999-1005.

Cotton, A., Schonstein, E.& Adams, R. (2006). "Use of Functional Capacity Evaluations by Rehabilitation Providers in NSW." Work **26**(3): 287-295.

Deen, M., Gibson, L.& Strong, J. (2002). "A Survey of Occupational Therapy in Australian Work Practice." Work **19**(3): 219-230.

Depoy, E.& Gitlin, L. (1998). Introduction to Research. Understanding and Applying Multiple Strategies. St Louis, Mosby.

Durand, M., Brassard, B., Hong, Q., Lemaire, J.& Loisel, P. (2008). "Responsiveness of the Physical Work Performance Evaluation, a Functional Capacity Evaluation, in Patients with Low Back Pain." Journal of Occupational Rehabilitation **18**(1): 58-67.

Durand, M., Loisel, P., Poitras, S., Mercier, R., Stock, S.& Lemaire, J. (2004). "The Interrater Reliability of a Functional Capacity Evaluation: The Physical Work Performance Evaluation." Journal of Occupational Rehabilitation **14**(2): 119-129.

Gouttebarge, V., Wind, H., Kuijer, P.& Frings-Dresen, M. (2004). "Reliability and Validity of Functional Capacity Evalutation Methods: A Systematic Review with Reference to Blankenship, Ergos Work Simulator, Ergo-Kit and Isernhagen Work System." International Archives of Occupational Environmental Health **77**(8): 527-537.

Gouttebarge, V., Wind, H., Kuijer, P. P., Sluiter, J. K.& Frings-Dresen, M. (2006). "Reliability and Agreement of 5 Ergo-Kit Functional Capacity Evaluation Lifting Tests in Subjects with Low Back Pain." Archives of Physical Medicine and Rehabilitation **87**(10): 1365-1370.

Gross, D., Battie, M.& Asante, A. (2007). "Evaluation of a Short Form Functional Capacity Evaluation: Less Maybe Best." Journal of Occupational Rehabilitation **17**(3): 422-435.

Holtgrefe, K.& Glenn, T. (2007). Principles of Aerobic Exercise. Therapeutic Exercise: Foundations and Techniques. C. Kisner & Colby, L. Philadelphia, F. A. Davis. **5th ed:** 231-49.

Innes, E.& Straker, L. (1999a). "Reliability of Work-Related Assessments." Work **13**(2): 107-124.

Innes, E.& Straker, L. (1999b). "Validity of Work-Related Assessments." Work **13**(2): 125-152.

Innes, E.& Straker, L. (2003). "Attributes of Excellence in Work Related Assessments." Work **20**(1): 63-76.

James, C.& Mackenzie, L. (2009). "Health Professional's Perceptions and Practices in Relation to Functional Capacity Evaluations: Results of a Quantitative Survey." Journal of Occupational Rehabilitation **19**(2): 203-211.

Jay, M., Lamb, J., Watson, R., Young, I., Fearon, F., Alday, J.& Tindall, A. (2000). "Sensitivity and Specificity of the Indicators of Sincere Effort of the Epic Lift Capacity Test on a Previously Injured Population." Spine **25**(11): 1405-12.

Jones, T.& Kumar, S. (2003). "Functional Capacity Evaluation of Manual Materials Handlers: A Review." Disability and Rehabilitation **25**(4-5): 179-191.

King, P., Tuckwell, N.& Barrett, T. (1998). "A Critical Review of Functional Capacity Evaluations." Physical Therapy **78**(8): 852.

Lee, G., Chan, C.& Hui-Chan, C. (2001). "Work Profile and Functional Capacity of Formwork Carpenters at Construction Sites." Disability and Rehabilitation **23**(1): 9-14.

Legge, J.& Burgess-Limerick, R. (2007). "Reliability of the Jobfit System Pre-Employment Functional Assessment Tool." Work **28**(4): 299-312.

Matheson, L., Rogers, L., Kaskutas, V.& Dakos, M. (2002). "Reliability and Reactivity of Three New Functional Assessment Measures." Work **18**(1): 41-50.

Portney, L. & Watkins, M. (2000). Foundations of Clinical Research: Applications to Practice. New Jersey, Prentice Hall Health.

Portney, L.& Watkins, M. (2009). Foundations of Clinical Research: Applications to Practice. Upper Saddle River, New Jersey, Pearson Prentice Hall.

Reesink, D., Jorritsma, W.& Reneman, M. (2007). "Basis for a Functional Capacity Evaluation Methodology for Patients with Work-Related Neck Disorders." Journal of Occupational Rehabilitation **17**(3): 436-449.

Reneman, M., Brouwer, S., Meinema, A., Dijkstra, P., Geertzen, J.& Groothoff, J. (2004). "Test-Retest Reliability of the Isernhagen Work Systems Functional Capacity Evaluation in Healthy Adults." Journal of Occupational Rehabilitation **14**(4): 295-305.

Reneman, M.., Jaegar, S., Westmaas, M.& Goeken, L. (2002). "The Reliability of Determining Effort Level of Lifting and Carrying in a Functional Capacity Evaluation." Work **18**(1): 23-27.

Soer, R., Gerrits, E.& Reneman, M. (2006). "Test-Retest Reliability of a WRULD Functional Capacity Evaluation in Healthy Adults." Work **26**(3): 273-280.

Ting, W., Wessel, J., Brintnell, S., Maikala, R.& Bhambhani, Y. (2001). "Validity of the Baltimore Therapeutic Equipment Work Simulator in the Measurement of Lifting Endurance in Healthy Men." American Journal of Occupational Therapy **55**(2): 184-90.

Tuckwell, N., Straker, L.& Barrett, T. (2002). "Test-Retest Reliability on Nine Tasks of the Physical Work Performance Evaluation." Work **19**(3): 243-253.

4.2 Manuscript: Inter and Intra-rater reliability of the Manual Handling component of the WorkHab Functional Capacity Evaluation

Citation: James, C., Mackenzie, L. and Capra, M. (early online). "Inter and Intra-rater reliability of the Manual Handling component of the WorkHab Functional Capacity Evaluation." Disability and Rehabilitation DOI:10.3109/09638288.2010.548896 (posted online: January 19, 2011).

ABSTRACT

The WorkHab Functional Capacity Evaluation (FCE) is widely used in Australian workplace injury management and occupational rehabilitation arenas however there is a lack of published literature regarding its reliability and validity.

Purpose

This study investigated the intra and inter-rater reliability of the manual handling component of this FCE.

Method

A DVD was produced containing footage of the manual handling components of the WorkHab conducted with four injured workers. Therapist raters (n=17) who were trained and accredited in use of the WorkHab FCE scored these components and 14 raters re-evaluated them after approximately 2 weeks. Ratings were compared using Intraclass Correlation Coefficients (ICC's), paired sample t-tests (intra-rater), chi-squared (inter-rater), and percentage agreement.

Results

Intra-rater agreement was high with ICC's for the manual handling components and manual handling score showing excellent reliability (0.94 – 0.98) and good reliability for identification of the safe maximal lift (ICC: 0.81).

Overall inter-rater agreement ranged from good to excellent for the manual handling components and safe maximal lift determination (ICC >0.9). Agreement for safe maximal lift identification was good.

Conclusions

Ratings demonstrated substantial levels of intra-rater and inter-rater reliability for the lifting components of the WorkHab FCE.

KEYWORDS: Functional Capacity Evaluation, Work, WorkHab Functional Capacity Evaluation, Reliability, Lifting.

4.2.1 INTRODUCTION

Functional Capacity Evaluations (FCE) are widely used in the area of work injury management and occupational rehabilitation to define an individual's functional work abilities or limitations (Gibson and Strong 1997; King et al. 1998; Jundt and King 1999; Lee et al. 2001). Lifting capacity is one component commonly assessed during an FCE (Jones and Kumar 2003). There are many different FCEs commercially available and clinics have also developed their own non- standardised, work specific FCEs (Jay et al. 2000; Ting et al. 2001; Matheson et al. 2002; Boadella et al. 2003; Durand et al. 2004; Gouttebarge et al. 2004; Reneman et al. 2004; Cotton et al. 2006; Durand et al. 2008; James and Mackenzie 2009b).

There are several approaches to assessment of capacity that can be used in the FCE (Gardener and McKenna 1999). These include: a metabolic approach based on quantification of physiological measures (oxygen uptake, physiological stress etc); a biomechanical approach which is based upon the ability to perform work within safe musculoskeletal or neuromuscular limits; a psychophysical approach which assesses the ability to perform work within perceived limitations (Gardener and McKenna 1999); and a combination of these. The latter acknowledges the complex inter-relationship between physical, cognitive, behavioural and contextual functioning.

The WorkHab Functional Capacity Evaluation is an Australian standardised assessment used in occupational rehabilitation to determine functional capacity. It incorporates physiological and kinematic performance measures including heart rate, pain, perceived exertion and the observation of biomechanics (Bradbury and Roberts 1998). There is however, limited evidence on the psychometric properties of this assessment tool (Innes and Straker 1999a; Innes and Straker 1999b; Innes 2006; Legge and Burgess-Limerick 2007). In the current occupational rehabilitation climate, evidence for the use of specific assessment tools is required with appropriate decisions being based on accurate and meaningful FCEs being essential for successful occupational rehabilitation. Evidence of psychometric properties of assessment tools provides confidence to consumers and informs best practice (Gardener and McKenna 1999).

Reliability of an assessment is considered a pre-requisite for accurate measurement and is defined as the extent to which a measure is consistent, free from error and demonstrates reproducibility over time. There are different aspects to reliability; test re-rest, inter and intra rater reliability. Inter-rater reliability has been identified as being best assessed when raters are able to measure a response during a single trial,

however the use of one or more raters in a 'real life' situation may not always be possible (Portney and Watkins 2009). In the case of an FCE, space, procedure and information used to determine results, such as observation, physiological measures and client comment may impact upon the ability of more than one rater to be involved. This technique has been used to assess reliability in some FCEs (Gardener and McKenna 1999; Gross and Battie 2002). Video recording of subjects performing activities provides an opportunity for multiple observers to rate the same performance with the same information and has been used to assess inter-rater reliability for several FCEs (Smith 1994; Isernhagen et al. 1999; Reneman et al. 2002; Legge and Burgess-Limerick 2007).

This study aimed to evaluate the intra-rater and inter-rater reliability of the manual handling (lifting) component of the WorkHab FCE to provide evidence for its use in practice.

4.2.2 METHOD

Study design

This was a cross sectional study design using therapist raters who were trained and accredited in use of the WorkHab FCE. Four injured workers who underwent an FCE, agreed to have the manual handling component of the evaluation video recorded. This provided a range of manual handling (floor to bench lifts, bench to bench lifts and bench to shoulder lifts) from the different evaluations to be used in the production of a DVD. The manual handling component of each FCE was divided into lifting segments. Each lifting segment represented three lift repetitions at one weight, and a total of 35 lifting segments were included on the DVD. The lift segments were presented in random order on the DVD. Therapist raters were asked to score each of the thirty five lift segments using the WorkHab manual handling scoring system (Bradbury and Roberts 1998) and to identify if the lift was at the individual's maximum ability.

Subjects

Injured workers, who were undertaking an FCE as part of their occupational rehabilitation program in relation to return to work, were recruited from one rehabilitation provider accredited with WorkCover (NSW). Health professionals who were accredited providers of the WorkHab FCE from a variety of workplaces were invited to participate.

Instruments

WorkHab

The manual handling component of the WorkHab FCE uses a modular box system, which allows boxes to be stacked at various heights. The subject is instructed to lift the load box (initially empty) from beginning (e.g. bench) to end height (e.g. shoulder) and return. This is repeated three times before additional weight is added to the load box, until the safe maximum lift is reached. The WorkHab FCE assessor reviews the heart rate readings as the assessment progresses (objective physiological measurements). They observe the biomechanics of the lift and record reported pain and ratings of client perceived exertion (effort). The assessor records the weight lifted and calculates a manual handling score. Stance, posture, leverage, torque and pacing comprise the manual handling score, which is based on the principles of safe manual handling, with each of these components being rated on a scale of 0-4 with '0' being no adherence and '4' being the highest safety score. The sum of the score for each components is recorded as the manual handling score for each subject (Bradbury and Roberts 1998). A higher score indicates more appropriate manual handling techniques are being used.

Procedures

Following ethics approval from the University Human Research Ethics Committee, health professionals working for one rehabilitation provider discussed the research with injured workers, who were being referred for an FCE as part of their occupational rehabilitation, and provided the information sheet and consent form. Prior to commencing the FCE, each subject gave informed consent and signed a consent form for the manual handling component of the FCE to be video-taped. The FCE assessors who worked for the rehabilitation provider also gave informed consent and signed a consent form agreeing to the manual handling component of the FCE they were conducting to be video-taped.

Accredited providers of the WorkHab FCE were sent an email informing them of the research and asking for expressions of interest to be involved in rating the DVD. Those providers that responded to the researcher were sent an information sheet and consent form. Providers that consented to be part of the study were sent the first DVD for rating. Following return of the first DVD and the rating score sheet, and after an approximate 2 week time-lag, a second DVD was sent for the provider to re-evaluate. The same manual handling lift segments were on the second DVD, however these were randomised in a different sequence to those on the first DVD.

Data analysis

All data was entered into SPSS (version18.0) for analysis. Descriptive analysis, Intraclass Correlation Coefficients (ICC's), 95% confidence intervals, paired sample t-tests (intra-rater), chi-square (inter-rater) and percentage agreement were calculated where appropriate. Intra-rater reliability, the level of agreement when the same therapist viewed the same clip on two different occasions, was calculated using the ICC – model 3. This a mixed model where the rater is considered the fixed effect and the subjects are considered the random effect and is appropriate for measuring intra-rater reliability as the measurements of a single rater cannot be generalised to other raters. The ICC used to determine inter-rater reliability was Model 2 where both raters and subjects are considered random effects (Portney and Watkins 2009). An ICC of 0.90 or more was considered a measure of excellent reliability, an ICC of 0.75 – 0.90 was considered good and an ICC of less than 0.75 was considered moderate to poor (Innes 1999b, Legge and Burgess-Limerick 2007, Portney and Watkins 2009).

Analysis was conducted on data for each of the manual handling components, stance, posture, leverage, torque and pace, (using a mean of individual raters' scores for each component) for each individual worker.

Sample size was calculated using the Confidence Intervals for Proportions - the Bayesian Method using Beta Distribution, using the lifts from four injured workers based on the number of lift segments included in each FCE.

4.2.3 RESULTS

Participants

Four injured workers who were undergoing a WorkHab FCE as part of their occupational rehabilitation program consented to have the manual handling component of the FCE video-taped. Seventeen accredited providers of the WorkHab FCE completed the inter-rater reliability component of the study and 14 completed the follow up evaluation for the intra-rater reliability.

Eighty eight percent of the study sample were occupational therapists (n= 15) and 12% physiotherapists (n=2). The mean number of years of professional experience was 11.2 years (SD: 7.3; Range: 2-27 years); the mean number of years of FCE experience was 8 years (SD: 6.1; Range: 1-20 years) and the mean number of years conducting WorkHab FCEs was 4.7 years (SD: 2.7; Range: 1-9 years).

Intra-rater agreement

The overall intra-rater agreement (n=14) for the manual handling score and for each of the manual handling components (stance, posture, leverage, torque and pace) showed excellent reliability (ICC: 0.91 – 0.97), (see table 1).

Table 1: Intra-rater agreement: ICC for Manual Handling components (N=14).

Manual Handling Component	ICC[a]	95% CI[b] of ICC	Interpretation of ICC
Manual Handling Score	0.97	0.94 : 0.99	Excellent
Stance	0.97	0.93 : 0.98	Excellent
Posture	0.91	0.82 : 0.96	Excellent
Leverage	0.97	0.94 : 0.99	Excellent
Torque	0.93	0.86 : 0.97	Excellent
Pace	0.96	0.91 : 0.98	Excellent
Safe Maximal Lift	0.81	0.58 : 0.94	Good

[a] ICC = Intraclass Correlation Coefficient (two way mixed).

[b] 95% CI = 95% Confidence interval.

Fifty two percent of all ratings were identical and 87% were within one score of each other on the 0-4 point scoring system.

Intra-rater percentage agreement for determining whether a lift was a safe maximal lift was moderate with 70% of all ratings being identical. The ICC for safe maximal lift showed a good level of reliability (ICC: 0.81).

Results calculated for each of the four injured workers (subjects) provides further information of the intra-rater agreement for the manual handling score with results showing good agreement (ICC:0.79 – 0.88), as seen in table 2.

Table 2: Intra-rater agreement: Manual Handling Score Results: means, difference, standard deviations 95% confidence intervals and ICC for each injured worker (subject). (N=14)

Manual Handling Component	Mean 1a (SD 1)	Mean 2[b] (SD 2)	Mean Difference (SD difference)	95% CI[c] of difference	ICC[d] (95% CI of ICC)	Interpretation of ICC
Injured worker 1 (n=8 lifts)	16.94 (1.61)	16.5 (2.01)	0.44 (1.16)	-0.24 to 1.11	0.88 (0.64 to 0.96)	Good
Injured worker 2 (n=8 lifts)	13.39 (2.14)	16.79 (1.87)	-3.4* (1.32)	-4.16 to -2.64	0.88 (0.62 to 0.96)	Good
Injured worker 3 (n=9 lifts)	18.03 (1.36)	16.63 (1.99)	-0.91* (1.40)	0.58 to 2.22	0.79 (0.34 to 0.93)	Good
Injured worker 4 (n=10 lifts)	16.47 (1.79)	16.51 (2.05)	-0.02 (1.31)	-0.78 to 0.73	0.87 (0.58 to 0.95)	Good

[a] Mean 1 = mean manual handling score from DVD 1

[b] Mean 2 = mean manual handling score from DVD 2

[c]95% CI = 95% Confidence interval.

[d]ICC = Intraclass correlation coefficient (two way mixed)

*Significant (two tailed) at p<0.05 (paired sample t-test)

Results for the individual manual handling components ranged from poor to excellent (table 3). There were significant differences ($p<0.05$) in the posture and leverage ratings (means) for subject 2 and in the torque rating for subject 3. Both subjects 2 and 3 also had significant differences in the manual handling score. Four of the raters (23%) scored the manual handling components for subject 2 significantly higher ($p<0.05$) at the second rating and 2 of the raters did so for subject 3.

Table 3: Intra-rater agreement-Manual Handling Component Results: mean difference (time1: time 2), standard deviation, 95% confidence intervals and ICC for each injured worker (subjects) (N=14).

	Stance		Posture		Leverage		Torque		Pace	
	Mean dif[a] (SD[c]) (95% CI[b] of difference)	ICC[d] (95%CI)	Mean dif[a] (SD[c]) (95% CI[b] of difference)	ICC[d] (95% CI)	Mean dif[a] (SD[c]) (95% CI[b] of difference)	ICC[d] (95% CI)	Mean dif[a] (SD[c]) (95% CI[b] of difference)	ICC[d] (95% CI)	Mean dif[a] (SD[c]) (95% CI[b] of difference)	ICC[d] (95% CI)
Injured worker 1 (n=8 lifts)	0.12 (0.35) (-0.07 to 0.32)	0.91 (0.71 to 0.97)	0.12 (0.32) (-0.06 to 0.31)	0.85 (0.55 to 0.95)	0.08 (0.32) (-0.10 to 0.26)	0.87 (0.59 to 0.95)	-0.07 (0.31) (-0.25 to 0.10)	0.79 (0.35 to 0.93	0.18 (0.27) (0.01 to 0.33)	0.84 (0.51 to 0.95)
Injured worker 2 (n=8 lifts)	-1.06 (0.51)* (-1.35 to-0.76)	0.81 (0.39 to 0.94)	-1.05 (0.52)* (-1.35 to-0.75)	0.3 (-1.19 to 0.77)	-0.81 (0.42)* (-1.05 to-0.56)	0.84 (0.49 to 0.94)	-0.23 (0.49) (-0.51 to 0.05)	0.70 (0.06 to 0.90)	0.24 (0.31) (-0.4 to -0.05)	0.84 (0.51 to 0.95)
Injured worker 3 (n=9 lifts)	-0.08 (0.41) (-0.32 to 0.15)	0.88 (0.64 to 0.96)	0.32 (0.45) (0.06 to 0.58)	0.52 (-0.48 to 0.84)	0.22 (0.37) (0.01 to 0.43)	0.85 (0.52 to 0.95)	0.61 (0.37)* (0.39 to 0.82)	0.46 (-0.67 to 0.83)	0.33 (0.28)* (0.16 to 0.5)	0.79 (0.36 to 0.93)
Injured worker 4 (n=10 lifts)	0.12 (0.44) (-0.12 to 0.38)	0.87 (0.76 to 0.94)	0.08 (0.45) (-0.17 to 0.35)	0.83 (0.68 to 0.93)	0.11 (0.27) (-0.05 to 0.26)	0.87 (0.75 to 0.94)	-0.16 (0.27) (-0.32 to-0.01)	0.82 (0.66 to 0.92)	-0.18 (0.33) (-0.38 to 0.1)	0.8 (0.62 to 0.92)

[a] Mean difference = mean difference of score in DVD 1 and DVD 2

[b]95% CI of difference – 95% Confidence interval of the mean difference

[c]SD = Standard deviation

[d]ICC = Intraclass correlation coefficient (two way mixed).

*Significant (two tailed) at p<0.05 (paired sample t-test)

Inter-rater agreement

Overall inter-rater agreement (n=17) for the manual handling score and for each of the manual handling components ranged from good to excellent with ICC's ranging from 0.77 to 0.91 (See table 4).

Table 4: Inter-rater agreement: ICC for Manual Handling components (N=17).

Manual Handling Component	ICC[a]	95% CI[b] of ICC	Interpretation of ICC
Manual Handling Score	0.90	0.79 : 0.96	Excellent
Stance	0.91	0.81 : 0.96	Excellent
Posture	0.79	0.58 : 0.92	Good
Leverage	0.91	0.81 : 0.96	Good
Torque	0.77	0.53 : 0.91	Good
Pace	0.82	0.63 : 0.93	Good
Safe Maximal Lift	0.9	0.80 : 0.96	Excellent

[a] ICC = Intraclass Correlation Coefficient (two way random).

[b] 95% CI = 95% Confidence interval.

Inter-rater percentage agreement of scoring for the safe maximal lift was identical in 68% of cases, and the ICC for determining a safe maximal lift was 0.9, indicating an excellent level of reliability.

The inter-rater agreement for each subject ranged from good to excellent (table 5). No significant difference between raters was found using chi-square analysis for each of the manual handling components.

Table 5. Inter-rater agreement for manual handling components for each injured worker (N=17).

	Stance		Posture		Leverage		Torque		Pace	
	Mean [a] (SD[c]) 95% CI[b] of mean	ICC[d] (95%CI)	Mean [a] (SD[c]) 95% CI[b] of mean	ICC[d] (95% CI)	Mean [a] (SD[c]) 95% CI[b] of mean	ICC[d] (95% CI)	Mean [a] (SD[c]) 95% CI[b] of mean	ICC[d] (95% CI)	Mean [a] (SD[c]) 95% CI[b] of mean	ICC[d] (95% CI)
Injured worker 1 (n=8 lifts)	3.21 0.62 2.88 :3.52	0.91 (0.82 to 0.96)	3.13 0.47 2.88 :3.37	0.84 (0.69 to 0.93)	3.32 0.39 3.11: 3.51	0.70 (0.42 to 0.87)	3.67 0.34 3.49 :3.85	0.84 (0.7 to 0.93)	3.75 0.27 3.61 :3.89	0.76 (0.53 to 0.89)
Injured worker 2 (n=10 lifts)	2.26 0.69 1.91 :2.62	0.92 (0.84 to 0.96)	2.30 0.46 2.06: 2.53	0.74 (0.49 to 0.89)	2.55 0.60 2.24 :2.86	0.85 (0.71 to 0.94)	3.08 0.58 2.77: 3.38	0.83 (0.68 to 0.93)	3.38 0.42 3.36 :3.40	0.81 (0.63 to 0.92)
Injured worker 3 (n=8 lifts)	3.22 0.68 2.87:3.58	0.92 (0.85 to 0.96)	3.56 0.37 3.36: 3.75	0.79 (0.59 to 0.91)	3.55 0.43 3.33 :3.77	0.83 (0.67 to 0.93)	3.87 0.22 3.76 :3.98	0.78 (0.56 to 0.91)	3.8 0.26 3.66 :3.93	0.85 (0.71 to 0.94)
Injured worker 4 (n=9 lifts)	3.21 0.50 2.95 :3.46	0.87 (0.75 to 0.95)	3.19 0.42 2.97: 3.41	0.83 (0.68 to 0.93)	3.37 0.45 3.14 :3.61	0.87 (0.75 to 0.95)	3.42 0.33 3.25 :3.60	0.82 (0.66 to 0.93)	3.28 0.38 3.08 :3.48	0.80 (0.62 to 0.92)

[a] Mean = Mean score for each of the manual handling components.

[b] Standard deviation = standard deviation of mean score

[c]95% CI = 95% Confidence interval of mean scores.

[d]ICC = Intraclass correlation coefficient (two way random).

4.2.4 DISCUSSION

Reliability of assessment tools has been identified as important for assessing the abilities of injured individuals (Innes and Straker 1999a). This study investigated the intra and inter-rater reliability of the manual handling components of the WorkHab FCE. The intraclass correlation coefficient measures the agreement between pairs of observations, and is accepted as a measure of reliability in relation to the discriminative capacity of a test (Portney and Watkins 2009).

Intra-rater reliability findings

The intra-rater agreement in this study, as shown with ICC values and percentage agreement, indicates that health professionals can make consistent judgements on the manual handling scoring system and when determining safe maximal lifts during a WorkHab FCE. When looking at the results for individual injured workers (subjects) the ICC values for the manual handling score (calculated from the sum of the components, stance, posture, leverage, torque and pace) was good. The different manual handling components; stance, leverage and pace also had good to excellent ICC values. Posture and torque, however, when analysed for each subject, had poor to good ICC values suggesting this was less reliably scored.

Posture is scored according to the maintenance of the normal lordosis throughout the assessment, with non maintenance resulting in lower scores; torque is scored according to the amount of rotation relative to the pelvis, particularly looking at spinal twisting in the low back (Bradbury and Roberts 1998). Raters in this study saw video-footage of lift segments from two angles (rear and side) however did not have the three dimensional vision that would be present in the clinical setting and this may have impacted upon the scoring of these components. Another reason for this difference could be due to the operational definitions for posture and torque. This may not be as detailed or clear to assessors as the operational definitions of other components of the manual handling scoring system. Use of operational definitions has been shown to improve reliability between and within raters (Gardener and McKenna 1999). Another possibility is that posture or torque is more complex and therefore more difficult to assess. It should be noted, however that difference in the posture scoring component did not reflect negatively upon the reliability of the overall manual handling score, nor did the difference in the torque score. The manual handling score is the aspect used most in clinical practice, rather than the individual components.

The intra-rater agreement for individual injured workers (subjects) for each of the manual handling components showed that injured worker (subject) 2 was rated

significantly higher at the re-assessment for stance, posture and leverage suggesting raters perceived a better manual handling technique was being employed on the second viewing of the DVD. These results were influenced by significant increases in ratings by 4 individual raters. Injured worker (subject) 3 however was rated significantly lower at the re-assessment, specifically for torque suggesting raters perceived a less good lifting technique was being employed at the second viewing of the DVD. Experienced providers of the WorkHab FCE were rating one testing occasion for each subject at two different times. When considered across the 14 ratings, the overall manual handling score for subjects 2 and 3 however, had good ICC values.

These findings demonstrate that the WorkHab FCE is a reliable measure when the same person acts as the assessor. This supports the findings of earlier research which found the test-retest reliability of the WorkHab FCE was substantial (James et al. 2010). Similar types of lifting tasks evaluated in other FCEs such as Isernhagen, Job Fit, Physical work performance evaluation and Ergo-kit have reported substantial or acceptable levels of intra-rater reliability (Reneman et al. 2002; Durand et al. 2004; Gouttebarge et al. 2005; Legge and Burgess-Limerick 2007), however direct comparison is not possible due to differences in testing procedures and measures of reliability used.

Evaluation of safe maximal lifts

Safe maximal lifts are commonly determined during FCEs to enable guidelines to be developed for individuals for return to work safely. Raters in our study identified if each lift segment was a safe maximal lift or not. The intra-rater agreement for determining safe maximal lift was good (ICC: 0.81), with percentage agreement being 70% (moderate). The inter-rater agreement for determining safe maximal lift was similar with an ICC of 0.90 (excellent) and percentage agreement of 68% which is moderate. Isernhagen (Isernhagen et al. 1999) used video analysis of lift segments to assess reliability of light, moderate and heavy lifts and found higher levels of intra and inter-rater reliability when analysing just light and heavy lifts, using a kappa statistic. In this study of the WorkHab FCE, raters were blinded to the sequence of the lift, and therefore were unable to consider the effects on lifting technique as weight increased. The WorkHab uses physiological and kinematic performance measures, including observation of biomechanics, for assessors to determine safe maximal lifts, and viewing the lifts out of sequence may have removed biomechanical cues typically used in the clinical setting. This may have impacted upon these results. Further investigation of rater determination of safe maximal lift during the WorkHab FCE is recommended.

Inter-rater reliability findings

The overall inter-rater reliability for the manual handling score and for each of its components (stance, posture, leverage, torque and pace) were good to excellent (ICC: 0.77 – 0.91). Confidence intervals for the manual handling score, stance, leverage and safe maximal lift were narrow suggesting high precision of these results (Portney and Watkins 2009). Although the ICC values for posture and torque were good, the confidence intervals were larger suggesting more variation in these results between raters. These findings are in line with other studies investigating inter-rater reliability of FCEs (Lechner et al. 1994; Gross and Battie 2002; Durand et al. 2004; Gouttebarge et al. 2006; Legge and Burgess-Limerick 2007), although due to differences in testing protocols and reliability measures used, direct comparisons between these results and other published research is not possible.

The inter-rater reliability for each individual injured worker showed a range of ICC values indicating good to excellent reliability for stance and pace (0.76 to 0.92) and moderate to good reliability for posture, leverage, and torque (0.70 to 0.87). As with intra-rater reliability, posture and torque was less reliably scored and further research to investigate the clinical reasoning used in determining scores is recommended.

Relationship of findings to clinical practice

In this study, the manual handling component of the FCE for the injured worker was divided into lift segments, with each lift segment representing three lift repetitions at the one weight. In clinical practice, the score for the manual handling components is determined after observing the whole lift – from minimum to safe maximum weight (Bradbury and Roberts 1998). The lift segments in our DVD were also randomised therefore raters were seeing each lift segment in isolation rather than as part of a whole lift. Randomising the lift segments enabled an objective evaluation of each of the manual handling components using just the information provided, without reference to what had occurred before or after the lift for each subject. This added rigour to the study, although the process was not clinically realistic. In clinical practice other information is available to the assessor such as verbal cues, facial expressions, and knowledge about the subject in regards to type of injury, type of job or reason for completing the FCE. Clinical reasoning skills used by therapists (raters) include gathering information about different aspects of the situation, perceiving and interpreting the cues and then on the basis of this information and of relevant (therapist) knowledge making a judgement or assessment (Hagedorn 2001). In clinical practice, raters will therefore have more information to guide the clinical reasoning used in the determination of the manual handling component score. The use of

pragmatic clinical reasoning by therapists conducting FCEs involves evaluating if an action is feasible in a given situation taking account of the context, resources, therapist knowledge, skills and interests and the wider organisational, socio-cultural and political considerations (James and Mackenzie 2009a). Investigation of the nature and processes of clinical reasoning when conducting an FCE is an area that deserves further exploration, particularly investigating the aspects of clinical reasoning used to determine ratings for the different components of manual handling within the WorkHab FCE. Studies that have used real time to evaluate rater agreement (Lechner et al. 1994; Gross and Battie 2002; Durand et al. 2004; Gouttebarge et al. 2006) (where this clinical information was available) and studies that used video recordings (where this information was not available) to evaluate rater agreement (Isernhagen et al. 1999; Reneman et al. 2002; Legge and Burgess-Limerick 2007) have however, both found substantial or acceptable levels of inter-rater agreement.

This study had 17 raters in the inter rater reliability component and 14 raters in the intra rater reliability component which is more than some previous studies of inter-rater reliability in FCEs (Reneman et al. 2002; Gouttebarge et al. 2006; Legge and Burgess-Limerick 2007). As more raters are included in a study, the chance of variation in ratings increases. Therefore, these findings indicate that the WorkHab FCE is a reliable measure to assess manual handling because different therapist-raters provide consistent ratings when assessing injured workers. Health professionals need to evaluate all the attributes of FCEs which includes safety, reliability, validity, practicality and utility, to ensure high quality standards of practice. Further studies investigating the validity of the WorkHab FCE are recommended.

4.2.5 CONCLUSION

This research investigated the intra-rater and inter-rater reliability of the manual handling scoring system, and safe maximal lift determination of the WorkHab FCE and included floor to bench, bench to bench and bench to shoulder lifts. The results found substantial levels of intra-rater agreement as shown with both ICC and percentage agreement results. The Inter-rater reliability results also showed substantial levels of reliability. In previous research it has been suggested that therapists are basing decisions on historical information and practices rather than evidence based practice and current research. The results of this research contribute to the growing evidence of FCEs in practice and will provide therapists with information to guide practice.

ACKNOWLEDGEMENTS

The authors would like to thank the injured workers who consented to have the manual handling component of the WorkHab FCE they were completing to be video-taped. Thanks go also to the WorkHab accredited providers who volunteered to participate in this study. This study was made possible with the support of a University of Newcastle grant (No GO187308).

Declaration of interest: The authors report no conflict of interest. The authors are responsible for the content and writing of this paper.

REFERENCES

Boadella, J., Sluiter, J.& Frings-Dresen, M. (2003). "Reliability of Upper Extremity Tests Measured by the Ergos Work Simulator: A Pilot Study." Journal of Occupational Rehabilitation **13**(4): 219-232.

Bradbury, S.& Roberts, D. (1998). Workhab Functional Capacity Evaluation Procedural Manual.WorkHab Australia.

Cotton, A., Schonstein, E.& Adams, R. (2006). "Use of Functional Capacity Evaluations by Rehabilitation Providers in NSW." Work **26**(3): 287-295.

Durand, M., Brassard, B., Hong, Q., Lemaire, J.& Loisel, P. (2008). "Responsiveness of the Physical Work Performance Evaluation, a Functional Capacity Evaluation, in Patients with Low Back Pain." Journal of Occupational Rehabilitation **18**(1): 58-67.

Durand, M., Loisel, P., Poitras, S., Mercier, R., Stock, S.& Lemaire, J. (2004). "The Interrater Reliability of a Functional Capacity Evaluation: The Physical Work Performance Evaluation." Journal of Occupational Rehabilitation **14**(2): 119-129.

Gardener, L.& McKenna, K. (1999). "Reliability of Occupational Therapists in Determining Safe Maximal Lifting Capacity." Australian Occupational Therapy Journal **46**(3): 110-19.

Gibson, L.& Strong, J. (1997). "A Review of Functional Capacity Evaluation Practice." Work **9**(1): 3-11.

Gouttebarge, V., Wind, H., Kuijer, P.& Frings-Dresen, M. (2004). "Reliability and Validity of Functional Capacity Evalutation Methods: A Systematic Review with Reference to Blankenship, Ergos Work Simulator, Ergo-Kit and Isernhagen Work System." International Archives of Occupational and Environmental Health **77**(8): 527-537.

Gouttebarge, V., Wind, H., Kuijer, P. P., Sluiter, J. K.& Frings-Dresen, M. H. (2005). "Intra- and Interrater Reliability of the Ergo-Kit Functional Capacity Evaluation

Method in Adults without Musculoskeletal Complaints." Archives of Physical Medicine and Rehabilitation **86**(12): 2354-2360.

Gouttebarge, V., Wind, H., Kuijer, P. P., Sluiter, J. K.& Frings-Dresen, M. H. (2006). "Reliability and Agreement of 5 Ergo-Kit Functional Capacity Evaluation Lifting Tests in Subjects with Low Back Pain." Archives of Physical Medicine and Rehabilitation **87**(10): 1365-1370.

Gross, D.& Battie, M. (2002). "Reliability of Safe Maximum Lifting Determinations of a Functional Capacity Evaluation." Physical Therapy **82**(4): 364-372.

Hagedorn, R. (2001). Foundations for Practice in Occupational Therapy. London, Churchill Livingstone.

Innes, E. (2006). "Reliability and Validity of Functional Capacity Evaluations: An Update." International Journal of Disability Management Research **1**(1): 135-148.

Innes, E.& Straker, L. (1999a). "Reliability of Work-Related Assessments." Work **13**(2): 107-124.

Innes, E.& Straker, L. (1999b). "Validity of Work-Related Assessments." Work **13**(2): 125-152.

Isernhagen, S., Hart, D.& Matheson, L. (1999). "Reliability of Independent Observer Judgments of Level of Lift Effort in a Kinesiophysical Functional Capacity Evaluation." Work **12**(2): 145-150.

James, C.& Mackenzie, L. (2009a). "The Clinical Utility of Functional Capacity Evaluations: The Opinion of Health Professionals Working within Occupational Rehabilitation." Work **33**(3): 231-9.

James, C.& Mackenzie, L. (2009b). "Health Professional's Perceptions and Practices in Relation to Functional Capacity Evaluations: Results of a Quantitative Survey." Journal of Occupational Rehabilitation **19**(2): 203-211.

James, C., Mackenzie, L.& Capra, M. (2010). "Test Retest Reliability of the Manual Handling Component of the Workhab Functional Capacity Evaluation in Healthy Adults." Disability & Rehabilitation **32**(22): 1863-1869.

Jay, M., Lamb, J., Watson, R., Young, I., Fearon, F., Alday, J.& Tindall, A. (2000). "Sensitivity and Specificity of the Indicators of Sincere Effort of the EPIC Lift Capacity Test on a Previously Injured Population." Spine **25**(11): 1405-12.

Jones, T.& Kumar, S. (2003). "Functional Capacity Evaluation of Manual Materials Handlers: A Review." Disability and Rehabilitation **25**(4-5): 179-191.

Jundt, J.& King, P. (1999). "Work Rehabilitation Programs: A 1997 Survey." Work **12**(2): 139-144.

King, P., Tuckwell, N.& Barrett, T. (1998). "A Critical Review of Functional Capacity Evaluations." Physical Therapy **78**(8): 852.

Lechner, D., Jackson, J., Roth, D.& Straaton, K. (1994). "Reliability and Validity of a Newly Developed Test of Physical Work Performance." Journal of Occupational Medicine **36**(9): 997-1004.

Lee, G., Chan, C.& Hui-Chan, C. (2001). "Work Profile and Functional Capacity of Formwork Carpenters at Construction Sites." Disability and Rehabilitation **23**(1): 9-14.

Legge, J.& Burgess-Limerick, R. (2007). "Reliability of the Jobfit System Pre-Employment Functional Assessment Tool." Work **28**(4): 299-312.

Matheson, L., Rogers, L., Kaskutas, V.& Dakos, M. (2002). "Reliability and Reactivity of Three New Functional Assessment Measures." Work **18**(1): 41-50.

Portney, L.& Watkins, M. (2009). Foundations of Clinical Research: Applications to Practice. Upper Saddle River, New Jersey, Pearson Prentice Hall.

Reneman, M., Brouwer, S., Meinema, A., Dijkstra, P., Geertzen, J.& Groothoff, J. (2004). "Test-Retest Reliability of the Isernhagen Work Systems Functional Capacity Evaluation in Healthy Adults." Journal of Occupational Rehabilitation **14**(4): 295-305.

Reneman, M., Jaegar, S., Westmaas, M.& Goeken, L. (2002). "The Reliability of Determining Effort Level of Lifting and Carrying in a Functional Capacity Evaluation." Work **18**(1): 23-27.

Smith, R. (1994). "Therapist's Ability to Identify Safe Maximal Lifting in Low Back Pain Patients During Functional Capacity Evaluation." Journal of Orthopaedic and Sports Physical Therapy **19**(5): 227-281.

Ting, W., Wessel, J., Brintnell, S., Maikala, R.& Bhambhani, Y. (2001). "Validity of the Baltimore Therapeutic Equipment Work Simulator in the Measurement of Lifting Endurance in Healthy Men." American Journal of Occupational Therapy **55**(2): 184-90.

4.3 Chapter Summary

The importance of health professionals using reliable and valid assessment tools has been highlighted in the literature (Portney and Watkins 2009) and reliability of assessment tools provides health professionals and other stakeholders with confidence that the measurements are accurate and meaningful. One study in this chapter has demonstrated substantial levels of test-retest reliability of the manual handling components of the WorkHab FCE in healthy adults. The intra-class correlation coefficient values were high (ICC:0.90-0.92); however, the 95% limits of agreement showed variations of ± 3-4kg in weight lifted. However, this was within ±2.5kg in the majority of cases, which is clinically more acceptable. This highlights the importance of health professionals utilising clinical judgement skills to interpret findings from the WorkHab FCE taking into consideration aspects that may contribute to variation such as specific tasks being assessed, aspects of the job and workplace issues. These results are comparable and in some instances better than results on other specific FCEs that have reported acceptable levels of rest -retest reliability (Reneman et al. 2004; Gouttebarge et al. 2006; Legge and Burgess-Limerick 2007).

The manual handling scoring scale of the WorkHab FCE showed good internal consistency (α=0.92), which indicates the items within this scale measure the same construct of manual handling technique. However, moderate inter-rater reliability was demonstrated for the overall manual handling score (ICC:0.74), with higher scores recorded in the re-test assessment. Clinically, an improvement in the manual handling score is positive because this indicates the

subject has learned and is utilising suitable lifting techniques. As discussed in the manuscript, the convenience sample of healthy university staff and students may have increased the likelihood of participants gaining skills in lifting more safely and thus impacting upon these results.

The improvements in manual handling score in the re-test did not impact upon the amount of weight lifted, as might be expected. There was a slight reduction in the mean weights lifted in the re-test however the standard deviation at the re-test was also smaller, which suggests a smaller range of weight being lifted. Overall this further reinforces the importance of clinical reasoning as part of safe maximal lift determination with health professionals considering all aspects of manual handling.

Overall study results demonstrated substantial levels of inter-rater and intra-rater reliability of the manual handling scoring system of the WorkHab FCE for the bench to bench, bench to shoulder, and floor to bench lifts and of the safe maximal lift determination for these lifts. These results are similar to those reported in other studies that have found substantial inter-rater and intra rater reliability in specific FCEs (Reneman et al. 2002b; Durand et al. 2004; Gouttebarge et al. 2005; Legge and Burgess-Limerick 2007).

Of the five manual handling scoring components (stance, posture, leverage, torque and pace), posture and torque showed greatest variation in both intra-rater and inter-rater results both for individual injured workers and in the type of lift. This suggests that further research focussed on the clinical reasoning used to determine scores for these principles of safe manual handling is required.

These studies provide evidence to clinicians that the administration procedures for the WorkHab FCE are dependable. Test re-test reliability in healthy adults was substantial. The WorkHab FCE was a reliable measure when the same person acts as assessor (intra-rater reliability) and a reliable measure to assess manual handling with different assessors providing consistent ratings in the assessment of injured workers (inter-rater reliability). Evidence of reliability allows confidence in the data collected and assists in ensuring high quality standards of practice when using the WorkHab FCE.

The next chapter discusses aspects of validity of the WorkHab FCE.

Chapter 5 The WorkHab Functional Capacity Evaluation: Validity

Overview: This chapter discusses validity as it relates to FCEs in general and presents the findings of two studies that investigated aspects of validity of the WorkHab FCE. These studies are presented in two manuscripts, the first of which reports on content validity of the WorkHab FCE and the second discusses construct validity in relation the manual handling component of the WorkHab FCE.

Validity also relates to the accuracy of an assessment and is concerned with the extent or degree to which an instrument measures what it is intended to measure (Innes and Straker 1999b; Portney and Watkins 2009). Validity suggests that an assessment is free from error and therefore implies that reliability has been established. A test can demonstrate reliability without necessarily having validity: for instance, an invalid assessment can be reliable. Reliability is therefore a pre-requisite to validity. However, the reliability of an assessment will also set the limits of validity in that an assessment with poor reliability or one that has erratic, inconsistent or imprecise results will have low validity. Conversely an assessment with strong reliability does not necessarily have strong validity (Schneider et al. 2007; Portney and Watkins 2009). Raters may agree on results in a simple test that indicates the test is reliable; however, the test itself may not relate to practice or the construct of interest and therefore is not valid. Validity needs to be related to practice that is intrinsically a complex environment for assessment. Evidence for validity of assessment tools requires the establishment of evidence about both the content of the instrument and

evidence of relationships between the measurement and other variables. Validity is therefore more difficult to establish than reliability, and raters may have more variety in their ratings if the complexity of practice is not accounted for within a measure (Schneider et al. 2007).

Test instruments normally examine issues such as whether results are able to predict future changes and the direction of these changes; whether results are capable of discriminating among individuals with and without particular variables; and also whether the test can evaluate change in a variable between assessment instances (Bruce et al. 2008; Portney and Watkins 2009). To this end, there are different types of validity that need to be addressed. Validity needs to be established for the use of the assessment in context such that validity can be established for a specific purpose related to practice (Schneider et al. 2007; Portney and Watkins 2009). The types of validity are: face validity, content validity, criterion related validity (concurrent validity and predictive validity) and construct validity.

Face validity indicates that an instrument appears to test what it is supposed to and is identified as the weakest form of validity. Face validity can be easily established for some assessments that measure the property of interest through direct observation. Face validity can be regarded as an all or none level of validity that is subjective and scientifically weak (Portney and Watkins 2009). Face validity suggests the assessment is acceptable to those who administer it, to those who are being tested or to those who will use the results, and is therefore a useful property of a test.

Content validity relates to the degree to which the items in an assessment seem to relate to the construct or content domain that is intended to be measured by the tool. Content validity is often evaluated with the use of an expert panel who review the instrument to determine if the content is appropriate and this can therefore be a subjective process (McKenzie et al. 1999; Portney and Watkins 2009). For FCEs, one approach used has been to examine how well the FCE covers the physical demands of the Dictionary of Occupational Titles taxonomy, or DOT, which is a US Department of Labor government publication (U.S.Department of Labor 1965; Lee and Chan 2003; Kersnovske et al. 2005). Many FCEs have been based on the DOT taxonomy and studies suggest content validity for these FCEs is deemed sufficient when they incorporate the various components of the DOT taxonomy (Fishbain et al. 1994; Lechner et al. 1994; Innes and Straker 1999b; Gibson et al. 2005). Examples of other FCEs where content validity has been established include the Gibson Approach to Functional Capacity Evaluation (GAPP-FCE), which was assessed using an expert panel with good results (Gibson and Strong 2002; Gibson et al. 2005). Content validity of the Assessment of Work Performance (AWP) tool was also studied using an expert panel with high levels of agreement identified for most aspects (Sandqvist et al. 2008).

Another approach to content validity evaluation is to develop FCE methodology for use with specific groups using literature and relevant research. This approach has been used to develop FCE methodology for work-related upper limb disorders (Reneman et al. 2005) and to develop FCE methodology for those with neck disorders (Reesink et al. 2007).

Criterion related validity indicates that the outcome of one assessment can be used as a substitute measure for an established reference standard criterion test or gold standard. Criterion related validity has been identified as the most practical and objective approach to validity testing (Bowling 2004; Portney and Watkins 2009); however, in the case of FCEs, no gold standard has been established for comparison (Gouttebarge et al. 2004; Wind et al. 2005). Criterion related validity can be separated into concurrent validity (the degree to which the outcomes of one assessment correlate with the outcomes of another when given at the same time) and predictive validity (the degree to which the assessment can be used to predict future performance) (Portney and Watkins 2009). Studies have demonstrated concurrent validity of specific FCEs with a variety of tools. The Ergo-kit FCE and the Instrument for Disability Risk (IDR) were studied in relation to the construction industry: results indicated concurrent validity between the isometric lifting tests and the IDR were poor, and results between the dynamic lifting tests and the IDR were moderate (Gouttebarge et al. 2009a). Tests on the Baltimore Therapeutic Equipment Work Simulator found that it overestimates lifting endurance in healthy men (Ting et al. 2001), and differences between safe lifting limits were also reported when comparing FCE results and the NIOSH guidelines (Kuijer et al. 2006b).

Studies that investigated the predictive validity of different FCEs are also reported in the literature. The recommendations from the Physical Work Performance Evaluation FCE were found to have substantial levels of agreement in relation to the prediction of return to work and the performance of actual work tasks at discharge, at three months and at six months (Lechner et al. 2008). The carrying lifting strength test of the Ergo-kit FCE showed a

moderate level of predictive validity on the IDR; however, other dynamic lifting tests of the Ergo-kit FCE had poor predictive validity (Gouttebarge et al. 2009a). In a study investigating predictive validity using the IWS with workers compensation claimants with upper limb disorders, better FCE performance was a weak predictor of faster benefit suspension (Gross and Battie 2006). The IWS was not related to recurrent back problems, self-reported disability or future pain in another study (Gross and Battie 2005b). For patients with chronic low back pain, better performance on the IWS was weakly predictive of faster recovery and performance on the floor to waist lift was as predictive as the whole assessment (Gross et al. 2004). In another study, better FCE performance was linked with a higher risk of recurrence of injury (Gross and Battie 2004).

The predictive validity of individual tasks commonly used as part of FCEs has also been reported. For instance, increasing the weight lifted during a floor to waist lift has been associated with a higher likelihood of return to work after six months for clients with chronic low back pain (Matheson et al. 2002a). In a study using the DOT components of an FCE, performance of the tasks of stooping, climbing, balancing, crouching, feeling, handling, lifting and carrying was associated with a higher chance of employment 30 months after treatment at a pain clinic when pain levels were also reported as low, which indicates elements of predictive validity (Fishbain et al. 1999).

Construct validity is the degree to which a theoretical construct is measured by an assessment. Construct validity can be separated into convergent validity (two measures believed to reflect the same phenomenon will be highly correlated) and discriminant validity (different results are expected from

measures that assess different characteristics) (Innes and Straker 1999b; Portney and Watkins 2009). Part of the establishment of construct validity is dependent upon content validity such that the content domain must be defined and representative of the construct to be measured. The theoretical context of constructs must also be established so that it can be assumed individuals with a particular condition would behave in a certain way under particular conditions. Therefore, construct validity provides evidence to support the theoretical framework behind the construct (Portney and Watkins 2009).

One study investigated the construct validity of lifting tests within the Ergo-kit FCE with construction workers. Poor results were identified with none of the lifting tests able to significantly discriminate between two groups of construction workers (one with high risk musculoskeletal injury risk and one with low risk). Convergent validity with self-reported pain measures was also poor (Gouttebarge et al. 2009b). In a study that investigated the IWS FCE, a moderate relationship between the IWS FCE and the Pain Disability Index (PDI) was found that supports construct validity for this tool as a functional measure (Gross and Battie 2003).

This chapter discusses content validity of the WorkHab FCE and construct validity of the safe maximal lift during the bench to shoulder lift of the WorkHab FCE. Content validity was studied using a cross sectional survey of health professionals who conduct FCEs in practice. The aim of this study was to determine the content validity of the WorkHab FCE by investigating the relevance of the FCE components to different work types, the difficulty of the FCE components to various client types and the relationship between the FCE components and the physical demands of the DOT with health professionals

(experts) who use FCEs in practice. As many FCEs are based on the DOT taxonomy, this survey included any health professional who conducts FCEs in practice and was not limited to those health practitioners accredited to conduct the WorkHab FCE specifically. This provided a larger population to sample. Those health professionals who are eligible to conduct FCEs within the Australian occupational rehabilitation were targeted (Occupational therapists, Physiotherapists and Exercise physiologists) and the focus of the survey was on musculoskeletal injuries. The results are discussed in **Manuscript 6 (5.1):** *"Content validity and the WorkHab FCE"* submitted for publication in Disability and Rehabilitation. The results of this study support content validity for the WorkHab FCE, specifically in relation to manual work and for vocational retraining purposes; however, further research into the relevance of some FCE items and the application of the WorkHab items to sedentary work is recommended.

Construct validity was assessed using an experimental study design. This study aimed to quantify and analyse muscle function with a view to establishing a physiological indicator for the safe maximal lift (SML) during the bench to shoulder lift of the WorkHab FCE. This laboratory based experimental study with healthy subjects aimed to determine if any change in muscle physiology or change in relationship between the physiology of the muscles in the upper limb and shoulder girdle for the bench to shoulder lift, as recorded by surface electromyography (SEMG), correlated with the safe maximal lift as determined by WorkHab evaluators. The determination of an end point of safe lifting has been the subject of FCE discussion and is one of the difficulties of FCEs identified by Innes and Straker (1998a). The determination of SML during the

WorkHab FCE was found to be reliable in the test-retest reliability study as discussed in the previous chapter and it was hypothesized that there would be a significant change in the physiological responses between muscles at the point of the SML. Results are discussed in **Manuscript 7 (5.2):** *"Physiological correlates of functional capacity evaluations: finding the safe maximal lift"* submitted for publication in Archives of Physical Medicine and Rehabilitation. The study identified there was a significant difference in muscle activity with increasing weight lifted, and a significant difference between the up lift (bench to shoulder) and the down lift (shoulder to bench) in most muscles. However, no significant difference was found in muscle activity before or after the SML, which indicates construct validity of the SML during the bench to shoulder lift of the WorkHab FCE was unable to be established using these SEMG physiological parameters.

5.1 Manuscript: Content Validity of the WorkHab Functional Capacity Evaluation

Citation: James, C., Mackenzie, L. and Capra, M. Content Validity of the WorkHab Functional Capacity Evaluation. Submitted for publication in Disability and Rehabilitation.

ABSTRACT

Introduction:

Evidence of reliability and validity is essential for accurate and meaningful measurement of functional capacity evaluations (FCEs). The WorkHab Functional Capacity Evaluation is an Australian standardised assessment commonly used in practice. However the validity of this tool is yet to be established. Content validity determines the degree to which the items in an instrument reflect the content domain being measured. This study aimed to evaluate the content validity of the WorkHab FCE.

Method:

A cross sectional survey of health professionals who conduct FCEs investigated their ratings of item relevance and item difficulty across different categories of work and injuries. The survey also explored the relationship of the WorkHab FCE components with the physical demand items of the Dictionary of Occupational Titles (DOT).

Results:

Therapists identified the WorkHab components of climbing, crawling and balance as less relevant for sedentary work. Most items were relevant for manual work and vocational retraining with over 90% agreement between raters.

The lifting components were perceived as the most difficult for those with a lower back injury, or an upper limb injury. Over 85% of therapists also rated lifting, carrying, climbing, standing, kneeling, crouching, squatting and crawling as difficult for clients with lower limb injuries.

Sixteen of the eighteen WorkHab FCE components had 100% agreement with the equivalent items on the DOT. The DOT demand of lift correlated highly with all the lifting components of the WorkHab FCE. The WorkHab FCE components that had poor agreement with the DOT were mainly sensory items such as see, feel, talk, and hear.

Conclusion:

The findings of this study support content validity for the WorkHab FCE specifically in relation to manual work and for vocational retraining purposes.

Keywords: Functional capacity evaluation; content validity; WorkHab FCE; Occupational rehabilitation.

5.1.1 INTRODUCTION

Functional capacity evaluation (FCE) is defined as “an evaluation of capacity of activities that is used to make recommendations for participation in work while considering the person’s body functions and structures, environmental factors, personal factors and health status” (Soer et al. 2008). An FCE is a performance measure that is used in occupational and vocational rehabilitation to make decisions about the capacity of a worker in relation to their work abilities. FCEs are used to screen potential employees as pre-employment assessments, to assess physical rehabilitation needs, to determine work readiness and job placement following injury, to facilitate return to work and to determine a person’s functional capacity for compensation or litigation reasons (King et al. 1998; Lee et al. 2001; Schonstein and Kenny 2001; Strong et al. 2004a; Gouttebarge et al. 2006; Innes 2006; Gross et al. 2007)

An FCE can be made up of a medical history, physical examination and a variety of work related performance tests, which are often founded on a taxonomy based on the US Department of Labor Dictionary of Occupational Titles (DOT) (U.S.Department of Labor 1965). The DOT taxonomy provides information about the physical demands and work factors for many jobs. There are 20 job factors which express the requirements of a job and the capacities a worker must have to meet or exceed those demands. Many FCEs have been based on the DOT taxonomy and studies suggest that content validity for these FCEs is deemed sufficient when they incorporate the various components of the DOT taxonomy (Fishbain et al. 1994; Lechner et al. 1994; Innes and Straker 1999b; Gibson et al. 2005).

Functional capacity evaluations and work-related assessments must have evidence of psychometric properties including reliability and validity, both of which are seen as important pre-requisites for accurate and meaningful measurement. These properties ideally should be established as the tool is being developed however historically, this has not been demonstrated with much FCE research (King et al. 1998; Innes and Straker 1999b; Innes and Straker 1999a; Ekbladh et al. 2004; Gouttebarge et al. 2004). More recently, psychometric evaluation of commercially available assessment tools has begun to appear in the literature (Ting et al. 2001; Matheson et al. 2002; Durand et al. 2004; Pransky and Dempsey 2004; Reneman et al. 2004).

The WorkHab FCE is an Australian standardised assessment used in occupational rehabilitation and work settings to determine functional capacity. It is widely used in the Australian context (Deen et al. 2002; Cotton et al. 2006; James and Mackenzie 2009b).

The WorkHab FCE consists of objective physiological measures, including heart rate; observations of the clients biomechanics and the clients reported pain and perceived exertion (effort) ratings during various activities (Bradbury and Roberts 1998). However there is limited evidence on the psychometric properties of this assessment tool (Innes and Straker 1999a; Innes and Straker 1999b; Innes 2006). Aspects of reliability of the WorkHab FCE have been undertaken (James et al. 2010), however evidence of validity is lacking.

Validity is defined as the extent to which an instrument measures what it is intended to measure (Portney and Watkins 2009). Content validity is defined as the degree to which the items in an instrument adequately reflect the content domain being measured (Portney and Watkins 2009). It has been noted that content validity of FCEs has rarely been undertaken (Innes and Straker 1999b; McKenzie et al. 1999). Content validity is often evaluated with the use of an expert panel who review the content of the instrument to determine if it contains appropriate content. In regards to FCEs, one approach used has been to examine how well the FCE covers the physical demands of the DOT taxonomy (Lee and Chan 2003; Kersnovske et al. 2005). This may be more appropriate for generic FCEs in comparison to job specific FCEs, as these DOT items consider a range of general physical demands (King et al. 1998). Another approach to content validity evaluation is to use literature and relevant research to develop FCE methodology to be used with specific groups. Reesink et al and Reneman et al have developed FCE methodologies for FCEs with work related upper limb disorders and neck disorders respectively (Reneman et al. 2005; Reesink et al. 2007), using this method of determining content validity.

The determination of content validity can be a subjective process, with the use of an expert panel to review an instrument and evaluate if it satisfies the content domain (McKenzie et al. 1999; Portney and Watkins 2009). Statistical methods have been developed to add to this process (Lawshe 1975; Lynn 1986; Thorn and Deitz 1989; Sireci 1998) where experts rate item-objective congruence, item relevance and item difficulty (Lawshe 1975; Thorn and Deitz 1989; Sireci 1998; McKenzie et al. 1999; Kersnovske et al. 2005).

Item-objective congruence is used to determine if an item matches the objective for which it was designed to measure and can be analysed using the index of item-objective congruence (Thorn and Deitz 1989). How relevant an item is to the domain of interest is also part of determining content validity. This is determined by judges rating the relevance of an item on a scale of not relevant to extremely relevant. Mean scores and percentage agreement are then calculated (Sireci 1998; Kersnovske et al. 2005). A

content validity ratio (CVR) can also be calculated to determine how essential items are to the domain (Lawshe 1975). Another aspect of content validity is item difficulty, or how difficult the item is perceived to be by the expert panel, for the population for which the test is designed.

The aim of this study was to determine the content validity of the WorkHab FCE by investigating the relevance of the FCE components to different work types, the difficulty of the FCE components to various client types and the relationship between the FCE components and the physical demands of the DOT, with health professionals (experts) who use FCEs in practice.

5.1.2 METHOD

Following ethics approval, a cross sectional on-line survey was used to gain the opinions of health professionals, who conduct FCEs in practice.

Procedure and participants

Health professionals who conduct FCEs in practice were invited to participate and included: accredited providers of the WorkHab FCE; members of the Australian Association of Occupational Therapists; members of the Exercise Physiology Association of Australia; members of the Australian Rehabilitation Provider Association; and members of the vocational rehabilitation special interest group of the British Association of Occupational Therapists. An email advertising an online survey and explaining the study was sent to these health professionals, by their member body, inviting them to complete the anonymous survey online.

Survey

The online survey was developed to include the components of the WorkHab FCE, which are also generic components of many commercially available FCEs. This was piloted with five health professionals. The pilot survey was modified according to feedback, prior to being distributed widely to health professionals who conduct FCEs in practice.

The survey included 15 questions within four sections: Demographic data of the participants; Relevance of components of the WorkHab FCE; Difficulty of the WorkHab FCE components; and relationship of the WorkHab FCE components with the Dictionary of Occupational Titles (DOT) physical demands (U.S.Department of Labor 1965). An attachment outlining the DOT physical demands definitions was included with the survey for health professionals to use as a resource.

Health professionals were asked to rate the relevance of items or components of the WorkHab FCE to different types of work: sedentary, manual and in respect to vocational retraining on a five point Likert style scale, as recommended (Sireci 1998), ranging from always relevant, to never relevant, with an additional option of not applicable.

Participants were asked to rate how difficult they perceived the components of the FCE were for clients with various injury locations: low back injury, upper limb injury and lower limb injury, also on a five point Likert scale from extremely difficult to not difficult.

Finally, respondents were asked to identify any relationship between the WorkHab FCE components and the DOT, scoring each component against the physical demands of the DOT on a matrix.

Completed surveys were collated using an online survey instrument (surveymonkey.com) and survey data was downloaded by the researcher for analysis.

Data analysis

All data was entered into SPSS (version18.0) for analysis. Descriptive analysis, 95% confidence intervals, standard deviations, t-test, construct validity ratios and percentage agreements were calculated where appropriate.

The content validity ratio was calculated using the formula: $CVR= (n^e –(N/_2)) / (N/_2)$ where n^e is the number of participants indicating ‘essential’ and N is the total number of participants (Lawshe 1975).

As construct validity ratios (CVR) require a dichotomous rating, the CVRs for the relevance and difficulty items were calculated by collapsing the scale of relevance ratings into essential (sometimes, often and always relevant) or not relevant, and for the difficulty ratings into essential (moderate, very and extremely difficult) or not difficult. Results for the relationship of the FCE components to the DOT were calculated as percentage agreement.

The minimum value of the CVR is based upon the number of participants, and has been calculated to consider results that might reasonably have occurred by chance. As this study included more than 40 participants, the minimum value was set at 0.29 to meet the 5% level of significance. According to Lawshe, only those items that meet the 5% level of significance are considered content valid and retained or included in a test (Lawshe 1975). However, any item perceived to be ‘essential’ by more than half of the participants has some degree of content validity with the more participants in agreement the greater the content validity.

Data analysis on the relevance and difficulty of items was conducted for respondents as a whole, and also to compare responses of health professionals with less than or more than 2 years FCE experience, and for those with or without post graduate qualifications.

5.1.3 RESULTS

Participants

One hundred and six participants completed the online survey. Of these 94% were occupational therapists (n=100), 4% physiotherapists (n=4), 1% nurses (n=1) and 1% exercise physiologists (n=1). This response is generally representative of the ratio of the different professionals working for rehabilitation providers in NSW (James and Mackenzie 2009b).

Ninety two percent of the participants worked in Australia with the majority (38%) identifying they worked in NSW (n=40). Four participants identified that they worked across two or more states in Australia, 6% of participants worked in the UK (n=7) and 1% in New Zealand (n=1). The mean number of years of professional experience was 14.2 years (range: 0.5-38yrs) and the mean number of years of FCE experience was 7.6 years (range: 0.5-26yrs). Not all participants completed all questions on the survey. Numbers are noted on the results tables indicating the number of participants who completed each section.

Analysis of the data identified that there was no significant difference between the groups of health professionals with less than or more than 2 years FCE experience, nor for those with or without post graduate qualifications in relation to the item relevance data, and limited differences in relation to item differences. Results are therefore presented on the data as a whole and any differences between groups are discussed.

Item Relevance

Item relevance was calculated for sedentary work, manual work and for vocational retraining. The relevance of the various components of the WorkHab FCE to sedentary work, manual work and vocational retraining can be seen in Table 1

Table 1: Relevance of WorkHab FCE components

WorkHab FCE component	Sedentary Work (n=76)				Manual Work (n=75)				Vocational Retraining (n=74)			
	Meana	SDb	Relevantc%	CVR	Meana	SDb	Relevantc%	CVR	Meana	SDb	Relevantc%	CVR
Grip strength	3.7	0.9	87	0.74	4.6	0.7	75	0.97	3.9	0.9	92	0.86
Lift FB	3.2	0.9	76	0.53	4.7	0.4	100	1	4.2	0.7	99	0.97
Lift BS	3.1	0.8	79	0.59	4.7	0.5	100	1	4.2	0.7	100	1
Lift BOH	2.8	0.8	63	0.27^	4.5	0.6	100	1	3.9	0.8	99	0.97
Carrying	3.5	0.9	85	0.7	4.7	0.5	100	1	4.2	0.7	100	1
Push/Pull	2.8	0.9	63	0.26^	4.5	0.7	97	0.95	4.1	0.8	97	0.95
Reaching	4.3	0.7	97	0.95	4.6	0.5	100	1	4.4	0.7	100	1
Standing	3.9	0.9	93	0.86	4.7	0.5	100	1	4.5	0.7	99	0.97
Sitting	4.8	0.3	100	1.0	3.8	0.9	95	0.92	4.5	0.7	99	0.97
Walking	4.2	0.8	96	0.92	4.7	0.5	100	1	4.5	0.7	100	1
Climbing	2.4	0.9	41	-0.17^	3.9	0.8	97	0.95	3.7	0.9	95	0.89
Stooping	3.1	0.9	80	0.59	4.1	0.8	99	0.97	3.9	0.9	96	0.92
Kneeling	2.6	0.8	55	0.11^	3.9	0.7	97	0.95	3.8	0.8	100	1
Balancing	2.5	1.0	44	-0.12^	4.1	0.8	97	0.95	3.7	0.9	92	0.84
Crouching	2.7	0.9	61	0.22^	3.9	0.7	99	0.97	3.7	0.9	92	0.84
Squatting	2.7	0.8	68	0.35	4.1	0.7	100	1	3.9	0.8	96	0.92
Crawling	1.8	0.87	13	-0.73^	3.3	0.8	87	0.73	3.3	1.0	78	0.57
Job simulation	3.7	1.3	82	0.65	4.1	1.2	95	0.89	3.8	1.2	93	0.86

aMean = Mean score for each of the FCE components (1= not relevant, 5= essential) b Standard deviation = standard deviation of mean score

cRelevance = % agreement for essential FCE components. ^ = Content Validity Ratio does not meet the 5% level

Item relevance: Sedentary work

Lifting bench to overhead; push/pull; kneeling and crouching were identified as having questionable relevance (more than half the participants perceiving this as essential suggesting some degree of content validity however, the CVRs were not at the 5% level of significance suggested for inclusion in the final test items (Lawshe 1975)). Climbing, balancing and crawling were identified as not relevant as shown by the negative CVR.

Item relevance: Manual work

All the FCE components were identified as being essential to manual work with 75% or more of the participants identifying these components as essential and all CVR's meeting the 5% level of significance.

Item relevance: Vocational Retraining

All the FCE components had CVRs above the 5% level of significance when considering the item relevance for vocational retraining. Most of the FCE components were identified by more than 90% of participants as being essential, with the exception of crawling which was identified by 78% of participants as essential.

Qualitative results for relevance

Twenty two participants provided qualitative comments regarding the relevance of the FCE components. Several themes emerged from this data: 1) the purpose and goal of the FCE; 2) the injury type and characteristics of the person; and 3) vocational retraining.

The purpose and goal of the FCE was a factor impacting upon the relevance of the various FCE components, with participants identifying the reason for the FCE (pre-employment screening; return to existing job; or returning to new job) will have an impact upon the relevance of various components. As one participant commented:- *"Whilst the components are relevant it does depend on the situation and goal of the FCE - pre-placement screening versus working out if someone is ready to go back to work".* Another participant commented*: "Relevance of tasks are dependent on the role the person may be undertaking and the specific job requirements".*

The injury type and characteristics of the person were also noted to impact upon the relevance of FCE components, with some components being more or less relevant depending upon the injury. As one participant commented: *"depends on the condition (i.e. trauma versus cumulative injury)"*.

The relevance of various FCE components to vocational retraining was commented upon by several participants. The relevance was noted as dependent upon the goal and the identification of appropriate retraining options. One participant stated: "*vocational retraining: all components are relevant if you are assessing someone to determine the type of retraining that best suits*" and another commented*:* "*vocational training can depend on the goal identified from the vocational assessment as to what tasks are relevant".*

One participant sums up the issue of relevance by commenting: *"the FCE always has to be relevant to the referral question and address these issues. This will determine the content and the type of tests included in the assessment".*

Item Difficulty

Item difficulty was rated according to clients with a lower back injury, an upper limb injury (including the shoulder, elbow, arm and hand) or a lower limb injury (including hip, knee, ankle and foot injury). The item difficulty of the various components of the WorkHab FCE to the different injury locations is presented in Table 2.

Table 2: Difficulty of WorkHab FCE components

WorkHab FCE component	Low Back Injury (n=72)				Upper Limb (n=72)				Lower Limb (n=71)			
	Meana	SDb	Relevantc %	CVR	Meana	SDb	Relevantc %	CVR	Meana	SDb	Relevantc %	CVR
Grip strength	1.4	0.7	6	-0.89^	3.6	1.0	92	0.83	2	0.3	1	-0.97^
Lift FB	3.8	0.8	96	0.92	3.5	0.9	88	0.75	3.9	0.8	99	0.97
Lift BS	3.2	0.9	81	0.61	3.9	0.8	96	0.92	2.8	1.1	56	0.11^
Lift BOH	3.7	0.8	97	0.94	4.2	0.8	97	0.94	2.8	1.1	60	0.19^
Carrying	3.2	0.7	85	0.69	3.6	0.9	92	0.83	3.4	0.9	88	0.75
Push/Pull	3.3	0.7	89	0.78	3.6	0.8	96	0.92	3.3	0.9	82	0.64
Reaching	2.8	0.9	61	0.22^	3.6	0.9	92	0.83	1.8	0.9	21	-0.58^
Standing	2.7	0.9	63	0.25^	1.3	0.8	8	-0.83^	3.3	0.9	87	0.75
Sitting	2.9	0.9	64	0.28^	1.2	0.6	97	0.94	2.1	1.0	23	-0.54^
Walking	2.3	0.9	37	-0.26^	1.4	0.6	8	-0.83^	3.5	0.9	87	0.75
Climbing	3.2	0.9	79	0.58	2.6	1.3	54	0.07^	4.1	0.8	99	0.97
Stooping	4.1	0.9	94	0.89	1.6	0.9	17	-0.67^	2.8	1.1	63	0.27^
Kneeling	2.8	0.9	63	0.25^	1.4	0.7	10	-0.81^	4.1	0.8	97	0.94
Balancing	2.7	1.0	59	0.18^	1.5	0.8	11	-0.78^	3.6	0.9	89	0.77
Crouching	3.5	0.8	89	0.77	1.5	0.7	13	-0.75^	4	0.8	96	0.92
Squatting	3.2	0.9	79	0.58	1.5	0.6	9	-0.83^	4.3	0.7	100	1.00
Crawling	3.5	0.9	89	0.77	2.9	1.2	79	0.57	4.1	0.8	99	0.97
Job simulation	2.4	1.4	66	0.32	2.9	1.5	60	0.20^	2.4	1.7	64	0.27^

a Mean = Mean score for each of the FCE components (1= not relevant, 5= essential), b Standard deviation = standard deviation of mean score c Relevance = % agreement for essential FCE components. ^ = Content Validity Ratio does not meet the 5% level

Item Difficulty: Lower Back Injury

When considering the CVR for reaching, standing, sitting, kneeling and balancing, these were not at the 5% level of significance, however it should be noted more than half of the participants perceived these components as difficult for this client group indicating some level of content validity. Grip strength, and walking were perceived as not difficult for clients suffering a low back injury as shown by the negative CVR's.

A significant difference was found when analysing the data for the groups of >2yrs FCE experience and < 2yrs FCE experience for the item of reaching (p=0.04), however in both groups more than half of participants identified this component as difficult (75% and 57% respectively) suggesting a level of content validity but the CVR's were lower than the 5% level of significance in both groups.

Item difficulty: Upper limb injury

Standing, walking, stooping, kneeling, balancing, crouching and squatting were perceived as not difficult for clients with an upper limb injury as shown with the negative CVR's. The CVR for climbing and for job simulation was not at the 5% level of significance, however more than half of the participants identified these as being difficult components for a FCE for person with an upper limb injury suggesting some level of content validity.

Grip strength and lifting floor to bench were rated significantly differently by those with less than or more than 2 years FCE experience (p=0.01 and p=0.004 respectively) however in both instances a high percentage of participants perceived these as difficult and CVR's were above the 5% level of significance. There was also a significant difference between the two groups of raters for crouching (p=0.01), however in this instance both groups gave a low mean score (2/5, 1.4/5 respectively) and there was a negative CVR for both groups suggesting this was perceived as not difficult by all.

Item difficulty: Lower limb injury

For clients with lower limb injuries (including hip, knee, ankle and foot injuries), lifting bench to shoulder, lifting bench to overhead, stooping, and job simulation were perceived as less difficult FCE items with CVR's not at the 5% level of significance. However, more than half of the participants perceived these items as difficult for someone with a lower limb injury and mean scores were 2.8/5 for all components with the exception of job simulation that had a mean score of 2.4/5. This suggests some level of content validity. Grip strength, reaching and sitting were not perceived as difficult for this client group as shown with the negative CVR's.

The difficulty of the FCE item "job simulation" had significantly different rating by participants with and without 2 yrs FCE experience (p=0.01), where participants with less experience gave higher ratings for difficulty. The CVR's were above the 5% level of significance in both instances indicating this component should be included in a content valid test. Difficulty ratings for the reaching FCE item were also significantly different between these groups (p=0.03) and CVR's not at the 5% level of significance in either group suggesting this item should not be included in a content valid test.

Qualitative results for items of difficulty

Twenty seven participants made some comments regarding the perceived difficulty of the FCE components and identified issues of generalisability according to the criteria used in the survey. As one participant commented:- *"depends very much on exact location and severity of injury".* Another participant commented: *"a hip injury would have a different functional implication with regard to sitting as compared to an ankle injury"* and another stated*: "someone with an upper limb injury to the hand has significantly different difficulties to someone with an upper limb injury to the shoulder".*

Generalising the item of job simulation was also identified as of concern with difficulty of job simulation varying according to the type of job simulation task and the specific injury location and severity. Climbing was identified as being too broad as this could be climbing ladders or stairs and the difficulty for these will vary according to injury type.

One participant summarised the issues of item difficulty with this quote: *"injury severity, general fitness, mental attitude and suitability of job task greatly affects this rating of FCE components".*

Relationship of FCE components to the DOT physical demands

The percentage agreement between the raters for the WorkHab FCE components and the DOT physical demands is presented in Table 3.

Table 3 Percentage agreement between WorkHab FCE components and the DOT physical demands (n=56).

WORKHAB ITEMS	DOT ITEMS																			
	Lift	Carry	Push	Pull	Reach	Stand	Sit	Walk	Climb	Stoop	Kneel	Crouch	Crawl	Balance	Handle	Finger	Feel	Talk	See	Hear
Grip St	79	79	70	73	28	4	13	2	32	0	0	0	9	4	93	75	66	5	27	7
Lift FB	100	69	20	20	60	87	0	46	4	53	18	80	4	73	82	51	58	7	71	13
Lift BS	100	73	29	27	82	86	2	44	6	13	0	9	2	78	91	53	60	7	71	13
Lift BOH	100	69	27	27	87	86	2	44	6	13	0	7	4	78	89	53	60	7	69	16
Carry	89	100	21	23	45	70	0	89	27	21	2	14	2	77	88	46	57	9	66	18
Push	23	25	100	34	63	77	13	84	13	23	4	4	7	70	75	36	59	9	73	20
Pull	23	25	30	100	68	75	9	84	16	18	2	4	7	68	79	39	55	9	66	21
Reach	32	29	34	32	100	50	46	20	32	21	13	18	21	55	58	55	50	7	66	7
Stand	30	30	27	27	23	100	9	43	29	25	13	13	4	70	9	7	13	9	41	16
Sit	0	0	8	8	38	9	100	4	2	11	2	13	2	42	28	32	34	21	42	28
Walk	26	36	28	26	17	64	6	100	42	4	4	6	6	79	9	6	26	17	62	28
Climb	11	11	15	17	47	57	2	64	100	8	6	15	11	89	60	26	53	11	64	26
Stoop	23	4	19	17	32	74	6	17	6	100	13	26	9	60	17	17	30	8	40	15
Kneel	8	6	2	2	17	40	11	15	9	23	100	38	32	72	15	9	30	8	40	13
Balance	42	48	39	42	42	75	35	60	58	37	40	50	39	100	10	4	23	8	40	17
Crouch	23	8	9	8	11	38	13	13	15	28	38	100	28	79	11	2	19	8	38	13
Crawl	4	2	2	0	23	23	6	8	6	25	56	50	100	60	15	4	40	8	48	19
Job Sim	88	85	77	77	82	88	88	92	77	79	82	82	71	88	88	82	82	68	77	77

Grip St= Grip strength Lift FB= Lift Floor to Bench Lift BS= Lift Bench to shoulder Lift BOH = Lift Bench to overhead Job Sim= Job simulation.

Sixteen of the eighteen WorkHab FCE components had 100% agreement in ratings with the equivalent items on the DOT, the exceptions being grip strength and job simulation. For the WorkHab item “grip strength” the highest agreement for DOT items was handle (90%), lift and carry (79%), finger (75%), pull (73%) and push (70%). Job simulation had a range of 68-92% agreement for each DOT item. The DOT items that had poor agreement with the WorkHab items were mainly sensory items such as see, feel, talk and hear.

The DOT demand of lift had 100% agreement with the WorkHab lifting components: lifting floor to bench, bench to shoulder and bench to overhead. Other components that had greater than 50% agreement were carry, reach, stand, balance, handle, finger, feel and see for all lifts, also stoop and crouch for the floor to bench lift. The relationship of the DOT with the lifting components of the WorkHab can be seen in figure 1.

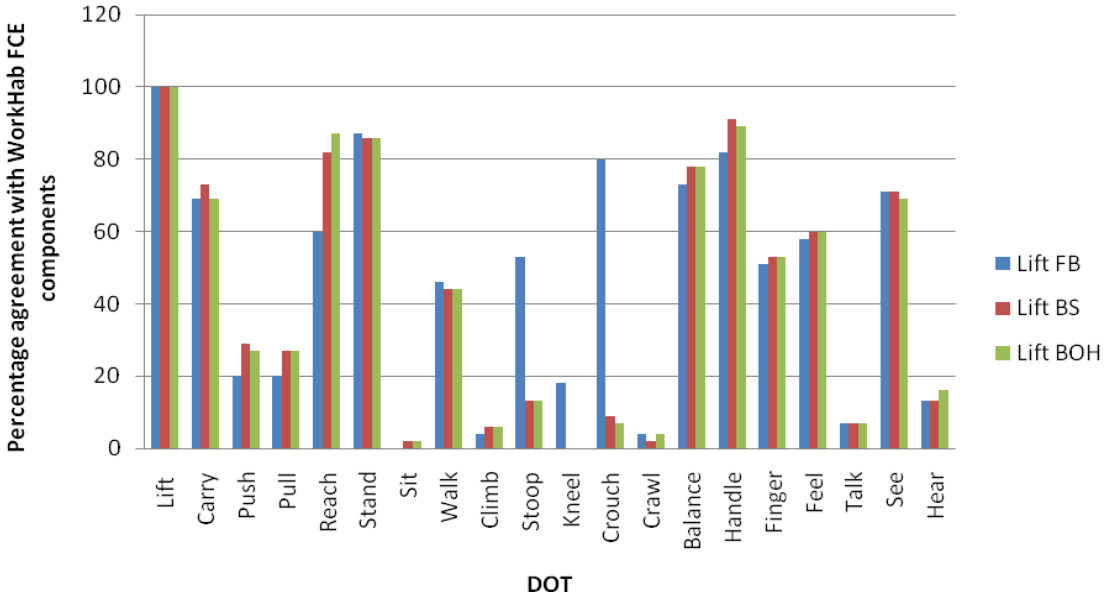


Figure 1 – Relationship of Lifting FCE components to DOT physical demands.

FB=Floor to bench
BS= Bench to shoulder
BOH= Bench to overhead

Qualitative comments indicated that all factors are relevant for job simulation and depended upon the nature of the injury and client’s job. Hearing and talking were acknowledged as useful to gauge communication ability but not essential as other methods of communication can be employed if required, such as sign language.

5.1.4 DISCUSSION

The aim of this study was to evaluate the content validity of the WorkHab FCE, by exploring item relevance, item difficulty and the relationship between the physical

demands of the DOT and the WorkHab FCE components. This process contributes to the evidence for content validity of the WorkHab FCE.

Role of experts in determining content validity

The use of experts is a common method to determine content validity (Portney and Watkins 2009), with an expert identified as having familiarity and expertise with the content and the knowledge and skill levels of the client population to be tested (Sireci 1998). Experts must be representative of those who are familiar with a topic, with demographic variables and specialisations needing consideration for appropriate membership to the expert panel. In this study health professionals who conduct FCEs of any type were invited to participate and complete the online survey. As the components of the WorkHab FCE are common to other assessments, this survey was not limited to those who conduct the WorkHab FCE specifically but also included health professionals who conduct FCEs other than WorkHab. This increased the number of potential participants and allowed a broader view of the content validity of the WorkHab FCE by a range of expert health professionals in this field. Lower numbers of 'experts' can provide significant levels of agreement (Lawshe 1975), however in this study the use of an online survey allowed wider access to appropriate health professionals. The NSW WorkCover (NSW) (WorkCover (NSW) 2002), recognition system for professionals to conduct FCEs was used as the basis for the 2 year time period used to compare participants by experience. Different skills regarding work and the use of FCEs may be expected for those who have been exposed to more practice experience. To further explore the expertise of participants, they were also grouped according to the possession of award recognised health professional post graduate qualifications. The rationale for distinguishing between these two groups was the assumption that those with post graduate qualifications would have advanced critical analysis and clinical reasoning skills. However when the data was analysed, there were limited differences between all the groups suggesting that both experience and post graduate qualifications did not influence the ratings participants gave. The findings from this study suggest that expertise in the practice skill of using FCEs may not necessarily be determined by longer experience or further academic qualifications suggesting practical training maybe all that is required which lends support to the WorkCover NSW guidelines (WorkCover (NSW) 2002). Where differences between groups were identified in this study, those with less experience were more conservative with responses however in all instances the CVR's were consistent, either above the 5% level of significance in both groups or below the 5% level of significance in both groups.

Item relevance

Health professionals rated the relevance of the FCE components to sedentary work, manual work and to vocational retraining. These broad work categories were chosen as FCEs are used for a variety of purposes.

The relevance ratings of the FCE components to sedentary work, manual work and vocational retraining were similar regardless of experience or post graduate qualifications.

Sedentary work was defined as work involving sitting most of the time, occasional walking and standing and occasional use of force to lift or move objects. Consistent with this definition, there was less agreement in the relevance ratings for the climbing, crawling and balance items, suggesting that these items may need further evaluation to be applied to sedentary work. Balance has been identified as an area of less relevance in other studies of FCEs. It was not identified as relevant in a study of the GAPP FCE for individuals with chronic low back pain, however it was not discussed in that study if it was relevant for any particular category of work (sedentary or manual) (Kersnovske et al. 2005). Halpern also identified the utility of balance as questionable in his study of a taxonomy relevant to low back impairments (Halpern 2001). Lawshe suggests the more participants who perceive an item as 'essential' the greater the degree of its content validity, and goes on to suggest any item perceived relevant by more than half the participants, has some degree of content validity (Lawshe 1975). As such the findings for sedentary work suggest eleven items of the WorkHab FCE (grip strength, lift floor to bench and bench to shoulder, carry, reach, stand, sit, walk, stoop, squat and job simulation) have acceptable relevance and therefore can be considered content valid.

Manual work was defined for this study as physically demanding work, with most of the FCE components being identified as essential by the experts. Similar results were found for the FCE components in relation to vocational retraining. The item of crawling although receiving a lower percentage score for both manual work and vocational retraining had a CVR that was above the 5% level of significance indicating this component is content valid. In a study of the DOT on formwork carpentry in Hong Kong, (manual work), crawling was identified as never involved (Lee and Chan 2003), which is supported in this study. However, sitting was also identified as never involved in the formwork carpentry study whereas the FCE sitting item was rated as relevant to manual work by participants in this study.

The findings in relation to item relevance suggest health professionals consider the FCE components relevant when completing a FCE for manual or vocational retraining with an acceptable level of content validity.

Previous research has identified that health professionals do not always use all components of an FCE, using parts as appropriate for different jobs and injury types (James and Mackenzie 2009b). This is consistent with the findings of this study that different items within the WorkHab FCE were rated as more relevant for different work types. Several authors (King et al. 1998; Innes and Straker 2003a) discuss the issue of generic versus specialist (or job specific) FCE testing, and the ability to alter the test depending upon the needs of the client, work and situation. The qualitative responses in this study also indicated the relevance of an item was dependent upon the purpose and goal of the FCE, the nature of the work, and the injury type. It has previously been found that FCEs conducted for those with no job (vocational retraining) were more generic (Innes and Straker 2003a). The 'experts' in our study identified all items as relevant for both manual and vocational retraining purposes, which supports the FCE research that suggests FCEs are more generic for those with no job.

Item difficulty

Health professionals rated the difficulty of the FCE components in relation to a client with a lower back injury, an upper limb injury (shoulder, elbow, arm and hand) and a lower limb injury (hip, knee, ankle and foot injury) to cover a range of injury types commonly assessed with an FCE.

The difficulty ratings of FCE items for people with a lower back injury suggested the most difficult tasks were the lifting components, carrying, push/pull. stooping, crouching, and crawling. This is consistent with the content validity findings in the GAPP FCE, where the lifting components were identified as the most difficult and therefore raised the most concern regarding safety (Kersnovske et al. 2005). Stooping has also previously been identified as a difficult item (Kersnovske et al. 2005).

In relation to an upper limb injury the items rated as most difficult were also the lifting components, carrying, push/pull, as well as reaching and grip strength. Reneman et al in his study of development of an FCE for upper limb disorders (Reneman et al. 2005), identified the overhead lift, reaching and grip strength as essential components which supports these findings.

The experts in our study agreed on the level of difficulty for items of the WorkHab FCE, indicating that some were more or less difficult depending upon the injury type.

Difficulty of items needs to be considered in relation to the safety of an injured person completing these activities. Safety has been identified as a critical issue for consideration when conducting FCEs (Innes and Straker 2003a; Gibson and Strong 2005). These results emphasise the findings of previous research that indicated health professionals wanted flexibility in an FCE to adapt the assessment in accordance with the client's injury and also in relation to the client's job (King et al. 1998; Innes and Straker 2003b).

Relationship of FCE components to the DOT physical demands

Health professionals were asked to evaluate the links between the DOT physical demand items and the WorkHab FCE components, percentage agreement was calculated for the responses. Percentage agreement is frequently used as a measure to determine content validity (Thorn and Deitz 1989; Sireci 1998) and has been used in the determination of content validity in other assessments (Halpern 2001; Kersnovske et al. 2005; Sandqvist et al. 2008; Spanjer et al. 2010).

As reported, sixteen of the eighteen WorkHab FCE components had 100% agreement with the equivalent items on the DOT. The DOT items that had poor agreement with the WorkHab FCE components were mainly sensory items such as see, feel, talk and hear. The qualitative comments suggested that the talk and hear items are useful to determine communication ability but were not necessarily essential. The DOT items of handle, finger and feel had fairly high levels of agreement with the WorkHab FCE equivalent items. The item of manipulating objects in Halpern's study of the functional taxonomy for low back impairments only had 43% agreement (Halpern 2001), however only one injury type was considered, whereas in this study the comparisons were in relation to the general FCE components. In relation to the FCE lifting components, participants identified these as related to all of the DOT physical demands to some extent, with those directly related to lifting having higher percentage agreement than those less directly related such as sitting, climbing, crawling, talking and hearing. These items were also identified in a study of the DOT on formwork carpentry in Hong Kong as being either never or occasionally performed which may support the lower percentage agreement of relationship with the DOT in this study (Lee and Chan 2003).

The levels of agreement between the FCE components and the DOT physical demands found in this study, give support to the WorkHab FCE components having acceptable content validity when using the DOT taxonomy as the basis for an FCE as has been done in other studies (Fishbain et al. 1994; Lechner et al. 1994; Innes and Straker 1999b; Gibson et al. 2005).

When considering the content validity of items within an FCE, the issue of flexibility to allow the health professional to adapt the assessment to a particular client injury and job has been identified in several studies as something health professionals value (King et al. 1998; Innes and Straker 2003b; Innes and Straker 2003a; Strong et al. 2004b; James and Mackenzie 2009b; James and Mackenzie 2009a), however this impacts on the standardisation and psychometric properties of an FCE. In the medico-legal arena, standardised tools have been identified as valuable (Innes and Straker 2002), however other research has supported the notion that there are several types of FCE; for those with a specific job, without a specific job; and a job or work capacity evaluation (Innes and Straker 2003a; Jones and Kumar 2003) and these will require the use of different components to match the specifics of the assessment. When considering the relevance ratings in this study for different work types, less WorkHab FCE items were identified as relevant to sedentary work, giving more support to the content validity of the WorkHab FCE for manual work and for vocational retraining purposes. This poses the question whether FCEs should be developed for specific work types. Several researchers have investigated the content validity of FCEs for different injuries or conditions. The GAPP FCE was developed specifically for use with those presenting with chronic low back pain (Gibson et al. 2005), and FCEs for work related neck disorders and work related upper limb disorders have also been investigated (Reneman et al. 2005; Reesink et al. 2007). However, FCEs for specific work types are not reported in the literature. The comprehensiveness of the DOT physical demands in relation to all work types and how relevant these are to specific real work environments may also be a limitation (Fishbain et al. 1994; Lysaght 1997; Lee and Chan 2003).

Limitations

One of the limitations of this study was that the survey did not ask how many FCEs health professionals had conducted therefore levels of experience were evaluated by the length of professional experience and length of time conducting FCEs. This may have underestimated the level of expertise of the participants in this study. In practice, the number of FCEs conducted over time is generally not recorded other than to generate invoices in the short-term (Gibson and Strong 2006). Additionally, details of the numbers of specific professionals involved in completing FCEs is not available and it is therefore unknown if the sample in this study was representative. Furthermore, participants rated relevance and difficulty on a five point Likert scale, and for analysis this data was collapsed into a dichotomous variable of essential/ (relevant or difficult) or not essential/ (relevant or difficult). This may have lead to an inflation of the estimates

of relevance and difficulty by raters. Another limitation of this study was that the index of item-objective congruence in relation to the DOT components could not be calculated as the on-line survey did not allow a -1, 0, +1 response as required to calculate the index (Thorn and Deitz 1989). However percentage agreement is frequently used as a measure to determine content validity (Thorn and Deitz 1989; Sireci 1998) and has been used in the determination of content validity in other assessments (Halpern 2001; Kersnovske et al. 2005; Sandqvist et al. 2008; Spanjer et al. 2010), as it was in this study.

5.1.5 CONCLUSION

This study analysed the item relevance, item difficulty and relationship of the FCE components of the WorkHab FCE to the physical demands of the DOT. The findings support content validity for the WorkHab FCE specifically in relation to manual work and for vocational retraining purposes. Further research into the relevance of the FCE items: Lift bench to overhead, push/pull, climb, balance, kneel and crawl, and the application of the WorkHab items to sedentary work requires additional examination. Although the content validity findings relate to items within the WorkHab FCE, many of these are generic components of several FCEs. Therefore these results on the content validity of WorkHab FCE items may be used as a starting point to evaluate other FCEs that assess a similar range of components.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the health professionals who piloted the survey and wish to thank all who completed the final survey.

REFERENCES

Bradbury, S.& Roberts, D. (1998). Workhab Functional Capacity Evaluation Procedural Manual. WorkHab Australia.

Cotton, A., Schonstein, E.& Adams, R. (2006). "Use of Functional Capacity Evaluations by Rehabilitation Providers in NSW." Work **26**(3): 287-295.

Deen, M., Gibson, L.& Strong, J. (2002). "A Survey of Occupational Therapy in Australian Work Practice." Work **19**(3): 219-230.

Durand, M., Loisel, P., Poitras, S., Mercier, R., Stock, S.& Lemaire, J. (2004). "The Interrater Reliability of a Functional Capacity Evaluation: The Physical Work Performance Evaluation." Journal of Occupational Rehabilitation **14**(2): 119-129.

Ekbladh, E., Haglund, L.& Thorell, L. (2004). "The Worker Role Interview - Preliminary Data on the Predictive Validity of Return to Work of Clients after an Insurance Medicine Investigation." Journal of Occupational Rehabilitation **14**(2): 131-141.

Fishbain, D., Abdel-Moty, E., Cutler, R., Khalil, T., Sadek, S., Rosomoff, R.& Rosomoff, H. (1994). "Measuring Residual Functional Capacity in Chronic Low Back Pain Patients Based on the Dictionary of Occupational Titles." Spine **19**(8): 872-880.

Gibson, L.& Strong, J. (2005). "Safety Issues in Functional Capacity Evaluation: Findings from a Trial of a New Approach for Evaluating Clients with Chronic Back Pain." Journal of Occupational Rehabilitation **15**(2): 237-251.

Gibson, L.& Strong, J. (2006). "Letter to the Editor (Authors' Response).Kersnovske S, Gibson L, Strong J. Item Validity of the Physical Demands from the Dictionary of Occupational Titles for Functional Capacity Evaluation of Clients with Chronic Back Pain. Work 24 (2005), 157-169." Work **26**(2): 231-233.

Gibson, L., Strong, J.& Wallace, A. (2005). "Functional Capacity Evaluation as a Performance Measure: Evidence for a New Approach for Clients with Chronic Back Pain." Clinical Journal of Pain **21**(3): 207-15.

Gouttebarge, V., Wind, H., Kuijer, P.& Frings-Dresen, M. (2004). "Reliability and Validity of Functional Capacity Evaluation Methods: A Systematic Review with Reference to Blankenship, Ergos Work Simulator, Ergo-Kit and Isernhagen Work System." International Archives of Occupational and Environmental Health **77**(8): 527-537.

Gouttebarge, V., Wind, H., Kuijer, P., Sluiter, J. & Frings-Dresen, M. (2006). "Reliability and Agreement of 5 Ergo-Kit Functional Capacity Evaluation Lifting Tests in Subjects with Low Back Pain." Archives of Physical Medicine and Rehabilitation **87**(10): 1365-1370.

Gross, D., Battie, M.& Asante, A. (2007). "Evaluation of a Short Form Functional Capacity Evaluation: Less Maybe Best." Journal of Occupational Rehabilitation **17**(3): 422-435.

Halpern, M. (2001). "Functional Assessment Taxonomy Relevant to Low-Back Impairments." Journal of Occupational Rehabilitation **11**(3): 201-215.

Innes, E. (2006). "Reliability and Validity of Functional Capacity Evaluations: An Update." International Journal of Disability Management Research **1**(1): 135-148.

Innes, E.& Straker, L. (1999a). "Reliability of Work-Related Assessments." Work **13**(2): 107-124.

Innes, E.& Straker, L. (1999b). "Validity of Work-Related Assessments." Work **13**(2): 125-152.

Innes, E.& Straker, L. (2002). "Workplace Assessments and Functional Capacity Evaluations: Current Practices of Therapists in Australia." Work **18**(1): 51-66.

Innes, E.& Straker, L. (2003a). "Attributes of Excellence in Work Related Assessments." Work **20**(1): 63-76.

Innes, E.& Straker, L. (2003b). "Workplace Assessments and Functional Capacity Evaluations: Current Beliefs of Therapists in Australia." Work **20**(3): 225-236.

James, C.& Mackenzie, L. (2009a). "The Clinical Utility of Functional Capacity Evaluations: The Opinion of Health Professionals Working within Occupational Rehabilitation." Work **33**(3): 231-9.

James, C.& Mackenzie, L. (2009b). "Health Professional's Perceptions and Practices in Relation to Functional Capacity Evaluations: Results of a Quantitative Survey." Journal of Occupational Rehabilitation **19**(2): 203-211.

James, C., Mackenzie, L.& Capra, M. (2010). "Test-Retest Reliability of the Manual Handling Component of the Workhab Functional Capacity Evaluation in Healthy Adults." Disability and Rehabilitation **32**(22): 1863-1869.

Jones, T.& Kumar, S. (2003). "Functional Capacity Evaluation of Manual Materials Handlers: A Review." Disability and Rehabilitation **25**(4-5): 179-191.

Kersnovske, S., Gibson, L.& Strong, J. (2005). "Item Validity of the Physical Demands from the Dictionary of Occupational Titles for Functional Capacity Evaluation of Clients with Chronic Low Back Pain." Work **24**(2): 157-169.

King, P., Tuckwell, N.& Barrett, T. (1998). "A Critical Review of Functional Capacity Evaluations." Physical Therapy **78**(8): 852.

Lawshe, C. (1975). "A Quantitative Approach to Content Validity." Personnel Psychology **28**(4): 563-575.

Lechner, D., Jackson, J., Roth, D.& Straaton, K. (1994). "Reliability and Validity of a Newly Developed Test of Physical Work Performance." Journal of Occupational Medicine **36**(9): 997-1004.

Lee, G., Chan, C.& Hui-Chan, C. (2001). "Work Profile and Functional Capacity of Formwork Carpenters at Construction Sites." Disability and Rehabilitation **23**(1): 9-14.

Lee, G. & Chan, C. (2003). "Use of Dictionary of Occupational Titles (DOT) on Formwork Carpentry-a Comparison between the United States and Hong Kong." Work **20**(2): 103-10.

Lynn, M. (1986). "Determination and Quantification of Content Validity." Nursing Research **35**(6): 382-5.

Lysaght, R. (1997). "Job Analysis in Occupational Therapy: Stepping into the Complex World of Business and Industry." American Journal of Occupational Therapy **51**(7): 569-75.

Matheson, L., Rogers, L., Kaskutas, V.& Dakos, M. (2002). "Reliability and Reactivity of Three New Functional Assessment Measures." Work **18**(1): 41-50.

McKenzie, J., Wood, M., Kotecki, J., Clark, J.& Brey, R. (1999). "Research Notes: Establishing Content Validity: Using Qualitative and Quantative Steps " American Journal of Health Behaviour **23**(4): 311-318.

Portney, L.& Watkins, M. (2009). Foundations of Clinical Research: Applications to Practice. Upper Saddle River, New Jersey, Pearson Prentice Hall.

Pransky, G.& Dempsey, P. (2004). "Practical Aspects of Functional Capacity Evaluations." Journal of Occupational Rehabilitation **14**(3): 217-229.

Reesink, D., Jorritsma, W.& Reneman, M. (2007). "Basis for a Functional Capacity Evaluation Methodology for Patients with Work-Related Neck Disorders." Journal of Occupational Rehabilitation **17**(3): 436-449.

Reneman, M., Brouwer, S., Meinema, A., Dijkstra, P., Geertzen, J.& Groothoff, J. (2004). "Test-Retest Reliability of the Isernhagen Work Systems Functional Capacity Evaluation in Healthy Adults." Journal of Occupational Rehabilitation **14**(4): 295-305.

Reneman, M., Soer, R.& Gerrits, E. (2005). "Basis for an FCE Methodology for Patients with Work-Related Upper Limb Disorders." Journal of Occupational Rehabilitation **15**(3): 353-63.

Sandqvist, J., Gullberg, M., Henriksson, C.& Gerdle, B. (2008). "Content Validity and Utility of the Assessment of Work Performance (AWP)." Work **30**(4): 441-450.

Schonstein, E.& Kenny, D. (2001). "The Value of Functional and Work Place Assessments in Achieving a Timely Return to Work for Workers with Back Pain." Work **16**(1): 31-38.

Sireci, S. (1998). "Gathering and Analysing Content Validity Data." Educational Assessment **5**(4): 299-321.

Soer, R., Cees, P., Groothoff, J., Geertzen, J.& Reneman, M. (2008). "Towards Consensus in Operational Definitions in Functional Capacity Evaluation: A Delphi Survey." Journal of Occupational Rehabilitation **18**(4): 389-400.

Spanjer, J., Krol, B., Brouwer, S., Popping, R., Groothoff, J.& van der Klink, J. (2010). "Reliability and Validity of the Disability Assessment Structured Interview (DASI): A Tool for Assessing Functional Limitations in Claimants." Journal of Occupational Rehabilitation **20**(1): 33-40.

Strong, S., Baptiste, S., Clarke, J.& Costa, M. (2004a). "Use of Functional Capacity Evaluations in Workplaces and the Compensation System: A Report on Workers' and Users' Perceptions." Work **23**(1): 67-77.

Strong, S., Baptiste, S., Cole, D., Clarke, J., Costa, M., Shannon, H., Reardon, R.& Sinclair, S. (2004b). "Functional Assessment of Injured Workers: A Profile of Assessor Practices." The Canadian Journal of Occupational Therapy **71**(1): (13-23).

Thorn, D.& Deitz, J. (1989). "Examining Content Validity through the Use of Content Experts." Occupational Therapy Journal of Research **69**(6): 334-346.

Ting, W., Wessel, J., Brintnell, S., Maikala, R.& Bhambhani, Y. (2001). "Validity of the Baltimore Therapeutic Equipment Work Simulator in the Measurement of Lifting Endurance in Healthy Men." American Journal of Occupational Therapy **55**(2): 184-90.

U.S.Department of Labor (1965). Dictionary of Occupational Titles. Washington, DC., U.S. Government printing office.

WorkCover (NSW). (2002). Rehabilitation Providers. Standards and Conditions of Accreditation. WorkCover. Sydney, WorkCover NSW.

5.2 Manuscript: Physiological correlates of functional capacity evaluations: finding the safe maximal lift

Citation: James, C., Mackenzie, L. and Capra, M.F. Physiological correlates of functional capacity evaluations: finding the safe maximal lift. Submitted for publication in Archives of Physical and Medical Rehabilitation.

ABSTRACT

Objectives: To quantify and analyse muscle function using surface electromyography (SEMG) with a view to establishing a physiological indicator for safe maximal lift (SML) during the bench to shoulder lift of the WorkHab functional capacity evaluation.

Design: Experimental

Setting: Laboratory

Participants: 20 healthy volunteers

Interventions: Not applicable

Main Outcome Measures: SEMG data for upper trapezius, mid deltoid, thoracic, brachioradialis and bicep muscles.

Results: There was a significant difference in muscle activity and duration of muscle activation with increasing weight lifted. There was a significant difference between the up lift (bench to shoulder) and the down lift (shoulder to bench) for all muscles except the brachioradialis. No significant difference was found in muscle activity before or after the SML.

Conclusions: Construct (convergent) validity of the bench to shoulder lift of this FCE was unable to be established using SEMG physiological parameters in this study.

Keywords: Electromyography, Lifting, Work capacity evaluation

5.2.1 INTRODUCTION

Functional Capacity Evaluations (FCEs) are performance measures designed to define the functional abilities and limitations of an individual in the context of safe, productive work tasks (King et al. 1998; Soer et al. 2008) and are commonly used with individuals who have suffered work related musculo-skeletal injuries (Strong et al. 2004).

The theoretical knowledge that informs the use of an FCE by practitioners is varied because different theoretical models have different foci (biomechanical, physiological, metabolic, psychophysical and kinesiophysical) (Abdel-Moty et al. 1993; Gibson and Strong 1997). The psychophysical approach places control with the participant who decides when to terminate the test thereby determining maximum function, however issues with injured participants determining their own safe limits have been identified (Abdel-Moty et al. 1993; Gibson and Strong 1997; Mitchell 2008). The kinesiophysical model utilises observation of movement patterns (biomechanics) and physiological performance to evaluate safe maximum function (Isernhagen 1992) and the evaluator controls the test and determines maximal capacity.

The WorkHab FCE is commonly used in the Australian occupational rehabilitation environment (Deen et al. 2002; Cotton et al. 2006; James et al. 2007; James and Mackenzie 2009) and is based upon the kinesiophysical approach with the evaluator observing physiological signs and biomechanical movement patterns to determine maximal capacity including safe maximal lifting (SML) load (Bradbury and Roberts 1998).

Client safety when conducting an FCEs has been identified as a critical issue (Innes and Straker 2003; Gibson and Strong 2005) and it has been suggested that therapists undertake a decision making process, that considers safety first, then determines the constructs of dependability and utility demonstrated by the FCE (Innes and Straker 2003).

The aim of this experimental study was to quantify and analyse muscle function using surface electromyography (SEMG) with a view to establishing a physiological indicator for safe maximal lift during the bench to shoulder lift of the WorkHab FCE. SEMG has been used in studies of occupational musculoskeletal disorders to quantify exposure or fatigue in different environments or with different equipment (Granström et al. 1985; Jensen et al. 1999; Davis et al. 2000; Hansson et al. 2000; Laursen et al. 2003; Cook et al. 2004).

This study was undertaken to establish construct validity, specifically convergent validity of the WorkHab FCE. Construct validity is the degree to which a theoretical construct is measured by an instrument and convergent validity is where two measures believed to reflect the same phenomenon are highly correlated (Portney and Watkins 2009).The study aimed to identify if there is a relationship between muscle activity and increasing weight lifted during the WorkHab FCE and to identify if there is any relationship between therapist determined safe maximal lift and a physiological indicator.

5.2.2 METHODS

Sample

Ethical clearance was obtained from the University Human Research Ethics Committee, following which subjects were recruited from staff and students of the university. Volunteers contacted the researcher directly, received an information statement, signed a consent form and arranged an assessment time. A convenience sample of 20 healthy adult volunteers was recruited.

Study design

This study used a laboratory based, experimental design. Participants completed the bench to shoulder lift of the WorkHab FCE.

The manual handling component of the WorkHab FCE uses a modular box system. Boxes are set at an appropriate height, and the subject is instructed to lift the load box (initially empty) from beginning to end height and return. This is repeated three times before additional weight is added to the load box. The FCE assessment follows a protocol of increasing load at each height until the safe maximum lifting limit is reached (Bradbury and Roberts 1998).

In this experimental study the height of the lift was relative to the subject's waist (for the bench component) and shoulder (for the shoulder component). The lifting protocol was explained to participants and they were instructed to lift with weight being incrementally increased after each 3 lifts. Participants were instructed to lift until they perceived they had reached their maximum abilities and could not lift more weight. The assessor observed the lifting, asked participants if they wished to complete another lift with each additional weight and monitored for excessive heart rate, however in this study they did not influence participants when to cease lifting.

Prior to commencing the FCE each subject completed a pre-assessment screening: a medical status questionnaire; height and weight and a blood pressure check.

Participants' muscles were palpated, skin prepared by shaving, abrading with sandpaper and cleaned with alcohol before disposable, self adhesive, surface electrodes (ADInstruments: MLA1010) were positioned parallel and either end of the muscle bulk of the brachioradialis, bicep, mid deltoid and upper trapezius muscles on the left side of the body, and approximately 2cm from the spinous process at the level of T6 and T9 for the thoracic paraspinal muscles. A single researcher trained in physiotherapy and manipulation skills marked each participant to minimise variation.

Data collection

Participants were video-taped using two Sony Handycam Camcorders (Model HRD-HC9E, Sony, Tokyo, Japan). The camera images were recorded digitally using Dartfish Pro-Suite (Dartfish, Lausanne, Switzerland). The cameras were set up to view the rear coronal and right sagittal planes during lifting.

An ADI Powerlab 8SP (ML 785, ADInstruments.com) in combination with a tower of four dual bioamplifiers (ML135) was used to collect surface electromyography data and transmit to a laptop for processing using LabChart software (version 7.1.2, ADInstruments.com).

Measurements

Maximal isometric voluntary contractions (MVC) were recorded for each muscle as a reference. Three resisted maximum voluntary contractions of 6 seconds each were recorded for each muscle. For the thoracic spine the participant laid prone, shoulders were abducted to 90º, elbow flexed to 90º, and pressure applied to the trunk at level T12/L1 as the participant was instructed to raise their arms and head in one movement into the 'aeroplane' position. MVC for the upper trapezius was recorded with resisted elevation of the shoulders. For the mid deltoid, the participant abducted the shoulder to 90º and was resisted at the wrist whilst trying to adduct the shoulder. The MVC for the biceps muscle was recorded with the elbow held at 90º flexion, resistance was applied at the wrist and stabilisation of the upper arm occurred whilst the participant attempted to flex the elbow. With the elbow flexed to 90°, forearm in neutral position and upper arm close to body, the participant was resisted in radial deviation to record the MVC of the brachioradialis muscle.

SEMG recordings were taken at a sampling rate of 1k/s with high (minimum cut off 0.3Hz) and low pass (maximum cut off 1kHz) filters. Following data collection, the raw data for each muscle was normalised and rectified using the Labchart software.

Root-mean-square (RMS) of the raw data in mV, the integral of the RMS trace in mV.s, mean power frequency of the raw data in Hz and duration in seconds for each set of 3 lifts at each weight was collected. For each lift up (bench to shoulder height) and down (shoulder height to bench) the mean of the RMS, integral, power and duration was calculated from the 3 lifts completed at each weight. RMS values were normalised and expressed as a percentage of the maximum voluntary contraction to aid comparison between individuals.

The video of each participant completing the lifting segment was de-identified and an expert panel of five occupational therapists experienced in the use of the WorkHab FCE determined the safe maximal lift (SML) for each individual. This point was used during analysis of the data.

Data analysis

Descriptive statistics were calculated for each participant. A generalised linear mixed model regression with a random intercept for the individual was used to analyse the last five lifts of SEMG data for each participant.

All statistical analyses were performed using the statistic package STATA V11.1 (StataCorp 2009).

5.2.3 RESULTS

Participants

The study sample consisted of 10 women and 10 men ranging in age from 21 years to 64 years, with a mean age of 39.5 years (SD 14.8)**.**

Relationship of muscle activity and increased weight

The duration of muscle activation with increased weight is shown in figure 1. In most muscles there is an incremental rise in the time of muscle activation with increasing weight lifted. The exception is the upper trapezius muscle where lift 3 saw a sharp increase in duration of activation. However, when this data was analysed using regression, there was no significant difference between lifts 1, and 2, 3 or 4 for all muscles but a significant difference was found between lift 1 and 5: Upper Trapezius ($p=0.006$); Mid Deltoid ($p=0.002$); Thoracic ($p=0.005$); Brachioradialis ($p= 0.024$) and Biceps ($p=0.0005$).

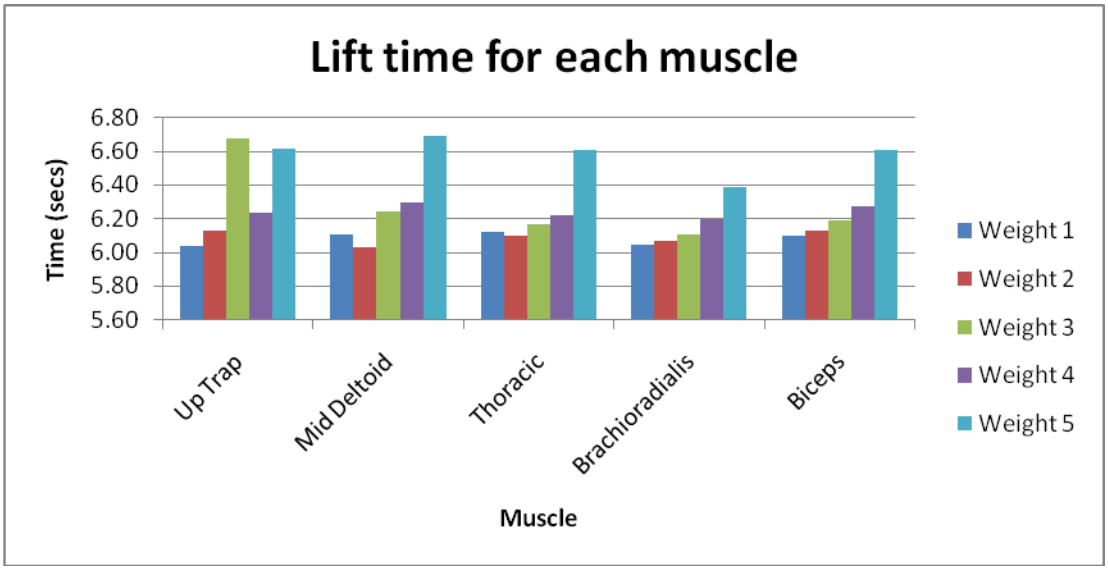


Figure 1: Lift time, muscle involvement and weight progression (N=20)

In all participants, muscle activity showed a linear relationship with weight. The greater the weight being lifted the higher the muscle activity as is shown in figure 2. This difference with weight was significant in all muscles as can be seen in Table 1.

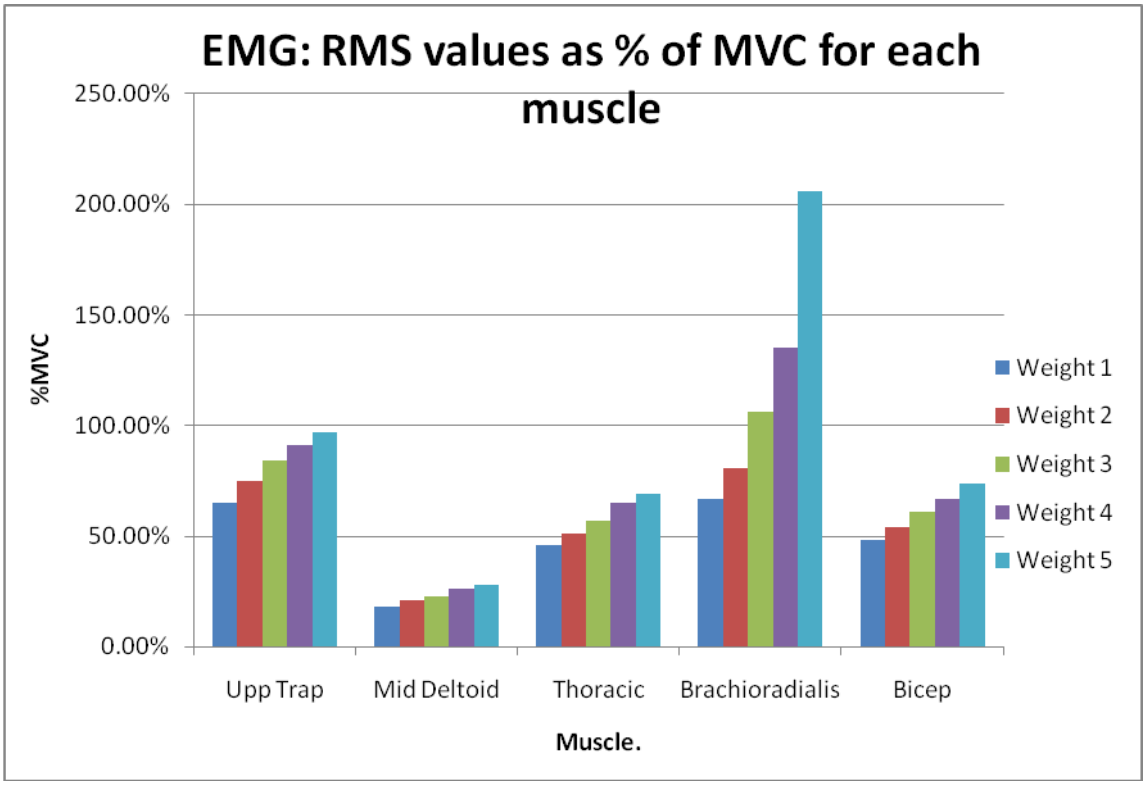


Figure 2: RMS as % of MVC for each muscle for each weight. (N=20)

Table 1: Differences between lifting load up and down, before and after SML and with increasing weight (N=20)

Muscle	Difference in lift up/down				Difference before and after SML				Difference with weight			
	Coef.	St Err	CI	P-value	Coef.	St Err	CI	P-value	Coef.	St Err	CI	P-value
Upper Trapezius	0.048	0.004	0.04:0.06	0.000	-0.001	0.006	-0.01:0.01	0.895	0.019	0.002	0.01:0.02	0.000
Mid Deltoid	0.031	0.002	0.02:0.03	0.000	-0.006	0.004	-0.01:0.002	0.146	0.014	0.001	0.01:0.02	0.000
Thoracic	0.079	0.014	0.05:0.11	0.000	0.0003	0.027	-0.53:0.05	0.989	0.592	0.008	0.04:0.07	0.000
Brachioradialis	0.049	0.204	-0.35:0.45	0.819	0.221	0.373	-0.51:0.95	0.555	0.275	0.122	0.03:0.51	0.024
Biceps	0.098	0.009	0.07:0.12	0.000	-0.006	0.0172	-0.04:0.02	0.704	0.064	0.005	0.05:0.07	0.000

Relationship between safe maximal lift, lift up, lift down and weight

Table 1 outlines the differences between lifting the load up (bench to shoulder height) and down (shoulder height to bench) for each muscle. There was a significant difference between the up lift and the down lift for all muscles except the brachioradialis ($p=0.819$). No significant difference was found in muscle activity before or after the SML, as determined by the expert panel, in any of the muscles studied.

5.2.4 DISCUSSION

This study demonstrated that there is an increase in muscle activity with increasing weight during the bench to shoulder lift of the WorkHab FCE. Other studies investigating the effect of load have also found increasing levels of muscle activity with increased load (Cole et al. 2004; Robertson et al. 2008; McBride et al. 2010). Incremental weight increases were used in this study which may account for the linear increase identified in the muscle activity levels. A relationship between the weight (load) lifted and the subsequent load placed upon the spine has also been reported, during a work capacity assessment (Cole et al. 2004) and in studies investigating load on joints (Arjmand and Shirazi-Adl 2005). Frequency of lift has also been reported to impact upon muscle activity in both low back muscles and shoulder muscles (Nielsen et al. 1998).

An incremental increase in duration of muscle activity (time) with increased weight lifted was found, with a significant increase between the lightest load and the heaviest load. Muscle activation time maybe affected by the frequency of lift. Lift frequency was not considered in this study, as with the procedures for the WorkHab FCE, participants completed the lifts in their own time, with no imposed number of lifts per minute as have been considered in other studies (Davis et al. 2000). However, during the WorkHab FCE the timing and pacing of lifts is considered as part of the manual handling score (Bradbury and Roberts 1998). Further studies investigating the relationship between muscular activity time and weight in other lifts and with an injured population are needed to determine if the findings in this study of increasing time with increasing weight is common across lifting as part of an FCE.

Differences in muscle activation when lifting up (bench to shoulder height) and down (shoulder height to bench) with loads were found in this study, with the difference being significant in the upper trapezius, mid deltoid, thoracic and bicep muscles. Differences between ascending and descending lifts have also been identified in studies investigating muscle function during full squats with the ascending lifts showing an

increased muscle activity level when compared to descending lifts (Robertson et al. 2008) in most muscles.

In this study there was no significant difference between the up and down of the lift in the brachioradialis muscle which, as an elbow flexor muscle, maybe the result of the elbows being flexed during both phases of the lift. The grip used during this lift may also have impacted upon the activity of this muscle. The WorkHab FCE uses load boxes with cut-out handles on the side of the box which dictate the hand grip required. The corresponding wrist position used with these handles may impact upon the brachioradialis muscle activity levels and could account for the lack of difference identified in this study. The presence of handles was found to have a significant difference on spinal loading in a study of the effects of box features during warehouse manual handling (Marras et al. 1999) . A review of the biomechanics during the bench to shoulder lift would also indicate if joint position may affect the level of muscle activity and the lack of difference between the up phase and the down phase of the lift in the brachioradialis. Biomechanics play a significant role in lifting and it has been shown that altered biomechanics during lifting tasks can cause increased joint stress and an increased risk of injury (Arjmand and Shirazi-Adl 2005).

In this study, there was no significant difference in muscle activity between the lifts immediately before and after the safe maximal lift, as determined by the expert panel. SML in the WorkHab FCE is determined based on heart rate and observation of biomechanical movement patterns however this study showed there was no relationship between the SML and the recorded physiological parameters, hence no physiological indicator was found that correlated with the SML.

SEMG has been used to determine subject effort. In a comprehensive muscular activity profile it was used in conjunction with range of motion testing, as a predictor of effort during FCE testing and compared to therapist ratings (Gatchel et al. 2009), with high levels of sensitivity suggesting this can be used as a method to identify muscular effort but this may not be linked to the determination of whether a lift is at a safe maximal level.

Safe maximal lifting limits have been proposed according to lifting height, frequency and worker characteristics (Snook and Ciriello 1991) and when considering the compression force on the spine (Konz 1982), however Cole found in his study of the loads on the spine during the work capacity assessment, the recommended limits were exceeded (Cole et al. 2004). Differences between safe lifting limits were also reported when comparing FCE and the NIOSH guidelines (Kuijer et al. 2006). The principles of

safe manual handling techniques are used to determine SML: a steady base of support; neutral spinal curves; loads kept close to the spine and within range of gravity where possible; no twisting; and movements that are smooth and controlled (Bradbury and Roberts 1998). Observation of the recruitment of upper extremity strength for the ability to control the lift and the ability to stabilise the lumbo-sacral spine without hyperextension is suggested for the shoulder lift during the WorkHab FCE (Bradbury and Roberts 1998). Other observations recommended for determination of safe maximal lift include, muscle bulging of prime movers, involuntary use of accessory muscles, altered body mechanics including counterbalancing, loss of equilibrium, increased base of support, decreased efficiency and smoothness of movement, cardiovascular signs of heart rate and breathing patterns and referred symptoms (Gross and Battie 2002). An operational definition of safe lift used in a study by Gardener and McKenna advised that an unsafe lift included observation of extremes of trunk or upper limb range of motion, poor control of the load and/ or the load not kept close to the body (Gardener and McKenna 1999).The end point of lifting in a waist to waist lift was reported as being mainly for biomechanical reasons particularly an increased lumbar lordosis in a study of the GAPP-FCE with persons with chronic low back pain (Gibson and Strong 2005). Good reliability of determining safe maximal lift during FCEs has been reported (Gardener and McKenna 1999; Isernhagen et al. 1999; Gross and Battie 2002; Reneman et al. 2002; James et al. 2010).

The lack of significant change in muscle activity identified in this study at the point of the safe maximal lift suggests that biomechanical determinates are being used by therapists as part of the clinical reasoning process to clarify a SML and these do not necessarily correlate with physiological indicators in the muscles of the upper arm during a bench to shoulder lift. Further studies investigating the biomechanical changes during the bench to shoulder lift are recommended to ascertain if there are quantifiable changes occurring that assist in the determination of a SML.

Limitations

In this study subjects were healthy individuals with no manual handling restrictions. Further studies on an injured population are needed to determine if these results are generalisable. The technique used to normalize data for comparison in this study was determination of the MVC and calculation of the RMS as a percentage of the MVC. The use of MVC is commonly used in SEMG studies, however it is noted that this is an isometric contraction and is being compared to a dynamic activity. Maximal activation of motor units is dependent upon many factors such as motivation, training and the specific muscle activation, with reports that an MVC can be 20-40% less than the true

maximum (Soderberg and Knutson 2000). In this study a standardized procedure with a qualified physiotherapist was used to minimize this limitation.

5.2.5 CONCLUSIONS

This study identified that there was a significant relationship between the weight lifted and muscle activity in the upper trapezius, mid deltoid, thoracic, brachioradialis and biceps muscles during a bench to shoulder lift of the WorkHab FCE. The study identified that there was a significant difference between muscle activity in the up lift (bench to shoulder height) and the down lift (shoulder to bench height) in these muscles with the exception of the brachioradialis. However, no significant differences were found in the recorded physiological parameters immediately before or after the safe maximal lift. Therefore convergent validity was unable to be established using these parameters.

The quantification of the safe maximal lift in clinical practice requires further investigation with a suggested focus upon the biomechanical patterns of lifting during the FCE and the specific clinical reasoning processes used by therapists in practice.

ACKNOWLEDGEMENTS

The authors would like to thank Suzanne Snodgrass for her physiotherapy expertise, and the individuals who volunteered to participate in this study.

REFERENCES

Abdel-Moty, E., Fishbain, D., Khalil, T., Sadek, S., Cutler, R., Steele Rosomoff, R.& Rosomoff H. (1993). "Functional Capacity and Residual Functional Capacity and Their Utility in Measuring Work Capacity." Clinical Journal of Pain **9**(3): 168-173.

Arjmand, N.& Shirazi-Adl, A. (2005). "Biomechanics of Changes in Lumbar Posture in Static Lifting." Spine **30**(23): 2637-48.

Bradbury, S.& Roberts, D. (1998). Workhab Functional Capacity Evaluation Procedural Manual. WorkHab Australia.

Cole, M., Grimshaw, P.& Burden, A. (2004). "Loads on the Lumbar Spine During a Work Capacity Assessment Test." Work **23**(2): 169-78.

Cook, C., Burgess-Limerick, R.& Papalia, S. (2004). "The Effect of Upper Extremity Support on Upper Extremity Posture and Muscle Activity During Keyboard Use." Applied Ergonomics **35**(3): 285-292.

Cotton, A., Schonstein, E.& Adams, R. (2006). "Use of Functional Capacity Evaluations by Rehabilitation Providers in NSW." Work **26**(3): 287-295.

Davis, K., Jorgensen, M.& Marras, W. (2000). "An Investigation of Perceived Exertion Via Whole Body Exertion and Direct Muscle Force Indicators During the Determination of the Maximum Acceptable Weight of Lift." Ergonomics **43**(2): 143 - 159.

Deen, M., Gibson, L.& Strong, J. (2002). "A Survey of Occupational Therapy in Australian Work Practice." Work **19**(3): 219-230.

Gardener, L.& McKenna, K. (1999). "Reliability of Occupational Therapists in Determining Safe Maximal Lifting Capacity." Australian Occupational Therapy Journal **46**(3): 110-19.

Gatchel, R., Ricard, M., Choksi, D., Mayank, J.& Howard, K. (2009). "The Comprehensive Muscular Activity Profile (CMAP): Its High Sensitivity, Specificity and Overall Classification Rate for Detecting Submaximal Effort on Functional Capacity Testing." Journal of Occupational Rehabilitation **19**(1): 49-55.

Gibson, L.& Strong, J. (1997). "A Review of Functional Capacity Evaluation Practice." Work **9**(1): 3-11.

Gibson, L.& Strong, J. (2005). "Safety Issues in Functional Capacity Evaluation: Findings from a Trial of a New Approach for Evaluating Clients with Chronic Back Pain." Journal of Occupational Rehabilitation **15**(2): 237-251.

Granström, B., Kvarnström, S.& Tiefenbacher, F. (1985). "Electromyography as an Aid in the Prevention of Excessive Shoulder Strain." Applied Ergonomics **16**(1): 49-54.

Gross, D.& Battie, M. (2002). "Reliability of Safe Maximum Lifting Determinations of a Functional Capacity Evaluation." Physical Therapy **82**(4): 364-372.

Hansson, G., Nordander, C., Asterland, P., Ohlsson, K., Strömberg, U., Skerfving, S.& Rempel, D. (2000). "Sensitivity of Trapezius Electromyography to Differences between Work Tasks - Influence of Gap Definition and Normalisation Methods." Journal of Electromyography and Kinesiology **10**(2): 103-115.

Innes, E.& Straker, L. (2003). "Attributes of Excellence in Work Related Assessments." Work **20**(1): 63-76.

Isernhagen, S. (1992). "Functional Capacity Evaluation:Rationale, Procedure, Utility of the Kinesiophysical Approach." Journal of Occupational Rehabilitation **2**(3): 157-168.

Isernhagen, S., Hart, D.& Matheson, L. (1999). "Reliability of Independent Observer Judgements of Level of Lift Effort in a Kinesiophysical Functional Capacity Evaluation." Work **12**(2): 145-50.

James, C.& Mackenzie, L. (2009). "Health Professional's Perceptions and Practices in Relation to Functional Capacity Evaluations: Results of a Quantitative Survey." Journal of Occupational Rehabilitation **19**(2): 203-211.

James, C., Mackenzie, L.& Capra, M. (2010). "Test-Retest Reliability of the Manual Handling Component of the Workhab Functional Capacity Evaluation in Healthy Adults." Disability and Rehabilitation **32**(22): 1863-1869.

James, C., Mackenzie, L.& Higginbotham, N. (2007). "Health Professionals' Attitudes and Practices in Relation to Functional Capacity Evaluations." Work **29**(2): 81-88.

Jensen, C., Finsen, L., Hansen, K.& Christensen, H. (1999). "Upper Trapezius Muscle Activity Patterns During Repetitive Manual Material Handling and Work with a Computer Mouse." Journal of Electromyography and Kinesiology **9**(5): 317-325.

King, P., Tuckwell, N.& Barrett, T. (1998). "A Critical Review of Functional Capacity Evaluations." Physical Therapy **78**(8): 852.

Konz, S. (1982). "NIOSH Lifting Guidelines." American Industrial Hygiene Association Journal **43**(12): 931-933.

Kuijer, W., Dijkstra, P., Brouwer, S., Reneman, M., Groothoff, J.& Geertzen, J. (2006). "Safe Lifting in Patients with Chronic Low Back Pain: Comparing FCE Lifting Task and NIOSH Lifting Guidelines." Journal of Occupational Rehabilitation **16**(4): 579-589.

Laursen, B., Søgaard, K.& Sjøgaard, G. (2003). "Biomechanical Model Predicting Electromyographic Activity in Three Shoulder Muscles from 3D Kinematics and External Forces During Cleaning Work." Clinical Biomechanics **18**(4): 287-295.

Marras, W., Granata, K., Davis, K., Allread, W.& Jorgensen, M. (1999). "Effects of Box Features on Spine Loading During Warehouse Order Selecting." Ergonomics **42**(7): 980-996.

McBride, J., Larkin, T., Dayne, A., Haines, T.& Kirby, T. (2010). "Effect of Absolute and Relative Loading on Muscle Activity During Stable and Unstable Squatting." International Journal of Sports Physiology and Performance. **5**(2): 177-83.

Mitchell, T. (2008). "Utilization of the Functional Capacity Evaluation in Vocational Rehabilitation " Journal of Vocational Rehabilitation **28**(1): 21-28.

Nielsen, P., Andersen, L.& Jørgensen, K. (1998). "The Muscular Load on the Lower Back and Shoulders Due to Lifting at Different Lifting Heights and Frequencies." Applied Ergonomics **29**(6): 445-450.

Portney, L.& Watkins, M. (2009). Foundations of Clinical Research: Applications to Practice. Upper Saddle River, New Jersey, Pearson Prentice Hall.

Reneman, M., Jaegar, S., Westmaas, M.& Goeken, L. (2002). "The Reliability of Determining Effort Level of Lifting and Carrying in a Functional Capacity Evaluation." Work **18**(1): 23-27.

Robertson, D., Wilson, J.& St Pierre, T. (2008). "Lower Extremity Muscle Functions During Full Squats." Journal of Applied Biomechanics **24**(4): 333-9.

Snook, S.& Ciriello, V. (1991). "The Design of Manual Handling Tasks: Revised Tables of Maximum Acceptable Weights and Forces." Ergonomics **34**(9): 1197-213.

Soderberg, G.& Knutson, L. (2000). "A Guide for Use and Interpretation of Kinesiologic Electromyographic Data." Physical Therapy **80**(5): 485-498.

Soer, R., van der Schans, C., Groothoff, J., Geertzen, J.& Reneman, M. (2008). "Towards Consensus in Operational Definitions in Functional Capacity Evaluation: A Delphi Survey." Journal of Occupational Rehabilitation **18**(4): 389-400.

StataCorp (2009). Stata Statistical Software: Version 11.1.College Station, Tx..

Strong, S., Baptiste, S., Clarke, J.& Costa, M. (2004). "Use of Functional Capacity Evaluations in Workplaces and the Compensation System: A Report on Workers' and Users' Perceptions." Work **23**(1): 67-77.

5.5 Chapter Summary

The usefulness of an assessment depends on the extent to which clinicians can rely on data as being accurate and meaningful and is necessary for drawing inferences from the data and to determine how the results of an assessment can be used. Validity is one component of measurement that is required to achieve this (Portney and Watkins 2009).

From this research, content validity for the WorkHab FCE is supported, specifically in relation to manual work and for vocational retraining purposes. The findings in relation to item relevance suggest health professionals consider the FCE components relevant for manual or vocational retraining; however, less relevance was identified for sedentary work. As has been highlighted earlier, FCEs are used for a variety of purposes, therefore defining intended outcomes, and what the results of an FCE actually mean in practice can be difficult and complicates any validity assessment. In these studies of the WorkHab FCE, aspects of lifting bench to overhead, push/pull, kneeling and crouching were identified as less relevant, and climbing, balancing and crawling were identified as not at all relevant for sedentary work. It has also been suggested that balance is an area of less relevance in FCEs in other studies (Halpern 2001; Kersnovske et al. 2005). The findings in relation to difficulty of FCE items for people with a lower back injury suggested that the most difficult tasks were the lifting components, carrying, push/pull, stooping, crouching, and crawling, which is consistent with the content validity findings of the GAPP-FCE (Kersnovske et al. 2005). Item difficulty in relation to the upper limb identified the lifting components of carrying, push/pull, as well as reaching and grip strength as

most difficult, which is supported by findings from the development of an FCE taxonomy for upper limb disorders (Reneman et al. 2005).

The levels of agreement between the FCE components and the DOT physical demands were high in this study. Sixteen out of the 18 components had 100% agreement, which gives support to the WorkHab FCE components having acceptable content validity when using the DOT taxonomy as the basis. The DOT taxonomy has also been used to determine content validity in other tools (Fishbain et al. 1994; Lechner et al. 1994; Innes and Straker 1999b; Gibson et al. 2005).

In the study that investigated construct (convergent) validity, there was a significant relationship with an increase in muscle activity with increasing weight lifted in the upper trapezius, mid deltoid, thoracic, brachioradialis and biceps muscles during a bench to shoulder lift of the WorkHab FCE. The study also identified that there was a significant difference between muscle activity in the up lift (bench to shoulder height) and the down lift (shoulder to bench height) in these muscles with the exception of the brachioradialis, which, as an elbow flexor muscle, may be the result of the elbows being flexed during both phases of the lift. The grip required to lift the boxes used as part of the WorkHab FCE may also have impacted upon the activity of the brachioradialis muscle. Despite these differences, no significant differences were found in the recorded physiological parameters, either in individual muscles or when comparing the relationships between the muscles immediately before or after the determination of the SML as had been hypothesized. Therefore, using these SEMG physiological parameters, the construct (convergent) validity of the SML during the bench to shoulder lift of the WorkHab FCE was unable to be

established in this study. Further studies to investigate other muscle involvement, particularly the muscles in the spine are recommended.

Chapter six presents a general discussion of the findings outlined in this thesis, the limitations of the studies, outlines future research directions and implications for clinical practice.

Chapter 6 Discussion

Overview: So far the thesis has discussed FCEs in general, reported on the current practice of health professionals on FCE usage, particularly to determine the level of usage of the WorkHab FCE, and presented studies that investigated aspects of reliability and validity of the WorkHab FCE. This chapter discusses the key research findings from completed studies and the limitations of this research. Implication of the results to health professionals who work in occupational rehabilitation settings and future directions for research in this area is also considered.

6.1 Key research findings within the thesis

Evidence about the psychometric properties of assessment tools used in occupational rehabilitation is needed to provide confidence to health professionals and other key stakeholders who rely upon these tools to present reliable and valid findings, and to inform best practice. In the current Australian occupational rehabilitation climate, where there is a large focus on litigation, evidence for the use of specific assessment tools is required to allow appropriate decisions to be made based on accurate and meaningful FCE results. The current usage of FCEs in the Australian context was investigated, which identified and confirmed that the WorkHab FCE is a commonly used tool in this environment. Following from this, studies related specifically to the reliability and validity of the WorkHab FCE were completed.

Current usage of FCEs

Using a population of health professionals in NSW, Australia, one of the studies conducted as part of this thesis identified the most commonly used FCE was a

non-standardised version. The popularity of non-standardised assessment use, in this environment, highlights that health professionals may not be using or have access to available evidence to inform their practice. Reliance upon non-standardised assessments is inconsistent with professional imperatives to use evidence based practice including the use of sound outcome measures to provide accurate and meaningful results that can be argued and justified within the legal system. The WorkHab FCE was the most popular standardised tool used by this population. This result provided justification to focus subsequent studies on the psychometric properties of the WorkHab FCE.

The studies conducted found that FCEs are applied differently dependent upon the reason for referral and client goal, the health professional's workplace procedures, policies and resources and the health professional's skill and experience. Health professionals also discussed adapting FCEs for specific purposes and to meet the goal of the assessment. This concurs with other research that identified different types of FCE (Innes and Straker 2002b; Pransky and Dempsey 2004). The adaptation of an assessment tool has implications for the reliability and validity of the assessment tool; however, this further highlights the issue that health professionals may not use evidence to inform their practice. Personal skills and experience of the health professional was identified as an important consideration and concurs with previous research findings that identified that health professionals need to have appropriate knowledge, training and expertise to conduct FCEs (Abdel-Moty et al. 1993; King et al. 1998; Innes and Straker 2002b; Innes and Straker 2002a; Chappell et al. 2006; Isernhagen 2009). However, the use of non-standardised

tools and adaptations to FCEs suggests knowledge and training is not informing evidence based practice in this area.

Clinical utility

Clinical utility for FCEs in general was investigated and provided information to assess the usefulness of the WorkHab FCE compared to FCEs in general. Clinical utility reflects the degree of conviction health professionals have about the usefulness of an assessment (Toomey et al. 1995) and confirms that the tool is related to the purpose for which it is used (Barbara and Whiteford 2005). In this study, participants indicated that they obtained relevant and useful clinical information from a FCE. The usefulness of FCEs was also reported in a study of 'return to work' case managers who perceived the FCEs to be a useful tool in the management of clients (Wind et al. 2006). However, despite health professionals indicating in this study that they use the results to predict return to work, predictive benefits of FCEs are inconclusive (Gross and Battie 2004; Gross et al. 2004). Health professionals in this study valued the flexibility of FCEs and chose to adapt them to suit the client, injury type and job, rather than use standardised measures, as has been found in previous studies (King et al. 1998; Innes and Straker 2003b; Innes and Straker 2003a; Strong et al. 2004b).This suggests that health professionals value the ability to adapt an FCE more than the standardised properties of the tool, which may compromise the reliability and validity of the assessment tool. This could be the result of health professionals lacking an understanding about the psychometric properties of assessment tools (Clemson and Fitzgerald 1998). Knowledge of published literature was rated as moderately important in this study, which suggests that health professionals base decisions on historical information and practices

rather than current evidence and research. Access to available evidence, lack of technical skills to search databases and a lack of knowledge of the critical appraisal of literature may also inhibit the application of evidence in practice (McCluskey and Cusick 2002). It is, however, hoped that undergraduate and post graduate education is providing students with the skills to search databases and critically analyse literature which when coupled with an appropriate access to materials will enhance the use of current evidence in practice.

Reliability

Reliability is the consistency of a measure (Portney and Watkins 2009). Test-retest, inter-rater and intra-rater reliability have been judged most important for FCEs with agreement between raters at the same time or over time being important. If results are too variable, results of the measurement of individual function will not be meaningful (Innes and Straker 1999a). Test-retest reliability was studied for the manual handling component of the WorkHab FCE with the level of rater agreement high for all lifts; however, the 95% limits of agreement found a variation of ± 3 to 4kg in weight lifted at the safe maximal lift, although a clinically more acceptable ± 2.5kg at the safe maximal lift in the majority of cases. This suggests that the administration procedures for the WorkHab FCE are dependable and the average performance by the subjects was relatively stable over the study. Similar types of lifting tasks evaluated in other FCEs such as the Ergo-kit, Isernhagen Work System and the Physical Work Performance Evaluation, have reported substantial or acceptable test-retest reliability (Reneman et al. 2002b; Tuckwell et al. 2002; Reneman et al. 2004; Gouttebarge et al. 2006; Legge and Burgess-Limerick 2007; Gibson et al.

2010). Intra-rater reliability for the manual handling component of the WorkHab FCE was also good, where an ICC of 0.90 or more was considered a measure of excellent reliability, an ICC of 0.75 – 0.90 was considered good and an ICC of less than 0.75 was considered moderate to poor (Portney and Watkins 2009). This indicates that health professionals can make consistent judgements on the manual handling scoring system and in determination of safe maximal lifts in this FCE. Similar types of lifting tasks evaluated in other FCEs such as Isernhagen Work System, Job Fit, Physical Work Performance Evaluation and Ergo-kit have reported substantial or acceptable levels of intra-rater reliability (Reneman et al. 2002b; Durand et al. 2004; Gouttebarge et al. 2005; Legge and Burgess-Limerick 2007). These studies demonstrate that the WorkHab FCE is a reliable measure when the same person acts as the assessor.

Inter-rater reliability for the manual handling score and for each of its components (stance, posture, leverage, torque and pace) was good to excellent, based on the ICC criteria noted above. The inter-rater reliability findings are in line with other studies that investigated inter-rater reliability of FCEs and support the reliability of rater agreement (Lechner et al. 1994; Gross and Battie 2002; Durand et al. 2004; Gouttebarge et al. 2006; Legge and Burgess-Limerick 2007).

Validity

Validity relates to the accuracy of an assessment and is concerned with the extent or degree to which an instrument measures what it is intended to measure (Innes and Straker 1999b; Portney and Watkins 2009). Content validity of the WorkHab FCE was studied and included analysis of item relevance, item difficulty and the relationship of the FCE components to the

physical demands of the DOT. Item relevance for manual work and for vocational retraining was high, which supports content validity of the WorkHab FCE for these purposes. However, the relevance of the WorkHab FCE components to sedentary work was less convincing, especially in relation to the items: lift bench to overhead, push/pull, climb, balance, kneel and crawl. Other studies have investigated content validity using expert panels with positive results (Gibson and Strong 2002; Sandqvist et al. 2008), although content validity for type of work was not reported in these studies. The relationship of the WorkHab FCE components to the equivalent items on the DOT also demonstrated positive results. The DOT taxonomy has been used as a basis to establish content validity in other FCE studies (Fishbain et al. 1994; Lechner et al. 1994; Innes and Straker 1999b) with similar results (Gibson et al. 2005). Some of the DOT items had poor agreement with the WorkHab FCE components (sensory items); however, the qualitative comments suggested that these items were considered useful to evaluate communication ability but were not necessarily essential. This supports the finding that health professionals are willing to adapt the FCE to suit the requirements, whether this is client or work-related as has been found in other FCE studies (King et al. 1998; Innes and Straker 2003b; Innes and Straker 2003a; Strong et al. 2004b). Results for item difficulty in relation to clients with low back injuries and upper limb injuries were consistent with the findings of other FCE research, specifically the GAPP-FCE and an upper limb FCE taxonomy (Kersnovske et al. 2005; Reneman et al. 2005).

Construct validity is the degree to which a theoretical construct is measured by an assessment (Portney and Watkins 2009). This study investigated the

construct (convergent) validity where two measures believed to reflect the same phenomenon are expected to be highly correlated. It was hypothesised that surface electromyography recordings would be able to correlate the muscle activity with the estimation of a SML during the bench to shoulder lift. It was hypothesised that there would be a peak of muscle activity as subjects reached the SML, or a change in the relationship between the various muscles involved at this point, as subjects recruited different muscles to compensate for the increase in weight. A significant relationship between the weight lifted and muscle activity during a bench to shoulder lift of the WorkHab FCE was identified, showing a gradual increase in muscle activity in all muscles as weight increased, as was expected. Despite a significant difference between muscle activity in the up lift (bench to shoulder height) and the down lift (shoulder to bench height), no significant differences were found in the recorded physiological parameters immediately before or after the SML. Using these SEMG physiological parameters, construct (convergent) validity of the SML during the bench to shoulder lift of the WorkHab FCE was unable to be established. SML is determined for the bench to shoulder lift based on observation of the recruitment of upper extremity strength for the ability to control the lift and stabilise the lumbo-sacral spine without hyperextension (Bradbury and Roberts 1998), plus observation of muscle bulging of prime movers, involuntary use of accessory muscles, altered body mechanics, loss of equilibrium, increased base of support, decreased efficiency and smoothness of movement, and cardiovascular signs of heart rate and breathing patterns (Gross and Battie 2002). As the results of SEMG recordings in this study were unable to be correlated with the SML, this technique could not quantify this ‘end

point'. No other studies that investigated SEMG responses to lifting during an FCE were located. Reliability of safe maximal lifting has been studied with acceptable levels of agreement (Smith 1994; Gardener and McKenna 1999; Isernhagen et al. 1999; Reneman et al. 2002b); however, evidence of the quantification of the SML is not available, which suggests further investigation to identify parameters that can quantify this 'end point' of a SML is needed.

6.2 Limitations within the program of research for the thesis

Research of the current usage of FCEs was limited to NSW, Australia, with a small sample size for the qualitative component and a small response rate for the cross sectional survey study. Therefore results cannot be generalised beyond this group. The response rate to the inter- and intra-rater study was small even though all accredited evaluators of the WorkHab FCE, both nationally and internationally, were invited to participate. The response rate to the content validity survey was also small, despite widespread invitations to participate again both nationally and internationally. Difficulties with participant recruitment to research is noted and discussed in chapter two. Limiting factors that impede clinicians (as participants) involvement in research may include a lack of time, heavy caseloads or time spent in other research activities (Weierbach et al. ; Kadushin 2001; Serxner et al. 2004; Lannin and Cusick 2006). In the area of occupational rehabilitation in Australia, health professionals have workload targets to meet that, when combined with a lack of time and heavy workloads, may impact upon their involvement in research activities. However, the value placed on research and the use of research to inform evidence based practice may not be high for these health professionals as a group, which is supported by the findings in the study on current usage of

FCEs where personal knowledge of published literature was identified as only moderately important. As has been found in previous studies, those clinicians with research qualifications (Lannin and Cusick 2006) were more effective at recruiting patients to research studies and that suggests these professionals are more aware of, and place more importance on, research as evidence to inform practice.

More Occupational Therapists (OT) than other professionals became involved in all aspects of this research. As an occupational therapist was conducting the research this may have influenced who responded to the research. However, it is believed that more OTs are employed in occupational rehabilitation than other disciplines in NSW, Australia, despite any direct evidence to support this. An approach was made to the largest provider of occupational rehabilitation in Australia to obtain data on the numbers of OTs, physiotherapists and other health professionals involved in occupational rehabilitation, but this information is classified as 'commercial in-confidence' and was therefore unable to be obtained. It is therefore not possible to estimate the degree of representativeness of the findings or the response rates for some of the studies conducted as part of this thesis.

Another limitation to the research was the use of healthy individuals for the test-retest reliability study and for the construct validity study. Other studies that investigated FCEs have also used a healthy population (Ting et al. 2001; Reneman et al. 2004; Soer et al. 2006a) because healthy individuals are often chosen for convenience and as a result of difficulty accessing an injured worker population. Healthy participants overcome the possibility of results being affected by changes in injury or recovery status which may affect performance

between testing sessions. In Australia, injured workers are usually being managed in a litigious workers compensation system making ethical recruitment and consent processes more complex. The perceived potential for aggravation of an injury or re-injury may also discourage injured workers from agreeing to participate in research. Issues with the recruitment of patients in research are well documented (Weierbach et al. ; Kadushin 2001; Serxner et al. 2004; Lannin and Cusick 2006).

Another limitation is the use of video-footage for the inter-rater and intra-rater reliability study. This was implemented for the study due to practical constraints of more than one rater present at the time of a FCE, which may create a more artificial environment that is not replicated in clinical situations. As a result, health professional raters reviewed video-footage of lift segments from two angles (rear and side). The disadvantage of this was that raters did not have the three dimensional vision that would be present in the clinical setting and did not have the opportunity to observe other clinical prompts such as verbal cues or to read facial expressions, as would be the case in the usual clinical assessment. Other studies to evaluate rater agreement have used real time (Lechner et al. 1994; Gross and Battie 2002; Durand et al. 2004; Gouttebarge et al. 2006) and video recordings (Isernhagen et al. 1999; Reneman et al. 2002b; Legge and Burgess-Limerick 2007) with both finding substantial or acceptable levels of inter-rater agreement.

6.3 Implications of thesis findings for clinical practice

The findings from this research indicate that health professionals value adaptability and flexibility of FCE tools and will choose to adapt the tool in

accordance with a client's injury and job rather than use standardised measures. However, the potential for a negative impact on any reliability and validity of an assessment must be determined when adaptations are made. Similarly, when health professionals use parts rather than the whole FCE, the standardised properties of the FCE are compromised along with any reliability and validity of the tool. Use of modified or non-standardised assessment tools may reduce the confidence that health professionals and other stakeholders have about the accuracy and meaningfulness of results. The credibility of the health professional and of the assessment results may also be damaged in any litigation case.

The value placed on adaptability and flexibility of FCEs may also be related to a lack of understanding of the psychometric properties of tools and, as Clemson and Fitzgerald found, issues of reliability and validity were not clearly understood by therapists (Clemson and Fitzgerald 1998). This coupled with the lack of published literature of the psychometric properties of specific tools supports the need for ongoing research in this area. Health professionals also rated a personal knowledge of published literature as only moderately important for their practice, which suggests that health professionals use historical information and practices as the basis for decisions rather than evidence based practice and current research. Evidence based practice involves clinical reasoning to integrate the clinical experience of the health professional, any preferences of the client, the resources available and the highest quality of evidence available (Bennett and Bennett 2000). Levels of evidence for practice are identified as ranging from level A or 1a (such as evidence from a randomised control trial) to level D or 5 (expert opinion without explicit critical

appraisal) and therefore it is acknowledged that expert opinion or clinical experience is valuable when considering evidence based practice, but this is less valued than results from research studies (Phillips et al. 2002). Health professionals need to evaluate all the attributes of FCEs including safety, reliability, validity, practicality and utility, to ensure high quality standards of practice. Health professionals also need to review and update their knowledge to provide the best possible evidence based care for clients. To access evidence as part of practice, health professionals need to have the skills to search databases and to appraise the available literature, and organisations need to promote the use of evidence and provide access to appropriate resources. However, the evidence also needs to be available. As discussed throughout this thesis, evidence for the psychometric properties of specific FCEs is limited and the studies completed as part of this project build evidence for the WorkHab FCE. As interventions and services need to be justified with evidence, which is required both by health professions as a whole and by those who pay for services (Muir Gray 1997), this is a key skill for health professionals to demonstrate and an important aspect for continuing professional development.

The difficulties in recruitment of participants and clinicians in research have been noted; nonetheless, this does impact on clinical practice. If research activities are limited, the evidence for practice is also limited. Health professionals need to be aware of the importance of research to contribute to this evidence base, and they need to be encouraged to become involved in research both as a clinician and to assist in the recruitment of injured participants to studies. Explicitly linking research to continuing professional

development requirements is a strategy that warrants further investigation, particularly as we enter a new chapter that includes national registration of Occupational Therapists in Australia. Maintenance of currency of skills and knowledge that is evaluated alongside registration and continuing professional development requirements provides an opportunity for the value of research to be highlighted to clinicians. This process is one way that health professionals could be encouraged to be involved in research activity with a direct and tangible benefit to them.

The results of the reliability studies demonstrate that the WorkHab FCE is a reliable measure both when the same person acts as the assessor and when different health professionals assess. Consistent findings were found both in the test-retest and intra- and inter-rater reliability for the manual handling component of this assessment, which gives confidence to clinicians and service users in the consistency of this component of the measure. However, further studies to investigate other aspects are needed to demonstrate reliability across all components of the WorkHab FCE.

Content validity for the WorkHab FCE is supported for manual work and for vocational retraining, but was less strong for sedentary work. Health professionals who use the WorkHab FCE in practice need to consider if the WorkHab FCE is the most appropriate tool to use for those injured workers returning to sedentary work; however, health professionals can have confidence that the WorkHab FCE has appropriate content for use with those clients involved in manual work or for those undertaking vocational retraining.

Construct validity of the SML during the bench to shoulder lift of the WorkHab FCE was unable to be established using the physiological parameters of SEMG because no significant differences were found in the recorded physiological parameters before or after the safe maximal lift. Health professionals should be aware that there was a significant relationship between the weight lifted and muscle activity, with a gradual increase of muscle activity as weight increased during the bench to shoulder lift. There was also a significant difference between muscle activity in the up lift (bench to shoulder height) and the down lift (shoulder to bench height), which suggests participants had to work harder to lift a weight against gravity than lowering in the direction of gravity. However, further studies to quantify the determinants of a SML using different parameters are required. The clinical judgement skills used by health professionals to determine SML and to determine the ratings for the manual handling scoring system of the WorkHab FCE is another aspect for further study. As a result of these findings, health professionals are encouraged to consider what clinical reasoning they use in these determinations.

6.4 Future directions for research

This research investigated the reliability and aspects of the validity of the WorkHab FCE, but additional aspects of validity, including criterion related validity (predictive and concurrent) and construct validity (convergent and discriminative), require evaluation to further determine psychometric properties of this tool. A range of populations, including injured workers, needs to be studied in relation to the use of this tool across different populations. This research focussed on the manual handling components of the WorkHab FCE, and further studies to investigate the other components (such as the sitting,

standing, climbing, reaching and balancing) of the tool are recommended. This research identified that the WorkHab FCE demonstrated content validity for manual work and vocational retraining, but further research of the application of the WorkHab FCE to sedentary work is required. This would include research to test the relevance of the following FCE items: lift bench to overhead, push/pull, climb, balance, kneel and crawl to sedentary work.

Another area for future research relates to the evaluation of a safe maximal lift. As previously discussed, no significant differences were found in the recorded physiological parameters before or after the SML in this research and thus using these SEMG physiological parameters convergent validity was unable to be established. However, the quantification of the SML in clinical practice requires further investigation and a focus on other factors used to determine the SML end point such as the biomechanical patterns of lifting during the FCE is recommended. The specific clinical reasoning processes used by health professionals in practice to determine a SML and to determine the different manual handling ratings within the WorkHab manual handling scoring system also warrants further investigation.

6.5 Conclusions

This study intended to evaluate the level of usage of the WorkHab FCE in occupational rehabilitation practice in Australia and it identified that the WorkHab FCE is commonly used in this environment. This justified the subsequent studies to determine the reliability and validity of the WorkHab FCE. Important measurement properties of the WorkHab FCE were investigated. Rater judgements: test-retest, inter- and intra-rater reliability of the manual

handling components of this tool appear to have acceptable levels of reliability. Research findings support content validity for the WorkHab FCE, specifically in relation to manual work and for vocational retraining purposes; however, construct (concurrent) validity of the SML during the bench to shoulder lift was unable to be established using physiological measures, and further investigation into this and other forms of validity is recommended.

Evidence of reliability and validity is essential for accurate and meaningful measurement of any evaluation including FCEs. This study investigated reliability and aspects of validity specifically for the WorkHab FCE with promising results. The publication of these studies will disseminate the results so that health professionals can access and be aware of this evidence of the WorkHab FCE. Further research to investigate other aspects of reliability and validity of the WorkHab FCE is needed for health professionals and service users to have confidence of consistency and of all the measurement properties of this tool. This is also needed for other commercially available FCE tools. The results from this research provide evidence of reliability and content validity for users of the WorkHab FCE. Research based and appropriate assessment tools are considered essential to ensure credible practice. There is an expectation by service users that health professionals use evidence to inform practice. In relation to FCE use, this can be questioned when the lack of evidence relating to the properties of many FCEs is considered. This further supports the need for ongoing research into this area of practice. The use of tools without evidence of psychometric properties has the potential for stakeholders to doubt the value of the information provided and to threaten the credibility of health professionals in the medico-legal arena should they need to justify their conclusions and

interventions in the litigious environment of the workers compensation system in Australia.

Health professionals are encouraged to become familiar with the psychometric properties of the WorkHab FCE and to evaluate the available evidence in relation to the specific tasks, jobs and workplaces applicable to their clients. This research contributes to, and provides evidence of reliability and content validity for the WorkHab FCE, adding assurance to the accuracy of measurements gained and how this tool can confidently be used in practice.

References

Abdel-Moty, E., Fishbain, D., Khalil, T., Sadek, S., Cutler, R., Steele-Rosomoff, R. & Rosomoff, H. (1993). "Functional Capacity and Residual Functional Capacity and Their Utility in Measuring Work Capacity." Clinical Journal of Pain **9**(3): 168-173.

American Journal of Occupational Therapy (2000). "Occupational Therapy Services in Facilitating Work Performance (Statement)." American Journal of Occupational Therapy **54**(6): 626-628.

Arjmand, N.& Shirazi-Adl, A. (2005). "Biomechanics of Changes in Lumbar Posture in Static Lifting." Spine **30**(23): 2637-48.

Barbara, A.& Whiteford, G. (2005). "Clinical Utility of the Handicap Assessment and Resource Tool: An Investigation of Its Use with the Aged People in Hospital." Australian Occupational Therapy Journal **52**(1): 17-25.

Barry, C., Britten, N., Barber, N., Bradley, C.& Stevenson, F. (1999). "Using Reflexivity to Optimize Teamwork in Qualitative Research." Qualitative Health Research **9**(1): 26-44.

Bennett, S. & Bennett, J. (2000). "The Process of Evidence- Based Practice in Occupational Therapy: Informing Clinical Decisions." Australian Occupational Therapy Journal **47**: 171-180.

Bland, J.& Altman, D. (1999). "Measuring Agreement in Method Comparison Studies." Statistical Methods in Medical Research **8**(2): 135-160.

Bloom, B. (1976). Human Characteristics and School Learning. New York, McGraw-Hill.

Boadella, J., Sluiter, J.& Frings-Dresen, M. (2003). "Reliability of Upper Extremity Tests Measured by the Ergos Work Simulator: A Pilot Study." Journal of Occupational Rehabilitation **13**(4): 219-232.

Borg, G. (1978). Subjective Effort in Relation to Physical Performance and Working Capacity. Psychology: From Research to Practice. H. Pick. New York, Plenum Publishing**:** 333-361.

Bowling, A. (2004). Research Methods in Health; Investigating Health and Health Services. London, McGraw Hill.

Bradbury, S.& Roberts, D. (1998). Workhab Functional Capacity Evaluation Procedural Manual. WorkHab Australia.

Branton, E., Arnold, K., Appelt, S., Hodges, M., Battie, M. & Gross, D. (2010). "A Short-Form Functional Capacity Evaluation Predicts Time to Recovery but Not Sustained Return-to-Work." Journal of Occupational Rehabilitation: 1-7.

Brouwer, S., Dijkstra, P., Stewart, R., Goeken, L., Groothoff, J.& Geertzen, J. (2005). "Comparing Self Report, Clinical Examination and Functional Testing in the Assessment of Work-Related Limitations in Patients with Chronic Low Back Pain." Disability and Rehabilitation **27**(17): 999-1005.

Brouwer, S., Reneman, M., Dijkstra, P., Groothoff, J., Schellekens, J. & Goeken, L. (2003). "Test-Retest Reliability of the Isernhagen Work Systems Functional Capacity Evaluation in Patients with Chronic Low Back Pain." Journal of Occupational Rehabilitation **13**(4): 207-18.

Brubaker, P., Fearon, F., Smith, S., McKibben, R., Alday, J., Andrews, S., Clarke, E. & Shaw, G. (2007). "Sensitivity and Specificity of the Blankenship FCE System's Indicators of Submaximal Effort." Journal of Orthopaedic & Sports Physical Therapy **37**(4): 161-8.

Bruce, N., Pope, D. & Stanistreet, D. (2008). Quantitative Methods for Health Research: A Practical Interactive Guide to Epidemiology and Statistics. Chicester, UK: Wiley.

Carpenter, C.& Hammell, K. (2000). Evaluating Qualitative Research. Using Qualitative Research: A Practical Introduction for Occupational and Physical Therapists. K. Hammell, Carpenter, C.& Dyck, I. Edinburgh, Churchill Livingstone**:** 107-119.

Chappell, I., Henry, A., McLean, A., Richardson, M. & Shivji, M. (2006). The Functional Capacity Evaluation: A Clinicians Guide. Ottawa, CAOT Publications ACE.

Clemson, L.& Fitzgerald, M. (1998). "Understanding Assessment Concepts within the Occupational Therapy Context." Occupational Therapy International **5**(1): 18-34.

Cole, M., Grimshaw, P.& Burden, A. (2004). "Loads on the Lumbar Spine During a Work Capacity Assessment Test." Work **23**(2): 169-78.

Cook, C., Burgess-Limerick, R.& Papalia, S. (2004). "The Effect of Upper Extremity Support on Upper Extremity Posture and Muscle Activity During Keyboard Use." Applied Ergonomics **35**(3): 285-292.

Cotton, A., Schonstein, E.& Adams, R. (2006). "Use of Functional Capacity Evaluations by Rehabilitation Providers in NSW." Work **26**(3): 287-295.

Davis, K., Jorgensen, M.& Marras, W. (2000). "An Investigation of Perceived Exertion Via Whole Body Exertion and Direct Muscle Force Indicators During the Determination of the Maximum Acceptable Weight of Lift." Ergonomics **43**(2): 143 - 159.

Davis, S. and Madden, S. (2006). The International Classification of Function and Health. Rehabilitation: The Use of Theories and Models in Practice. S. Davis. London, Churchill Livingstone

Deen, M., Gibson, L.& Strong, J. (2002). "A Survey of Occupational Therapy in Australian Work Practice." Work **19**(3): 219-230.

Depoy, E.& Gitlin, L. (1998). Introduction to Research. Understanding and Applying Multiple Strategies. St Louis, Mosby.

Durand, M., Brassard, B., Hong, Q., Lemaire, J.& Loisel, P. (2008). "Responsiveness of the Physical Work Performance Evaluation, a Functional Capacity Evaluation, in Patients with Low Back Pain." Journal of Occupational Rehabilitation **18**(1): 58-67.

Durand, M., Loisel, P., Poitras, S., Mercier, R., Stock, S.& Lemaire, J. (2004). "The Interrater Reliability of a Functional Capacity Evaluation: The Physical Work Performance Evaluation." Journal of Occupational Rehabilitation **14**(2): 119-129.

Dusik, L., Menard, M., Cooke, C., Fairburn, S. & Beach, G. (1993). "Concurrent Validity of the Ergos Work Simulator versus Conventional Functional Capacity Evaluation Techniques in a Workers' Compensation Population." Journal of Occupational Medicine **35**(8): 759-767.

Ekbladh, E., Haglund, L.& Thorell, L. (2004). "The Worker Role Interview - Preliminary Data on the Predictive Validity of Return to Work of Clients after an Insurance Medicine Investigation." Journal of Occupational Rehabilitation **14**(2): 131-141.

Fishbain, D., Abdel-Moty, E., Cutler, R., Khalil, T., Sadek, S., Rosomoff, R.& Rosomoff, H. (1994). "Measuring Residual Functional Capacity in Chronic Low Back Pain Patients Based on the Dictionary of Occupational Titles." Spine **19**(8): 872-880.

Fishbain, D., Cutler, R., Rosomoff, H., Khalil, T., Abdel-Moty, E. & Steele-Rosomoff, R. (1999). "Validity of the Dictionary of Occupational Titles Residual Functional Capacity Battery." Clinical Journal of Pain **15**(2): 102-110.

Fisher, G. (1999). "Administration and Application of the Worker Role Interview: Looking Beyond Functional Capacity." Work **12**(1): 13-22.

Frings-Dresen, M. & Sluiter, J. (2003). "Development of a Job-Specific FCE Protocol: The Work Demands of Hospital Nurses as an Example." Journal of Occupational Rehabilitation **13**(4): 233-48.

Gardener, L.& McKenna, K. (1999). "Reliability of Occupational Therapists in Determining Safe Maximal Lifting Capacity." Australian Occupational Therapy Journal **46**(3): 110-19.

Gatchel, R., Ricard, M., Choksi, D., Mayank, J.& Howard, K. (2009). "The Comprehensive Muscular Activity Profile (CMAP): Its High Sensitivity, Specificity and Overall Classification Rate for Detecting Submaximal Effort on Functional Capacity Testing." Journal of Occupational Rehabilitation **19**(1): 49-55.

Gibson, L. & Strong, J. (1996). "The Reliability and Validity of a Measure of Perceived Functional Capacity for Work in Chronic Back Pain." Journal of Occupational Rehabilitation **6**(3): 159-175.

Gibson, L. & Strong, J. (1997). "A Review of Functional Capacity Evaluation Practice." Work **9**(1): 3-11.

Gibson, L. & Strong, J. (2002). "Expert Review of an Approach to Functional Capacity Evaluation." Work **19**(3): 231-242.

Gibson, L. & Strong, J. (2003). "A Conceptual Framework of Functional Capacity Evaluation for Occupational Therapy in Work Rehabilitation." Australian Occupational Therapy Journal **50**(2): 64-71.

Gibson, L.& Strong, J. (2005). "Safety Issues in Functional Capacity Evaluation: Findings from a Trial of a New Approach for Evaluating Clients with Chronic Back Pain." Journal of Occupational Rehabilitation **15**(2): 237-251.

Gibson, L.& Strong, J. (2006). "Letter to the Editor (Authors' Response).Kersnovske S, Gibson L, Strong J. Item Validity of the Physical Demands from the Dictionary of Occupational Titles for Functional Capacity Evaluation of Clients with Chronic Back Pain. Work 24 (2005), 157-169." Work **26**(2): 231-233.

Gibson, L., Dang, M., Strong, J. & Khan, A. (2010). "Test Retest Reliability of the GAPP Functional Capacity Evaluation in Healthy Adults." Canadian Journal of Occupational Therapy **77**(1): 38-48.

Gibson, L., Strong, J.& Wallace, A. (2005). "Functional Capacity Evaluation as a Performance Measure: Evidence for a New Approach for Clients with Chronic Back Pain." Clinical Journal of Pain **21**(3): 207-15.

Gouttebarge, V., Kuijer, P., Wind, H., van Duivenbooden, C., Sluiter, J. & Frings-Dresen, M. (2009a)."Criterion-Related Validity of Functional Capacity Evaluation Lifting Tests on Future Work Disability Risk and Return to Work in the Construction Industry." Occupational and Environmental Medicine **66**(10): 657-63.

Gouttebarge, V., Wind, H., Kuijer, P. & Frings-Dresen, M. (2004). "Reliability and Validity of Functional Capacity Evaluation Methods: A Systematic Review with Reference to Blankenship, Ergos Work Simulator, Ergo-Kit and Isernhagen Work System." International Archives of Occupational and Environmental Health **77**(8): 527-537.

Gouttebarge, V., Wind, H., Kuijer, P. P., Sluiter, J. K.& Frings-Dresen, M. H. (2005). "Intra- and Interrater Reliability of the Ergo-Kit Functional Capacity Evaluation Method in Adults without Musculoskeletal Complaints." Archives of Physical Medicine and Rehabilitation **86**(12): 2354-2360.

Gouttebarge, V., Wind, H., Kuijer, P., Sluiter, J. & Frings-Dresen, M. (2009b). "Construct Validity of Functional Capacity Evaluation Lifting Tests in Construction Workers on Sick Leave as a Result of Musculoskeletal Disorders." Archives of Physical Medicine and Rehabilitation **90**(2): 302-8.

Gouttebarge, V., Wind, H., Kuijer, P., Sluiter, J.& Frings-Dresen, M. (2006). "Reliability and Agreement of 5 Ergo-Kit Functional Capacity Evaluation Lifting Tests in Subjects with Low Back Pain." Archives of Physical Medicine and Rehabilitation **87**(10): 1365-1370.

Granström, B., Kvarnström, S.& Tiefenbacher, F. (1985). "Electromyography as an Aid in the Prevention of Excessive Shoulder Strain." Applied Ergonomics **16**(1): 49-54.

Gross, D. & Battie, M. (2002). "Reliability of Safe Maximum Lifting Determinations of a Functional Capacity Evaluation." Physical Therapy **82**(4): 364-372.

Gross, D. & Battie, M. (2003). "Construct Validity of a Kinesiophysical Functional Capacity Evaluation Administered within a Worker's Compensation Environment." Journal of Occupational Rehabilitation **13**(4): 287-295.

Gross, D. & Battie, M. (2004). "The Prognostic Value of Functional Capacity Evaluation in Patients with Chronic Low Back Pain: Part 2: Sustained Recovery." Spine **29**(8): 920-4.

Gross, D. & Battie, M. (2005a). "Factors Influencing Results of Functional Capacity Evaluations in Workers' Compensation Claimants with Low Back Pain." Physical Therapy **85**(4): 315-322.

Gross, D. & Battie, M. (2005b). "Functional Capacity Evaluation Performance Does Not Predict Sustained Return to Work in Claimants with Chronic Back Pain." Journal of Occupational Rehabilitation **15**(3): 285-294.

Gross, D. & Battie, M. (2006). "Does Functional Capacity Evaluation Predict Recovery in Workers' Compenstaon Claimants with Upper Extremity Disorders?" Occupational and Environmental Medicine **63**: 404-410.

Gross, D. (2004). "Measurement Properties of Performance-Based Assessment of Functional Capacity." Journal of Occupational Rehabilitation **14**(3): 165-73.

Gross, D., Battie, M. & Asante, A. (2006). "Development and Validation of a Short-Form Functional Capacity Evaluation for Use in Claimants with Low Back Disorders." Journal of Occupational Rehabilitation **16**(1): 53-62.

Gross, D., Battie, M. & Cassidy, J. (2004). "The Prognostic Value of Functional Capacity Evaluation in Patients with Chronic Low Back Pain: Part 1: Timely Return to Work." Spine **29**(8): 914-9.

Gross, D., Battie, M.& Asante, A. (2007). "Evaluation of a Short Form Functional Capacity Evaluation: Less Maybe Best." Journal of Occupational Rehabilitation **17**(3): 422-435.

Hagedorn, R. (2001). Foundations for Practice in Occupational Therapy. London, Churchill Livingstone.

Halpern, M. (2001). "Functional Assessment Taxonomy Relevant to Low-Back Impairments." Journal of Occupational Rehabilitation **11**(3): 201-215.

Hansson, G., Nordander, C., Asterland, P., Ohlsson, K., Strömberg, U., Skerfving, S.& Rempel, D. (2000). "Sensitivity of Trapezius Electromyography to Differences between Work Tasks - Influence of Gap Definition and Normalisation Methods." Journal of Electromyography and Kinesiology **10**(2): 103-115.

Higginbotham, N., Albrecht, G.& Connor, L. (2001). Health Social Science: A Transdisciplinary and Complexity Perspective. Melbourne, Oxford Press.

Holtgrefe, K. & Glenn, T. (2007). Principles of Aerobic Exercise. Therapeutic Exercise: Foundations and Techniques. C. Kisner and Colby, L. Philadelphia, F. A. Davis. **5th Ed:** 231-49.

Horneij, E., Holmstram, E., Hemborg, B., Isberg, P.-E. & Ekdahl, C. (2002). "Inter-Rater Reliability and between-Days Repeatability of Eight Physical Performance Tests." Advances in Physiotherapy **4**(4): 146-160.

Hussey, S., Sabonis-Chafee, B. and Clifford O'Brien, J., Eds. (2007). Introduction to Occupational Therapy. Missouri, Mosby Elsevier.

IJmker, S., Gerrits, E. and Reneman, M. (2003). "Upper Lifting Performance of Healthy Young Adults in Functional Capacity Evaluations: A Comparison of Two Protocols." Journal of Occupational Rehabilitation **13**(4): 297-305.

Imms, C. (2006). "The International Classification of Functioning, Disability and Health: They're Talking Our Language." Australian Occupational Therapy Journal **53**(2): 65-66.

Innes, E. (2006). "Reliability and Validity of Functional Capacity Evaluations: An Update." International Journal of Disability Management Research **1**(1): 135-148.

Innes, E. and Straker, L. (1998a). "A Clinician's Guide to Work-Related Assessments: 1 -- Purposes and Problems." Work **11**(2): 183-189.

Innes, E. and Straker, L. (1998b). "A Clinician's Guide to Work-Related Assessments: 2 -- Design Problems." Work **11**(2): 191-206.

Innes, E. and Straker, L. (1998c). "A Clinician's Guide to Work-Related Assessments: 3 -- Administration and Interpretation Problems." Work **11**(2): 207-219.

Innes, E. and Straker, L. (1999a). "Reliability of Work-Related Assessments." Work **13**(2): 107-124.

Innes, E. and Straker, L. (1999b). "Validity of Work-Related Assessments." Work **13**(2): 125-152.

Innes, E. and Straker, L. (2002a). "Strategies Used When Conducting Work -Related Assessments." Work **19**(1): 149-65.

Innes, E. and Straker, L. (2002b). "Workplace Assessments and Functional Capacity Evaluations: Current Practices of Therapists in Australia." Work **18**(1): 51-66.

Innes, E. and Straker, L. (2003a). "Attributes of Excellence in Work Related Assessments." Work **20**(1): 63-76.

Innes, E. and Straker, L. (2003b). "Workplace Assessments and Functional Capacity Evaluations: Current Beliefs of Therapists in Australia." Work **20**(3): 225-236.

Isernhagen, S. (1992). "Functional Capacity Evaluation:Rationale, Procedure, Utility of the Kinesiophysical Approach." Journal of Occupational Rehabilitation **2**(3): 157-168.

Isernhagen, S. (2009). "Functional Capacity Testing: What's New? What's Different?" The Interdisciplinary Journal of Rehabilitation **June**: 20-23.

Isernhagen, S., Hart, D.& Matheson, L. (1999). "Reliability of Independent Observer Judgments of Level of Lift Effort in a Kinesiophysical Functional Capacity Evaluation." Work **12**(2): 145-150.

James, C. & Mackenzie, L. (2009a). "The Clinical Utility of Functional Capacity Evaluations: The Opinion of Health Professionals Working within Occupational Rehabilitation." Work **33**(3): 231-9.

James, C. & Mackenzie, L. (2009b). "Health Professional's Perceptions and Practices in Relation to Functional Capacity Evaluations: Results of a Quantitative Survey." Journal of Occupational Rehabilitation **19**(2): 203-211.

James, C., Mackenzie, L. & Higginbotham, N. (2007). "Health Professionals' Attitudes and Practices in Relation to Functional Capacity Evaluations." Work **29**(2): 81-88.

James, C., Mackenzie, L.& Capra, M. (2010). "Test Retest Reliability of the Manual Handling Component of the Workhab Functional Capacity Evaluation in Healthy Adults." Disability & Rehabilitation **32**(22): 1863-1869.

Jay, M., Lamb, J., Watson, R., Young, I., Fearon, F., Alday, J.& Tindall, A. (2000). "Sensitivity and Specificity of the Indicators of Sincere Effort of the EPIC Lift Capacity Test on a Previously Injured Population." Spine **25**(11): 1405-12.

Jensen, C., Finsen, L., Hansen, K.& Christensen, H. (1999). "Upper Trapezius Muscle Activity Patterns During Repetitive Manual Material Handling and Work with a Computer Mouse." Journal of Electromyography and Kinesiology **9**(5): 317-325.

Jones, T.& Kumar, S. (2003). "Functional Capacity Evaluation of Manual Materials Handlers: A Review." Disability and Rehabilitation **25**(4-5): 179-191.

Jundt, J.& King, P. (1999). "Work Rehabilitation Programs: A 1997 Survey." Work **12**(2): 139-144.

Kadushin, G. (2001). "Clinical Trails: A Wider Lens 2." Health and Social Work **26**(3): 203-6.

Kersnovske, S., Gibson, L.& Strong, J. (2005). "Item Validity of the Physical Demands from the Dictionary of Occupational Titles for Functional Capacity Evaluation of Clients with Chronic Low Back Pain." Work **24**(2): 157-169.

Kilmer, D., Aitkens, S., Wright, N. & McCrory, M. (2000). "Simulated Work Performance Tasks in Persons with Neuropathic and Myopathic Weakness." Archives of Physical Medicine and Rehabilitation **81**(7): 938-942.

King, P., Tuckwell, N.& Barrett, T. (1998). "A Critical Review of Functional Capacity Evaluations." Physical Therapy **78**(8): 852.

Konz, S. (1982). "NIOSH Lifting Guidelines." American Industrial Hygiene Association Journal **43**(12): 931-933.

Kuijer, W., Brouwer, S., Reneman, M., Dijkstra, P., Groothoff, J., Schellekens, J. & Geertzen, J. (2006a). "Matching FCE Activities and Work Demands: An Exploratory Study." Journal of Occupational Rehabilitation **16**(4): 469-483.

Kuijer, W., Dijkstra, P., Brouwer, S., Reneman, M., Groothoff, J.& Geertzen, J. (2006). "Safe Lifting in Patients with Chronic Low Back Pain: Comparing FCE Lifting Task and NIOSH Lifting Guidelines." Journal of Occupational Rehabilitation **16**(4): 579-589.

Lannin, N. & Cusick, A. (2006). "Factors Affecting Patient Recruitment in an Acute Rehabilitation Randomized Controlled Trial." American Journal of Occupational Therapy **60**(2): 177-181.

Laursen, B., Søgaard, K.& Sjøgaard, G. (2003). "Biomechanical Model Predicting Electromyographic Activity in Three Shoulder Muscles from 3D Kinematics and External Forces During Cleaning Work." Clinical Biomechanics **18**(4): 287-295.

Law, M. (1998). Client- Centered Occupational Therapy. Thorofare, Slack.

Law, M. (2002). Evidence Based Rehabilitation: A Guide to Practice. Thorofare, Slack.

Law, M., Baum, C. & Dunn, W. (2001). Measuring Occupational Performance: Supporting Best Practice in Occupational Therapy. Thorofare, Slack.

Lawshe, C. (1975). "A Quantitative Approach to Content Validity." Personnel Psychology **28**(4): 563-575.

Lechner, D., Jackson, J., Roth, D.& Straaton, K. (1994). "Reliability and Validity of a Newly Developed Test of Physical Work Performance." Journal of Occupational Medicine **36**(9): 997-1004.

Lechner, D., Page, J. & Sheffield, G. (2008). "Predictive Validity of a Functional Capacity Evaluation: The Physical Work Performance Evaluation." Work **31**(1): 21-25.

Lee, G. & Chan, C. (2003). "Use of Dictionary of Occupational Titles (DOT) on Formwork Carpentry-a Comparison between the United States and Hong Kong." Work **20**(2): 103-10.

Lee, G., Chan, C. & Hui-Chan, C. (2001a). "Consistency of Performance on the Functional Capacity Assessment: Static Strength and Dynamic Endurance." American Journal of Physical Medicine and Rehabilitation **80**(3): 189-95.

Lee, G., Chan, C. & Hui-Chan, C. (2001b). "Work Profile and Functional Capacity of Formwork Carpenters at Construction Sites." Disability and Rehabilitation **23**(1): 9-14.

Legge, J.& Burgess-Limerick, R. (2007). "Reliability of the Jobfit System Pre-Employment Functional Assessment Tool." Work **28**(4): 299-312.

Lemstra, M., Olszynski, W. & Enright, W. (2004). "The Sensitivity and Specificity of Functional Capacity Evaluations in Determining Maximal Effort: A Randomized Trial." Spine **29**(9): 953-9.

Lygren, H., Dragesund, T., Joensen, J., Ask, T. & Moe-Nilssen, R. (2005). "Test-Retest Reliability of the Progressive Isoinertial Lifting Evaluation (PILE)." Spine **30**(9): 1070-1074.

Lynn, M. (1986). "Determination and Quantification of Content Validity." Nursing Research **35**(6): 382-5.

Lysaght, R. (1997). "Job Analysis in Occupational Therapy: Stepping into the Complex World of Business and Industry." American Journal of Occupational Therapy **51**(7): 569-75.

Maher, H. (2006). "The Functional Capacity Evaluation." AAOHN Journal **54**(9): 420-420.

Marras, W., Granata, K., Davis, K., Allread, W.& Jorgensen, M. (1999). "Effects of Box Features on Spine Loading During Warehouse Order Selecting." Ergonomics **42**(7): 980-996.

Matheson, L. (1996). "Relationships among Age, Body Weight, Resting Heart Rate and Performance in a New Test of Lift Capacity." Journal of Occupational Rehabilitation **6**(4): 225-237.

Matheson, L. (2003). The Functional Capacity Evaluation. Disability Evaluation. G. Andersson, Demeter, S. and Smith, G. Chicago, IL, Mosby Yearbook.

Matheson, L., Isernhaben, S. and Hart, D. (2002a). "Relationships among Lifting Ability, Grip Force, and Return to Work." Physical Therapy **82**(3): 249.

Matheson, L., Kaskutas, V., McCowan, S., Shaw, H. and Webb, C. (2001). "Development of a Database of Functional Assessment Measures Related to Work Disability." Journal of Occupational Rehabilitation **11**(3): 177-199.

Matheson, L., Rogers, L., Kaskutas, V. and Dakos, M. (2002b). "Reliability and Reactivity of Three New Functional Assessment Measures." Work **18**(1): 41-50.

Mathiowetz, V. (1993). "Role of Physical Performance Component Evaluations in Occupational Therapy Functional Assessment." American Journal of Occupational Therapy **47**(3): 225-30.

Matoushek, N. (2008). "Evaluating Functional Capacity." The Interdisciplinary Journal of Rehabilitation **21**(8): 30-33.

McBride, J., Larkin, T., Dayne, A., Haines, T.& Kirby, T. (2010). "Effect of Absolute and Relative Loading on Muscle Activity During Stable and Unstable Squatting." International Journal of Sports Physiology and Performance. **5**(2): 177-83.

McCluskey, A. & Cusick, A. (2002). "Strategies for Introducing Evidence-Based Practice and Changing Clinician Behaviour: A Manager's Toolbox." Australian Occupational Therapy Journal **49**: 63-70.

McFadden, S., MacDonald, A., Fogarty, A., Le, S. & Merritt, B. (2010). "Vocational Assessment: A Review of the Literature from an Occupation-Based Perspective." Scandinavian Journal of Occupational Therapy **17**(1): 43-48.

McFadyen, A. & Pratt, J. (1997). "Understanding the Statistical Concepts of Measures of Work Performance" British Journal of Occupational Therapy **60**(6): 279-284.

McGuire, S. (1995). Occupational Rehabilitation and Functional Capacity Evaluation in Australia. Comprehensive Guide to Work Injury Management. S. Isernhagen. Gaithersburg Md, Aspen Publishers.

McKenzie, J., Wood, M., Kotecki, J., Clark, J.& Brey, R. (1999). "Research Notes: Establishing Content Validity: Using Qualitative and Quantative Steps " American Journal of Health Behaviour **23**(4): 311-318.

McMillan, I. (2006). Assumptions Underpining a Biomechanical Frame of Reference in Occupational Therapy. Foundations for Practice in Occupational Therapy. E. Duncan. Edinburgh, Elsevier Churchill Livingstone. **4:** 255-275.

Menard, M. & Hoens, A. (1994). "Objective Evaluation of Functional Capacity: Medical, Occupational, and Legal Settings." Journal of Orthopaedic and Sports Physical Therapy **19**(5): 249-60.

Mitchell, T. (2008). "Utilization of the Functional Capacity Evaluation in Vocational Rehabilitation " Journal of Vocational Rehabilitation **28**(1): 21-28.

Morse, J.& Field, P.(1995). Nursing Research: The Application of Qualitative Approaches. London, Chapman Hall.

Moyers, T. (1999). "The Guide to Occupational Therapy Practice." American Journal of Occupational Therapy **53**: 247-322.

Muir Gray, J. (1997). Evidence-Based Healthcare: How to Make Health Policy and Management Decisions. London, Churchill Livingstone.

Nielsen, P., Andersen, L.& Jørgensen, K. (1998). "The Muscular Load on the Lower Back and Shoulders Due to Lifting at Different Lifting Heights and Frequencies." Applied Ergonomics **29**(6): 445-450.

"NSW Workers Compensation Act (1987)." from http:www.legislation.nsw.gov.au/fullhtml/inforce/act+70+1987+FIRST+0+N?#pt.3-div.2-sec.40a.

Patton, M. (2002). Qualitative Evaluation and Research Methods. Newbury Park, CA, Sage Publications.

Phillips, B., Ball, C., Sackett, D., Haynes, B., Straus, S. & McAlister, F. (2002). Levels of Evidence. London, EBOC - NHS.

Portney, L. & Watkins, M. (2000). Foundations of Clinical Research: Applications to Practice. New Jersey, Prentice Hall Health.

Portney, L.& Watkins, M. (2009). Foundations of Clinical Research: Applications to Practice. Upper Saddle River, New Jersey, Pearson Prentice Hall.

Pransky, G.& Dempsey, P. (2004). "Practical Aspects of Functional Capactiy Evaluations." Journal of Occupational Rehabilitation **14**(3): 217-229.

Ratzon, N., Jarus, T. & Catz, A. (2007). "The Relationship between Work Function and Low Back Pain History in Occupationally Active Individuals." Disability and Rehabilitation **29**(10): 791-796.

Reesink, D., Jorritsma, W.& Reneman, M. (2007). "Basis for a Functional Capacity Evaluation Methodology for Patients with Work-Related Neck Disorders." Journal of Occupational Rehabilitation **17**(3): 436-449.

Reneman, M. & Dijkstra, P. (2003). "Introduction to the Special Issue on Functional Capacity Evaluations: From Expert Based to Evidence Based." Journal of Occupational Rehabilitation **13**(4): 203-6.

Reneman, M., Brouwer, S., Meinema, A., Dijkstra, P., Geertzen, J.& Groothoff, J. (2004). "Test-Retest Reliability of the Isernhagen Work Systems Functional Capacity Evaluation in Healthy Adults." Journal of Occupational Rehabilitation **14**(4): 295-305.

Reneman, M., Dijkstra, P., Westmaas, M. & Goeken, L. (2002a). "Test-Retest Reliability of Lifting and Carrying in a 2-Day Functional Capacity Evaluation." Journal of Occupational Rehabilitation **12**(4): 269-75.

Reneman, M., Jaegar, S., Westmaas, M.& Goeken, L. (2002). "The Reliability of Determining Effort Level of Lifting and Carrying in a Functional Capacity Evaluation." Work **18**(1): 23-27.

Reneman, M., Joling, C., Soer, E. & Goeken, L. (2001). "Functional Capacity Evaluation: Ecological Validity of Three Static Endurance Tests." Work **16**: 227-234.

Reneman, M., Jorritsma, W., Schellekens, J. & Goeken, L. (2002c). "Concurrent Validity of Questionnaire and Performance Based Disability Measurements in Patients with Chronic Non Specific Low Back Pain." Journal of Occupational Rehabilitation **12**(3): 119-129.

Reneman, M., Kool, J., Oesch, P., Geertzen, J., Battie, M.& Gross, D. (2006). "Material Handling Performance of Patients with Chronic Low Back Pain During Functional Capacity Evaluation: A Comparison between Three Countries." Disability and Rehabilitation **28**(18): 1143-1149.

Reneman, M., Kuijer, W., Brouwer, S., Preuper, H., Groothoff, J., Geertzen, J. & Dijkstra, P. (2006b). "Symptom Increase Following a Functional Capacity Evaluation in Patients with Chronic Low Back Pain: An Explorative Study of Safety." Journal of Occupational Rehabilitation **16**(2): 197-205.

Reneman, M., Soer, R.& Gerrits, E. (2005). "Basis for an FCE Methodology for Patients with Work-Related Upper Limb Disorders." Journal of Occupational Rehabilitation **15**(3): 353-63.

Robertson, D., Wilson, J.& St Pierre, T. (2008). "Lower Extremity Muscle Functions During Full Squats." Journal of Applied Biomechanics **24**(4): 333-9.

Ruan, C., Haig, A., Geisser, M., Yamakawa, K. & Buchholz, R. (2001). "Functional Capacity Evaluations in Persons with Spinal Disorders: Predicting Poor Outcomes on the Functional Assessment Screening Test (Fast)." Journal of Occupational Rehabilitation **11**(2): 119-32.

Rustenburg, G., Kuijer, P. & Frings-Dresen, M. (2004). "The Concurrent Validity of the Ergos Work Simulator and the Ergo-Kit with Respect to Maximum Lifting Capacity." Journal of Occupational Rehabilitation **14**(2): 107-18.

Sandqvist, J., Bjork, M., Gullberg, M., Henriksson, C. & Gerdle, B. (2009). "Construct Validity of the Assessment of Work Performance (AWP)." Work **32**(2): 211-218.

Sandqvist, J., Gullberg, M., Henriksson, C.& Gerdle, B. (2008). "Content Validity and Utility of the Assessment of Work Performance (AWP)." Work **30**(4): 441-450.

Schneider, Z., Whitehead, D. & Elliot, D., Eds. (2007). Nursing & Midwifery Research: Methods and Appraisal for Evidence Based Practice. Sydney, Elsevier.

Schonstein, E.& Kenny, D. (2001). "The Value of Functional and Work Place Assessments in Achieving a Timely Return to Work for Workers with Back Pain." Work **16**(1): 31-38.

Serxner, S., Anderson, D. & Gold, D. (2004). "Building Program Participation: Strategies for Recruitment and Retention in Worksite Health Promotion Programs." American Journal of Health Promotion **18**(4): 1-6.

Simmonds, M. (2002). "Physical Function in Patients with Cancer: Psychometric Characteristics and Clinical Usefulness of a Physical Performance Test Battery." Journal of Pain and Symptom Management **24**(4): 404-414.

Sireci, S. (1998). "Gathering and Analysing Content Validity Data." Educational Assessment **5**(4): 299-321.

Smeets, R., Hijdra, H., Kester, A., Hitters, M. & Knottnerus, J.-A. (2006). "The Usability of Six Physical Performance Tasks in a Rehabilitation Population with Chronic Low Back Pain." Clinical Rehabilitation **20**(11): 989.

Smith, R. (1994). "Therapist's Ability to Identify Safe Maximal Lifting in Low Back Pain Patients During Functional Capacity Evaluation." Journal of Orthopaedic and Sports Physical Therapy **19**(5): 227-281.

Snook, S.& Ciriello, V. (1991). "The Design of Manual Handling Tasks: Revised Tables of Maximum Acceptable Weights and Forces." Ergonomics **34**(9): 1197-213.

Soanes, C.& Stevenson, A. E., Eds. (2005). Oxford Dictionary of English. Oxford, Oxford University Press.

Soderberg, G.& Knutson, L. (2000). "A Guide for Use and Interpretation of Kinesiologic Electromyographic Data." Physical Therapy **80**(5): 485-498.

Soer, R., Cees, P., Groothoff, J., Geertzen, J.& Reneman, M. (2008). "Towards Consensus in Operational Definitions in Functional Capacity Evaluation: A Delphi Survey." Journal of Occupational Rehabilitation **18**(4): 389-400.

Soer, R., Gerrits, E.& Reneman, M. (2006). "Test-Retest Reliability of a WRULD Functional Capacity Evaluation in Healthy Adults." Work **26**(3): 273-280.

Soer, R., Poels, B., Geertzen, J. & Reneman, M. (2006b). "A Comparison of Two Lifting Assessment Approaches in Patients with Chronic Low Back Pain." Journal of Occupational Rehabilitation **16**(4): 639-46.

Spanjer, J., Krol, B., Brouwer, S., Popping, R., Groothoff, J.& van der Klink, J. (2010). "Reliability and Validity of the Disability Assessment Structured Interview (DASI): A Tool for Assessing Functional Limitations in Claimants." Journal of Occupational Rehabilitation **20**(1): 33-40.

Stamm, T., Cieza, A., Machold, K., Smolen, J. & Stucki, G. (2005). "Exploration of the Link between Conceptual Occupational Therapy Models and the International Classification of Functioning, Disability and Health." Australian Occupational Therapy Journal **53**(1): 9-17.

Statacorp (2003). Statastatistical Software: Release 8.0. College Station, TX, Stata Corporation.

StataCorp (2009). Stata Statistical Software: Version 11.1.College Station, Tx..

Streibelt, M., Blume, C., Thren, K., Reneman, M. & Mueller-Fahrnow, W. (2009). "Value of Functional Capacity Evaluation Information in a Clinical Setting for Predicting Return to Work." Archives of Physical Medicine and Rehabilitation **90**(3): 429-434.

Strong, S., Baptiste, S., Clarke, J.& Costa, M. (2004a). "Use of Functional Capacity Evaluations in Workplaces and the Compensation System: A Report on Workers' and Users' Perceptions." Work **23**(1): 67-77.

Strong, S., Baptiste, S., Cole, D., Clarke, J., Costa, M., Shannon, H., Reardon, R. & Sinclair, S. (2004b). "Functional Assessment of Injured Workers: A Profile of Assessor Practices." The Canadian Journal of Occupational Therapy **71**(1): (13-23).

Suto, M. (2000). Issues Related to Data Collection. Using Qualitative Research: A Practical Introduction for Occupational and Physical Therapists. K. Hammell, Carpenter, C.&Dyck, I. Edinburgh, :Churchill Livingstone**:** 25-46.

Thorn, D.& Deitz, J. (1989). "Examining Content Validity through the Use of Content Experts." Occupational Therapy Journal of Research **69**(6): 334

Ting, W., Wessel, J., Brintnell, S., Maikala, R.& Bhambhani, Y. (2001). "Validity of the Baltimore Therapeutic Equipment Work Simulator in the Measurement of Lifting Endurance in Healthy Men." American Journal of Occupational Therapy **55**(2): 184-90.

Toeppen-Sprigg, B. (2000). "Importance of Job Analysis with Functional Capacity Matching in Medical Case Management: A Physician's Perspective." Work **15**(2): 133-37.

Toomey, M., Nicolson, D.& Carswell, A. (1995). "The Clinical Utility of the Canadian Occupational Performance Measure." Canadian Journal of

Tramposh, A. (1992). "The Functional Capacity Evaluation: Measuring Maximal Work Abilities." Occupational Medicine: State of the Art Reviews **7**(1): 113-24.

Tuckwell, N., Straker, L.& Barrett, T. (2002). "Test-Retest Reliability on Nine Tasks of the Physical Work Performance Evaluation." Work **19**(3): 243-253.

U.S.Department of Labor (1965). Dictionary of Occupational Titles. Washington, DC., U.S. Government printing office.

Velozo, C. (1993). "Work Evaluations; Critique of the State of the Art of Functional Assessment of Work." American Journal of Occupational Therapy **47**: 203-209.

Vert, M. (2002). "Functional Capacity Evaluation." Orthopedics **25**(10): 1094.

Visanji, E. & Oldham, J. (2001). "Patient Recruitment in Clinical Trials: A Review of Literature." Physical Therapy Reviews **6**: 141-150.

Weierbach, F., Glick, D., Fletcher, K., Rowlands, A. & Lyder, C. "Nursing Research and Participant Recruitment: Organizational Challenges and Strategies." Journal of Nursing Administration **40**(1): 43-48.

Wind, H., Gouttebarge, V., Kuijer, P. & Frings-Dresen, M. (2005). "Assessment of Functional Capacity of the Musculoskeletal System in the Context of Work, Daily Living, and Sport: A Systematic Review." Journal of Occupational Rehabilitation **15**(2): 253-272.

Wind, H., Gouttebarge, V., Kuijer, P., Sluiter, J.& Frings-Dresen, M. (2006). "The Utility of Functional Capacity Evaluation: The Opinion of Physicians and Other Experts in the Field of Return to Work and Disability Claims." International Archives of Occupational and Environmental Health **79**(6): 528-534.

Wind, H., Gouttegbarge, V., Kuijer, P., Sluiter, J.& Frings-Dresen, M. (2008). "Complementary Value of Functional Capacity Evaluation for Physicians in Assessing the Physical Work Ability of Workers with Musculoskeletal Disorders." International Archives of Occupational and Environmental Health **82** (4): 435-443.

WorkCover (NSW). (2002). Rehabilitation Providers. Standards and Conditions of Accreditation. WorkCover. Sydney, WorkCover NSW.

WorkCover(Victoria) (2006). Occupational Rehabilitation Service Standards, WorkCover Victoria.

WorkHab International. (2008). "About Workhab FCE Systems." Retrieved 0202, 2009, from http://www.workhab.com/.

World Health Organisation (2001). The International Classification of Functioning, Disability and Health (ICF). Geneva: World Health Organisation.

Appendices

Appendix 1 Co-author statements

Manuscript 1: (3.1) Health Professionals' Attitudes and Practices in relation to Functional Capacity Evaluations

Citation: James, C., Mackenzie, L. and Higginbotham, N. (2007). "Health Professionals' Attitudes and Practices in relation to Functional Capacity Evaluations." Work **29**(2): 81-88.

STATEMENT FROM CO-AUTHORS RELATING TO A PAPER PUBLISHED WITH CAROLE JAMES.

The Rules Governing Research Higher Degrees (Rule 000830) allow for a thesis to be submitted in the form of a series of published papers. Where this approach is used a signed statement is required from co-authors.

Statement from co-author:

I, Dr Lynette MacKenzie, attest that Research Higher Degree candidate CAROLE JAMES contributed to the paper/publication entitled:

1. "Health Professionals' Attitudes and Practices in relation to Functional Capacity Evaluations" published in the Journal: WORK: A Journal of Prevention, Assessment & Rehabilitation. 2007, 29(2), 81-88.

by conducting and writing- up the literature review, collecting the data, undertaking data analysis, description and interpretation of results and writing methodology, results, discussion and conclusions.

Dr LYNETTE MACKENZIE (Co-Author)

Date:

CAROLE JAMES (Candidate)

Date:

(PROFESSOR JOHN ROSTAS (Assistant Dean Research Training (ADRT))
Date:

STATEMENT FROM CO-AUTHORS RELATING TO A PAPER PUBLISHED WITH CAROLE JAMES.

The Rules Governing Research Higher Degrees (Rule 000830) allow for a thesis to be submitted in the form of a series of published papers. Where this approach is used a signed statement is required from co-authors.

Statement from co-author:

I, Nick Higginbotham, attest that Research Higher Degree candidate CAROLE JAMES contributed to the paper/publication entitled:

1. "Health Professionals' Attitudes and Practices in relation to Functional Capacity Evaluations" published in the Journal: WORK: A Journal of Prevention, Assessment & Rehabilitation. 2007, 29(2), 81-88.

by conducting and writing- up the literature review, collecting the data, undertaking data analysis, description and interpretation of results and writing methodology, results, discussion and conclusions.

(Signature of Co-Author)

Date:

CAROLE JAMES

Date:

Manuscript 2: (3.2) Health Professionals' Perceptions and Practices in relation to Functional Capacity Evaluations – Results of a Quantitative Survey

Citation: James, C. and Mackenzie, L. (2009). "Health Professional's Perceptions and Practices in Relation to Functional Capacity Evaluations: Results of a Quantitative Survey." Journal of Occupational Rehabilitation **19**(2): 203-211

STATEMENT FROM CO-AUTHORS RELATING TO A PAPER PUBLISHED WITH CAROLE JAMES.

The Rules Governing Research Higher Degrees (Rule 000830) allow for a thesis to be submitted in the form of a series of published papers. Where this approach is used a signed statement is required from co-authors.

Statement from co-author:

I, Dr Lynette Mackenzie, attest that Research Higher Degree candidate CAROLE JAMES contributed to the paper/publication entitled:

1. "Health Professional's Perceptions and Practices in Relation to Functional Capacity Evaluations: Results of a Quantitative Survey" published in the Journal of Occupational Rehabilitation. 2009, 19(2), 203-211

by conducting and writing- up the literature review, developing the survey instrument, collecting the data, undertaking data and statistical analysis, description and interpretation of results and writing methodology, results, discussion and conclusions.

Dr LYNETTE MACKENZIE (Co-Author)

Date:

CAROLE JAMES (Candidate)

Date:

(PROFESSOR JOHN ROSTAS (Assistant Dean Research Training (ADRT))

Date:

Manuscript 3: (3.3) The Clinical Utility of Functional Capacity Evaluations: the Opinion of Health Professionals working within Occupational Rehabilitation

Citation: James, C. and Mackenzie, L. (2009). "The Clinical Utility of Functional Capacity Evaluations: The Opinion of Health Professionals Working within Occupational Rehabilitation." Work **33**(3): 231-9.

STATEMENT FROM CO-AUTHORS RELATING TO A PAPER PUBLISHED WITH CAROLE JAMES.

The Rules Governing Research Higher Degrees (Rule 000830) allow for a thesis to be submitted in the form of a series of published papers. Where this approach is used a signed statement is required from co-authors.

Statement from co-author:

I, Dr Lynette MacKenzie, attest that Research Higher Degree candidate CAROLE JAMES contributed to the paper/publication entitled:

1. "The Clinical Utility of Functional Capacity Evaluations: the opinion of Health Professionals working within Occupational Rehabilitation" published in WORK: A Journal of Prevention, Assessment & Rehabilitation. 2009, 33(3),231-239

by conducting and writing- up the literature review, developing the survey instrument, collecting the data, undertaking data analysis, description and interpretation of results and writing methodology, results, discussion and conclusions.

Dr LYNETTE MACKENZIE (Co-Author)

Date:

CAROLE JAMES (Candidate)

Date:

PROFESSOR JOHN ROSTAS (Assistant Dean Research Training (ADRT))

Date:

Manuscript 4: (4.1) Test - Retest Reliability of the Manual Handling Component of the WorkHab Functional Capacity Evaluation in Healthy Adults

Citation: James, C., Mackenzie, L. and Capra, M. (2010). "Test-Retest Reliability of the Manual Handling Component of the WorkHab Functional Capacity Evaluation in Healthy Adults." Disability and Rehabilitation **32**(22): 1863-1869.

STATEMENT FROM CO-AUTHORS RELATING TO A PAPER PUBLISHED WITH CAROLE JAMES.

The Rules Governing Research Higher Degrees (Rule 000830) allow for a thesis to be submitted in the form of a series of published papers. Where this approach is used a signed statement is required from co-authors.

Statement from co-author:

I, Dr Lynette Mackenzie, attest that Research Higher Degree candidate CAROLE JAMES contributed to the paper/publication entitled:

1. "Test - Retest reliability of the Manual Handling component of the WorkHab Functional Capacity Evaluation" published in the journal 'Disability and Rehabilitation' 2010, 32(22):1863-1869.

by conducting and writing- up the literature review, collecting the data, undertaking data and statistical analysis, description and interpretation of results and writing methodology, results, discussion and conclusions.

Dr LYNETTE MACKENZIE (Co-Author)

Date:

CAROLE JAMES (Candidate)

Date:

PROFESSOR JOHN ROSTAS (Assistant Dean Research Training (ADRT))

Date:

STATEMENT FROM CO-AUTHORS RELATING TO A PAPER PUBLISHED WITH CAROLE JAMES.

The Rules Governing Research Higher Degrees (Rule 000830) allow for a thesis to be submitted in the form of a series of published papers. Where this approach is used a signed statement is required from co-authors.

Statement from co-author:

I, Professor Mike Capra, attest that Research Higher Degree candidate CAROLE JAMES contributed to the paper/publication entitled:

1. "Test - Retest reliability of the Manual Handling component of the WorkHab Functional Capacity Evaluation" published in the journal 'Disability and Rehabilitation' 2010, 32(22):1863-1869.

by conducting and writing- up the literature review, collecting the data, undertaking data and statistical analysis, description and interpretation of results and writing methodology, results, discussion and conclusions.

PROFESSOR MIKE CAPRA (Co-Author)

Date:

CAROLE JAMES (Candidate)

Date:

PROFESSOR JOHN ROSTAS (Assistant Dean Research Training (ADRT))

Date:

Manuscript 5: (4.2) Inter and Intra-rater Reliability of the Manual Handling Component of the WorkHab Functional Capacity Evaluation

Citation: James, C., Mackenzie, L. and Capra, M. (early online). “Inter and Intra- rater reliability of the Manual Handling component of the WorkHab Functional Capacity Evaluation.” Disability and Rehabilitation.

DOI:10.3109/09638288.2010.548896 (posted online: January 19, 2011).

STATEMENT FROM CO-AUTHORS RELATING TO A PAPER PUBLISHED WITH CAROLE JAMES.

The Rules Governing Research Higher Degrees (Rule 000830) allow for a thesis to be submitted in the form of a series of published papers. Where this approach is used a signed statement is required from co-authors.

Statement from co-author:

I, Dr Lynette Mackenzie, attest that Research Higher Degree candidate CAROLE JAMES contributed to the paper/publication entitled:

1. "Inter and Intra rater reliability of the Manual Handling component of the WorkHab Functional Capacity Evaluation" accepted for publication in Disability and Rehabilitation.

by conducting and writing- up the literature review, collecting the data, undertaking data and statistical analysis, description and interpretation of results and writing methodology, results, discussion and conclusions.

Dr LYNETTE MACKENZIE (Co-Author)

Date:

CAROLE JAMES (Candidate)

Date:

PROFESSOR JOHN ROSTAS (Assistant Dean Research Training (ADRT))

Date:

STATEMENT FROM CO-AUTHORS RELATING TO A PAPER PUBLISHED WITH CAROLE JAMES.

The Rules Governing Research Higher Degrees (Rule 000830) allow for a thesis to be submitted in the form of a series of published papers. Where this approach is used a signed statement is required from co-authors.

Statement from co-author:

I, Professor Mike Capra, attest that Research Higher Degree candidate CAROLE JAMES contributed to the paper/publication entitled:

1. "Inter and Intra rater reliability of the Manual Handling component of the WorkHab Functional Capacity Evaluation" accepted for publication in Disability and Rehabilitation.

by conducting and writing- up the literature review, collecting the data, undertaking data and statistical analysis, description and interpretation of results and writing methodology, results, discussion and conclusions.

PROFESSOR MIKE CAPRA (Co-Author)

Date:

CAROLE JAMES (Candidate)

Date:

PROFESSOR JOHN ROSTAS (Assistant Dean Research Training (ADRT))

Date:

Manuscript 6: (5.1) Content Validity of the WorkHab Functional Capacity Evaluation

Citation: James, C., Mackenzie, L. and Capra, M. (under review). "Content Validity of the WorkHab Functional Capacity Evaluation." Disability and Rehabilitation.

STATEMENT FROM CO-AUTHORS RELATING TO A PAPER PUBLISHED WITH CAROLE JAMES.

The Rules Governing Research Higher Degrees (Rule 000830) allow for a thesis to be submitted in the form of a series of published papers. Where this approach is used a signed statement is required from co-authors.

Statement from co-author:

I, Dr Lynette MacKenzie, attest that Research Higher Degree candidate CAROLE JAMES contributed to the paper/publication entitled:

1. "Content validity of the WorkHab Functional Capacity Evaluation" submitted for publication in Disability and Rehabilitation.

by conducting and writing- up the literature review, collecting the data, undertaking data and statistical analysis, description and interpretation of results and writing methodology, results, discussion and conclusions.

Dr LYNETTE MACKENZIE (Co-Author)

Date:

CAROLE JAMES (Candidate)

Date:

PROFESSOR JOHN ROSTAS (Assistant Dean Research Training (ADRT))

Date:

STATEMENT FROM CO-AUTHORS RELATING TO A PAPER PUBLISHED WITH CAROLE JAMES.

The Rules Governing Research Higher Degrees (Rule 000830) allow for a thesis to be submitted in the form of a series of published papers. Where this approach is used a signed statement is required from co-authors.

Statement from co-author:

I, Professor Mike Capra, attest that Research Higher Degree candidate CAROLE JAMES contributed to the paper/publication entitled:

1. "Content validity of the WorkHab Functional Capacity Evaluation" submitted for publication in Disability and Rehabilitation.

by conducting and writing- up the literature review, collecting the data, undertaking data and statistical analysis, description and interpretation of results and writing methodology, results, discussion and conclusions.

PROFESSOR MIKE CAPRA (Co-Author)

Date:

CAROLE JAMES (Candidate)

Date:

PROFESSOR JOHN ROSTAS (Assistant Dean Research Training (ADRT))

Date:

Manuscript 7: (5.2) Physiological correlates of Functional Capacity Evaluations: Finding the Safe Maximal Lift.

Citation: James, C., Mackenzie, L. and Capra, M. (under review). "Physiological correlates of Functional Capacity Evaluations: Finding the Safe Maximal Lift." submitted for publication in Archives of Physical Medicine and Rehabilitation.

STATEMENT FROM CO-AUTHORS RELATING TO A PAPER PUBLISHED WITH CAROLE JAMES.

The Rules Governing Research Higher Degrees (Rule 000830) allow for a thesis to be submitted in the form of a series of published papers. Where this approach is used a signed statement is required from co-authors.

Statement from co-author:

I, Dr Lynette MacKenzie, attest that Research Higher Degree candidate CAROLE JAMES contributed to the paper/publication entitled:

1. "Physiological correlates of functional capacity evaluations: finding the safe maximal lift" submitted for publication in Archives of Physical Medicine and Rehabilitation.

by conducting and writing- up the literature review, collecting the data, undertaking data and statistical analysis, description and interpretation of results and writing methodology, results, discussion and conclusions.

DR LYNETTE MACKENZIE (Co-Author)

Date:

CAROLE JAMES (Candidate)

Date:

PROFESSOR JOHN ROSTAS (Assistant Dean Research Training (ADRT))

Date:

STATEMENT FROM CO-AUTHORS RELATING TO A PAPER PUBLISHED WITH CAROLE JAMES.

The Rules Governing Research Higher Degrees (Rule 000830) allow for a thesis to be submitted in the form of a series of published papers. Where this approach is used a signed statement is required from co-authors.

Statement from co-author:

I, Professor Mike Capra, attest that Research Higher Degree candidate CAROLE JAMES contributed to the paper/publication entitled:

1. "Physiological correlates of functional capacity evaluations: finding the safe maximal lift" submitted for publication in Archives of Physical Medicine and Rehabilitation.

by conducting and writing- up the literature review, collecting the data, undertaking data and statistical analysis, description and interpretation of results and writing methodology, results, discussion and conclusions.

PROFESSOR MIKE CAPRA (Co-Author)

Date:

CAROLE JAMES (Candidate)

Date:

PROFESSOR JOHN ROSTAS (Assistant Dean Research Training (ADRT))

Date:

Appendix 2: Poster presentations

Poster 1: James, C., Mackenzie, L. and Capra, M. (2006). Health Professionals' Attitudes and Practices in Relation to Functional Capacity. 4th Congress of the World Federation of Occupational Therapists. Sydney, NSW, Australia.

THE UNIVERSITY OF NEWCASTLE
AUSTRALIA

Health professional's attitudes and practices in relation to functional capacity evaluations

Carole James, Dr Lynette Mackenzie & Professor Mike Capra

School of Health Sciences, Faculty of Health, The University of Newcastle

Context:

Functional Capacity Evaluations (FCE's) are part of practice in Work injury prevention and rehabilitation and are designed to define an individual's functional abilities or limitations in the context of safe, productive work tasks (King, Tuckwell & Barrett, 1998). They are commonly used with individuals who have suffered work related musculo-skeletal injuries. There are many different FCE's available commercially and many clinics have developed their own non standardised, work specific FCE's. All FCE's attempt to measure functional performance objectively.

However, few published studies appraise the reliability and validity of the assessments, to establish if this objectivity is achieved (Innes and Straker 1999a, & 1999b. King , Tuckwell & Barrett, 1998). Innes and Straker (2002) studied the current practice of therapists in Australia in relation to work assessment and found that assessments were grouped into; Work assessments, FCE (no job) and FCE (job). A variety of factors affecting the type, purpose and characteristics of the assessment and their influences and constraints were described; however, details of specific FCE's being used were not included. Therapists reported using triangulation of data to confirm results from the assessments:- however, further investigation of factors influencing clinical judgments were not discussed.

Purpose

This research had 3 aims:

1). to investigate the attitudes and practices of health professionals in relation to FCE use,

2). to identify why health professionals chose a particular FCE, and

3). to identify what factors influence health professionals' clinical judgments, when providing results and recommendations for the individual being assessed.

Methods

A quantitative cross sectional study design was used to survey health professionals, who conduct FCE's and who were working for rehabilitation providers in NSW. Following ethics approval, a small qualitative study to investigate the attitudes and practices of 5 health professionals in relation to FCE usage was conducted. Using the information gathered a pilot survey was then developed and sent to health professionals within the Hunter region who had previously agreed to participate. After modification, the survey was sent to 219 WorkCover (NSW) accredited rehabilitation providers, with the Managers asked to forward information to health professionals who conduct Functional Capacity Evaluations.

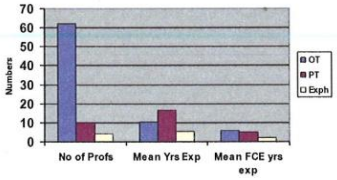
Data analysis

Surveys were coded and data were entered into STATA for descriptive analysis.

Participants

79 surveys were returned from 65 different providers, this being a response rate of 29.68%.
Of those who replied, 82% were Occupational Therapists, 13% Physiotherapists and 5% Exercise Physiologists
Table 1 provides some description of the sample.

Table 1: Descriptive information on Sample (n=77)

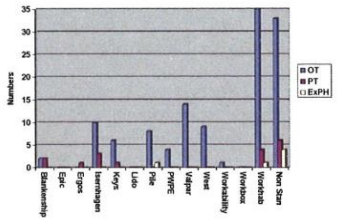


Results

The results showed 11 different FCE's were being used throughout NSW. Table 2 outlines FCE usage by profession.

What FCE's were being used in NSW.

Table 2: Usage by Profession & FCE type (n=77)



Of these 11 FCE's, the proportions of therapists using part or all of the components of the seven most used assessments is outlined in Table 3 & 4.

Utilisation of whole or parts of FCE.

Table 3: Utilisation of ALL components of FCE

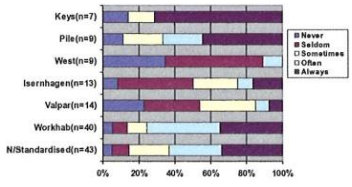
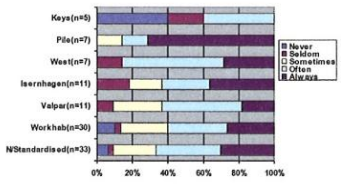


Table 4: Utilisation of PARTS of FCE



Utilisation issues.

40% of therapists indicated they were unlikely to use identical FCE's with all clients, whereas 36% indicated they would.

Therapists adapted the FCE conducted to suit the clients' injury (82%), and 80% indicated they would adapt the FCE in response to the clients' job.

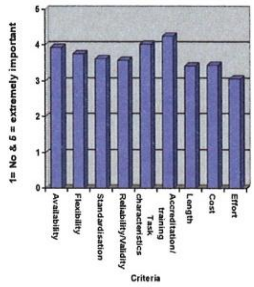
FCE's / Workplace assessments

33% of therapists indicated they conducted an FCE **AND** a workplace assessment often. However, 47% of therapists indicated they often conducted Workplace assessments **INSTEAD** of FCE's.

Factors affecting selection of FCE.

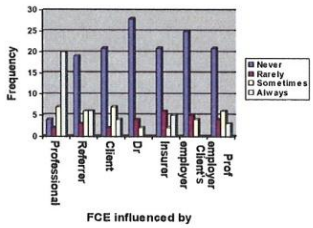
The importance of different criteria/issues in the selection of an FCE is outlined in Table 5.

Table 5: Mean score of importance of different qualities/criteria in choice of FCE



40% of therapists had a choice in the type of FCE that was conducted. 60% had no choice. Table 6 shows the frequency of different influences on the choice of FCE used for those therapists that did have a choice.

Table 6: Influences for FCE's (n=31)



Discussion

This study found 11 different FCE's were being used within NSW. Many therapists were using parts of, rather than the whole assessment and adapted the assessment in response to the client injury and job.

Standardisation, reliability and validity were identified as important criteria when considering FCE use. However, task characteristics and flexibility within the FCE were also identified as important criteria which relate to the adaptation of the FCE to suit the individual, and may impact upon reliability and validity of the assessment.

Further research is needed to establish how FCE's can be used in a valid and reliable way in everyday practice.

The University of Newcastle
AUSTRALIA

References

Innes, E., & Straker,L. (1999a). Reliability of work-related assessments. *Work*. 13, 107-124.
Innes, E., & Straker,L. (1999b). Validity of work-related assessments. *Work*. 13, 125-152.
Innes, E., & Straker,L. (2003). Workplace assessments and functional capacity evaluations: Current beliefs of therapists in Australia. *Work*. 20, 225-236.
Innes, E., & Straker,L. (2002). Workplace assessments and functional capacity evaluations: Current practices of therapists in Australia. *Work*. 18, 51-66.
King, P.M., Tuckwell, N., & Barrett, T. (1998). A critical review of functional capacity evaluations. Physical Therapy. *Physical Therapy*. 78, 852-866.

Poster 2: James, C. and Mackenzie, L. (2008). The Clinical Utility of Functional Capacity Evaluations: The Opinion of Health Professionals Working within Occupational Rehabilitation. OT Australia 23rd National Conference. Melbourne, Australia.

The Clinical Utility of Functional Capacity Evaluations: the opinion of Health Professionals working in Occupational Rehabilitation.

Carole James, Dr Lynette Mackenzie & Professor Mike Capra
School of Health Sciences, Faculty of Health, The University of Newcastle

Context:

Functional Capacity Evaluations (FCE's) are part of practice in Work injury prevention and rehabilitation and are designed to define an individual's functional as or limitations in the context of safe, productive bilitiework tasks (King, Tuckwell & Barrett, 1998). They are commonly used with individuals who have suffered work related musculo-skeletal injuries. There are many different FCE's available commercially and many clinics have developed their own non standardised, work specific FCE's. All FCE's attempt to measure functional performance objectively.

Innes and Straker (2003) suggest therapists using FCE's undertake a decision making process, considering safety first , then the constructs of dependability (quantitative attributes and the concept of reliability) and utility (qualitative attributes and the concept of validity) demonstrated by the
FCE used.

Clinical utility reflects the degree of conviction therapists have about the usefulness of an assessment (Toomey, Nicolson, & Carswell,1995). Clinical utility confirms that the tool is related to the purpose for which it is used (Barbara & Whiteford, 2005), and includes features such as accuracy, comprehensiveness, credibility, flexibility, practicality, relevance, suitability, value and adaptability. The clinical utility of an assessment tool provides important information about the usefulness of the tool and the ease with which a therapist can conduct the assessment.

Purpose

This study explored the clinical utility of FCE's using a questionnaire distributed to health professionals that investigated their attitudes, practices and perceptions about their use of FCE's. The study addressed the research question:
What are therapists' views on the clinical utility of FCE's and do these perceptions differ between professional groups?

Methods

A quantitative cross sectional study design was used to survey health professionals, who conduct FCE's and who were working for rehabilitation providers in NSW.
Following ethics approval, a pilot survey was developed based on findings from an earlier qualitative study (James, Mackenzie & Higginbotham(2007). This was sent to health professionals within the Hunter region who had previously agreed to participate. After modification, the survey which consisted of 60 items within 6 sections: demographic data, type of FCE used, FCE choice, FCE usage, perceived consequences of using FCE's and perceptions of FCE's, was sent to 219 WorkCover (NSW) accredited rehabilitation providers, with the Managers asked to forward information to health professionals who conduct Functional Capacity Evaluations.

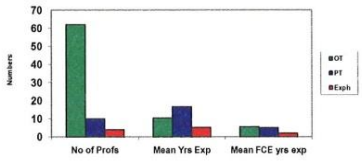
Data analysis

Surveys were coded and data were entered into STATA for descriptive analysis. Cronbach's alpha was used to measure the internal consistency of the scales and to evaluate the homogeneity of the items for each subscale identified in the survey. A Cronbach's alpha of between .70 and .90 was considered to indicate sufficient internal consistency and indicates the items within the scale are measuring the same construct (Depoy & Gitlin,1998). Analysis of variance (ANOVA) was used to determine any differences between the professional groups on continuous scores from scale items

Participants

79 surveys were returned from 65 different providers, this being a response rate of 29.68%. Of those who replied, 82% were Occupational Therapists, 13% Physiotherapists and 5% Exercise Physiologists
Table 1 provides some description of the sample.

Table 1: Descriptive information on Sample (n=77)



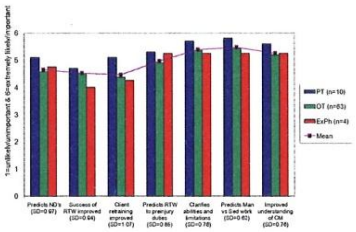
Results

The results are presented for the four subscales included in the survey instrument: usefulness and relevance; adaptability and flexibility; therapist perceived requirements and issues in practice.

Usefulness and Relevance.

Therapists indicated they obtained relevant and useful clinical information from the FCE used in their workplace. Table 2 outlines what therapists used FCE results for.

Table 2 - Usefulness and Relevance of FCE's (N=77).



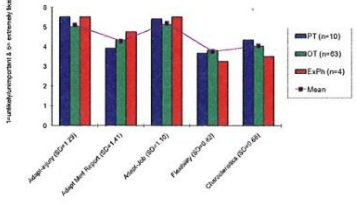
Key for figure 1: ND= Normal duties, RTW = Return to Work, Man = Manual, Sed = Sedentary, CM = Case Manager

Cronbach's alpha for this scale was 0.82.
Mean value = 5.04 (95%CI: 4.86-5.23).No significant difference was detected between groups.

Adaptability and Flexibility.

This subscale included items that investigated the clinical utility issues of adapting FCE's and flexibility of FCE's in practice.
Adequate internal consistency was identified – Cronbach's alpha = 0.7.
Overall mean = 4.47 (95%CI= 4.22-4.71).
No significant difference between groups detected.
Table 3:

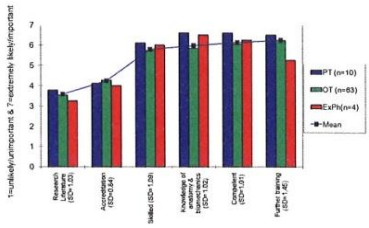
Table 3: Adaptability and Flexibility of FCE's (N=77)



Therapist perceived requirements.

This group of items included requirements for therapists to be able to use FCE's effectively in practice. Table 4 outlines the results:

Table 4:Therapist perceived requirements to administer FCE's (N=77)

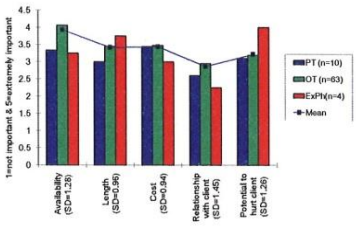


Cronbach's alpha indicated an adequate internal consistency of 0.74 for this subscale.
The mean value was 5.32 (95%CI: 5.07-5.57). No significant difference was detected between groups.

Issues in Practice.

The issues relating to using FCE's in practice were presented individually in the survey not in a scale. The results are outlined in Table 5:

Table 5: Issues in practice administering FCE's (N=77)



Discussion

This study did not identify any differences in opinion related to the clinical utility of FCE's between professional groups or experience despite this being an expectation as a result of the diversity of undergraduate theoretical approaches to professional roles. The results suggest a level of consistency in how FCE's are currently being used in practice across NSW (Australia) and with attitudes towards their clinical utility.

Therapists reported the FCE assisted in predicting RTW and manual vs sedentary duties.

All professional groups indicated a tendency to adapt FCE's based on the client, job and injury rather than adhering consciously to professionally defined models of practice or using standardised measures. Flexibility was also identified as important for all professional groups.

These factors impact upon the reliability and validity of an assessment tool, however from this study it can be suggested that therapists are adapting FCE's in practice despite the risks to the reliability and validity of the findings.

References

Barbara, A., & Whiteford, G. (2005). Clinical utility of the Handicap Assessment and Resource Tool: an investigation of its use with the aged people in hospital, *Australian Occupational Therapy Journal* 52[1], 17-25.
Depoy, E., & Gitlin, L. (1998) Introduction to Research. Understanding and applying Multiple Strategies (2nd ed.).St Louis: Mosby.
Innes, I., & Straker,L. (2003) Attributes of excellence in work related assessments, *Work* 20. 63-76.
King, P.M., Tuckwell, N., & Barrett, T. (1998). A critical review of functional capacity evaluations. Physical Therapy. *Physical Therapy*. 78, 852-866.
Portney, L., & Watkins, M. (2000).Foundations of Clinical Research: Applications to practice (2nd ed.). New Jersey: Prentice Hall Health
Toomey, M., Nicolson, M., & Carswell, A. (1995). The clinical utility of the Canadian Occupational Performance Measure, *Canadian Journal of Occupational Therapy* 62[5],242-249.Barbara,

Poster 3: James, C., Mackenzie, L. and Capra, M. (2010). Evidence for Functional Capacity Evaluations – the Test-Retest Reliability of the Workhab FCE. 15th Congress of the World Federation of Occupational Therapists. Santiago, Chile.

Evidence for Functional Capacity Evaluations: The test-retest reliability of the WorkHab FCE

Carole James[1], Dr Lynette Mackenzie[2], Professor Mike Capra[3]
[1]School of Health Sciences, Faculty of Health, University of Newcastle
[2] Faculty of Health Sciences, University of Sydney
[3] Faculty of Health Sciences, University of Queensland

Context

Functional capacity evaluations (FCEs) are an integral part of work injury prevention and occupational rehabilitation. FCEs are designed to provide a comprehensive, performance based assessment, to define the functional abilities and limitations of an individual in the context of safe, productive work tasks [1-3].
The WorkHab Functional Capacity Evaluation (FCE) is one of many FCEs currently available and is widely used in the Australian workplace injury management and occupational rehabilitation arena [4-5], however there is limited published literature on the psychometric properties of this FCE [6]. The WorkHab FCE is constructed of a series of tests looking at mobility, strength, cardiovascular fitness, tolerance to postures and movements, and manual handling abilities and uses physiological and kinematic performance measures (heart rate, pain, perceived exertion and the observation of biomechanics) [7].

Purpose

To evaluate the test retest reliability and agreement between two testing occasions for the manual handling component (floor to bench, bench to shoulder and bench to bench lifts) of the WorkHab FCE in healthy adults using one assessor.

Participants

A convenience sample of 25 healthy adult volunteers from a University staff and student population were recruited. The study sample consisted of 19 women and 6 men ranging in age from 19 years to 54 years, with a mean age of 29 years (SD :12.0). The mean weight was 66.4kg (SD: 10.4) and the mean height was 167.9cm (SD: 8.4).

Methodology

Following ethics approval the study was advertised to University staff and students. Prior to commencing the FCE participants gave informed consent and underwent a pre-assessment screening. The bench to bench, bench to shoulder and floor to bench (relative to participants waist & shoulder) manual handling components of the WorkHab FCE were completed (fig 1). The participant was instructed to lift the load box (initially empty) from start to end position and return three times before additional weight was added to the load box. The safe maximal lifting limits were determined for each lift.
Two testing sessions were held a week apart with the time of day being kept constant where possible. The FCE was conducted by one Occupational Therapist who was trained and accredited as a WorkHab assessor, with ten years experience in conducting FCEs. The participants were asked to perform to their maximum abilities. Information collected by the assessor included the weight lifted in kg.

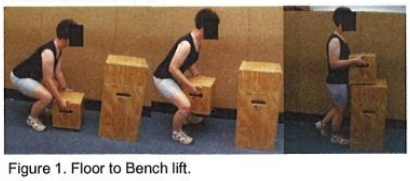


Figure 1. Floor to Bench lift.

References

[1]King PM, Tuckwell N, Barrett TE. A critical review of functional capacity evaluations. Physical Therapy. 1998;78(8):852
[2]Gouttebarge V, Wind H, Kuijer PP, Sluiter JK, Frings-Dresen MH. Reliability and Agreement of 5 Ergo-Kit Functional Capacity Evaluation Lifting Tests in Subjects With Low Back Pain. Archives of Physical Medicine and Rehabilitation. 2006;87(10):1365-70
[3]Lee G, Chan C, Hui-Chan C. Work profile and functional capacity of formwork carpenters at construction sites. Disability and Rehabilitation. 2001;23(1):9-14
[4]Cotton A, Schonstein E, Adams R. Use of Functional Capacity Evaluations by rehabilitation providers in NSW. Work. 2006;26:287-95.
[5]James C, Mackenzie L. Health Professional's Perceptions and Practices in Relation to Functional Capacity Evaluations. Results of a Quantitative Survey. Journal of Occupational Rehabilitation. 2009;19(2):203-11.
[6]Innes E, Straker L. Reliability of work-related assessments. Work. 1999;13(2):107-24
[7]Bradbury S, Roberts D. WorkHab Functional Capacity Evaluation Procedural Manual WorkHab Australia 1998.
[8]Bland JM, Altman DG. Measuring agreement in method comparison studies. Statistical Methods in Medical Research. 1999 06;8(2):135-60
[9]Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. 3rd ed. Upper Saddle River, New Jersey: Pearson/Prentice Hall 2009
[10]Reneman M, Brouwer S, Meesema A, Dijkstra P, Geertzen J, Groothoff J. Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in healthy adults. Journal of Occupational Rehabilitation. 2004 Dec 2004 14(4):295-305.
[11]Legge J, Burgess-Limerick R. Reliability of the JobFit system Pre-employment Functional assessment tool. Work. 2007;28:299-312
[12]Tuckwell NL, Straker L, Barrett TE. Test-retest reliability on nine tasks of the Physical Work Performance Evaluation. Work. 2002 11;18(3):243-53.

Data Analysis

Descriptive analysis, one way random Intraclass Correlation Coefficients (ICC's), 95% confidence intervals, limits of agreement [8], paired sample t-test, kappa (weighted for ordinal data) and percentage agreement were calculated where appropriate using SPSS (version 16.0). Percentage agreement, as a measure of agreement, can be used to determine reliability and kappa is a chance corrected measure of agreement considering both the proportion of observed agreements and the proportion expected by chance [9].

Results

The means, standard deviations, limits of agreement, 95% confidence intervals and ICC's for the three lifts are presented in Table 1:

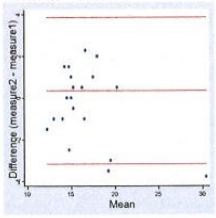
Table 1: Results of means, differences, standard deviations, 95% confidence intervals and ICC's for lifts.

Lift	Mean 1[a] (SD)	Mean 2[b] (SD)	Mean dif (SD)	95% CI[d] of difference	Limits of agreement	ICC[f]	95% CI of ICC	Interpretation of ICC
Floor to Bench	17.0 (3.9)	16.8 (3.6)	0.18 (2.1)	0.68 - 1.04	-4.0 - 5.2	0.92	0.82 – 0.96	Excellent
Bench to Shoulder	13.3 (3.3)	13.0 (2.6)	0.31 (1.7)	-0.41 - 1.03	-3.0 to 4.0	0.90	0.78 - 0.96	Excellent
Bench to Bench	16.8 (4.7)	16.8 (3.8)	-0.02 (2.5)	-1.07 -1.04	-3.7 to 2.2	0.91	0.79 - 0.96	Excellent

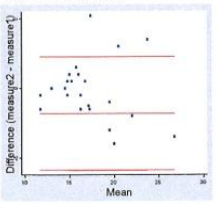
[a]: Mean 1 = group mean in first session (kg). [b] Mean 2 = group mean in second session (kg). [c] SD = Standard deviation. [d] 95% CI = 95% Confidence interval. [e] ratio between limits of agreement and mean score x100%
[f] ICC = Intraclass Correlation Coefficient (one way random). *Significant (two tailed) at p<0.05.

The limits of agreement of the three lifts ranged from ± 3 to 4kg variation in weight lifted. Figure 2 shows the distributions of the limits of agreement, where it can be seen that the majority of lifts were within ± 2.5kg difference, however lower levels of agreement are seen with higher mean loads.

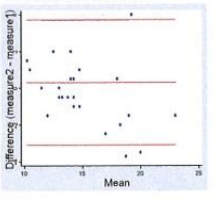
Figure 2:
Bench to Bench lift



Floor to Bench lift



Bench to shoulder lift



Discussion

The importance of reliable and valid assessment tools being used by therapists to identify abilities and limitations of individuals is highlighted in the literature [6]. Test-retest reliability is used to determine the consistency of a measure from one testing occasion to another [10]. Test-retest reliability for the lifting components of the WorkHab FCE was expressed by Intraclass Correlation Coefficients (ICC's) which is a measure of between- subject variance and within subject variance and is an accepted measure of reliability in relation to the discriminative capacity of a test. For the bench to bench, the floor to bench and the bench to shoulder lifts, the test-retest reliability was high (ICC's: 0.90-0.92). 95% limits of agreement were calculated as a descriptive measure of agreement with results being considered from a clinical rather than a statistical interpretation. Clinically a variation of ± 3 to 4kg in weight lifted maybe appropriate in some situations where heavier loads are lifted, however, when looking at the distribution of the limits of agreement (Figure 2), the majority of lifts were within ± 2.5kg difference, which is clinically more acceptable. The results of any FCE need to be interpreted and applied to the specific tasks, job and workplace for the individual, using clinical judgement skills. This variation can therefore be considered acceptable when considering functional ability for return to work and implies that this is an acceptable clinical measure of agreement for this purpose.
These findings suggest the administration procedures for the WorkHab FCE are dependable and the average performance by the subjects was relatively stable over the study period. Similar types of lifting tasks evaluated in other FCE's have reported substantial or acceptable test-retest reliability [2, 10,11,12].
The results of the lifting found substantial levels of test-retest reliability with this sample. It can therefore be concluded, that these lifts in the WorkHab FCE are reliable in healthy adults. Further research is recommended to establish other forms of reliability and validity of this assessment tool, using a range of client samples. The results from this study contribute to the growing evidence of FCE's in practice and the importance of reliability and validity in work related assessments.

Acknowledgements : This research was made possible with a grant from the University of Newcastle (GO187308) and thanks go to all the participants who willingly gave of their time and to WorkHab Australia who donated the equipment .

Appendix 3: Questionnaire - Health Professionals' Attitudes and Practices in relation to Functional Capacity Evaluations

Health Professionals' Attitudes and Practices in relation to Functional Capacity Evaluations

Researcher – **Carole James,**
Supervisors - **Dr Lynette MacKenzie**
Dr Nick Higginbotham
c/o Discipline of Occupational Therapy, Faculty of Health,
University of Newcastle,
Box 19, Hunter Building, University Drive, Callaghan, NSW 2308.
Tel: 02 49 216398
Fax: 02 49 216998

This project has been approved by the University's Human Research Ethics Committee, Approval No. H-576-0503

Should you have any concerns about your rights as a participant in this research or you have a complaint about the manner in which the research is conducted, it may be given to the researcher. If an independent person is preferred, to the Human Research Ethics Officer, Research Office, The Chancellery, The University of Newcastle, University Drive, Callaghan NSW 2308. Tel: (02) 49216333.

BACKGROUND INFORMATION:

1. Please tick your profession:

Physiotherapist □
Occupational Therapist □
Occupational Health Nurse □
Exercise Physiologist □
Other – please specify _____.

2. How many years clinical experience do you have? _____Years.

3. How many years experience conducting FCE's? _____Years.

Below is a list of Assessment Types (FCE's). For each type please indicate whether or not you use it. If YES, please continue along the row and circle the appropriate responses about how often you use all components, selected components and the number of times each month you use it. If no, please skip to the next assessment type.

Assessment type	**Use?**	**I use all components of this Ax:**	**I use only parts of this Ax:**	**The number of times on average I use this Ax each month is:**
4.Blankenship	Yes →	Always Often Sometimes	Always Often Sometimes	<1 1-2 3-5
	No ↓	Seldom Never	Seldom Never	>5
5.Epic	Yes →	Always Often Sometimes	Always Often Sometimes	<1 1-2 3-5
	No ↓	Seldom Never	Seldom Never	>5
6.Ergos	Yes →	Always Often Sometimes	Always Often Sometimes	<1 1-2 3-5
	No ↓	Seldom Never	Seldom Never	>5
7.Isernhagen	Yes →	Always Often Sometimes	Always Often Sometimes	<1 1-2 3-5
	No ↓	Seldom Never	Seldom Never	>5
8.Keys	Yes →	Always Often Sometimes	Always Often Sometimes	<1 1-2 3-5
	No ↓	Seldom Never	Seldom Never	>5

Assessment type	**Use?**	**I use all components of this Ax:**	**I use only parts of this Ax:**	**The number of times on average I use this Ax each month is:**
9. Lido Worksheet.	Yes ➔ No ↓	Always Often Sometimes Seldom Never	Always Often Sometimes Seldom Never	<1 1-2 3-5 >5
10. Pile	Yes ➔ No ↓	Always Often Sometimes Seldom Never	Always Often Sometimes Seldom Never	<1 1-2 3-5 >5
11.PWPE	Yes ➔ No ↓	Always Often Sometimes Seldom Never	Always Often Sometimes Seldom Never	<1 1-2 3-5 >5
12.Valpar	Yes ➔ No ↓	Always Often Sometimes Seldom Never	Always Often Sometimes Seldom Never	<1 1-2 3-5 >5
13.West	Yes ➔ No ↓	Always Often Sometimes Seldom Never	Always Often Sometimes Seldom Never	<1 1-2 3-5 >5
14.Workability	Yes ➔ No ↓	Always Often Sometimes Seldom Never	Always Often Sometimes Seldom Never	<1 1-2 3-5 >5
15.Workbox	Yes ➔ No ↓	Always Often Sometimes Seldom Never	Always Often Sometimes Seldom Never	<1 1-2 3-5 >5
16. Workhab	Yes ➔ No ↓	Always Often Sometimes Seldom Never	Always Often Sometimes Seldom Never	<1 1-2 3-5 >5
17.Non standardised	Yes ➔ No ↓	Always Often Sometimes Seldom Never	Always Often Sometimes Seldom Never	<1 1-2 3-5 >5
18.Other – please specify.	Yes ➔ No ↓	Always Often Sometimes Seldom Never	Always Often Sometimes Seldom Never	<1 1-2 3-5 >5

19. How frequently would you conduct a workplace assessment AND a FCE with the same client?

Always Often Sometimes Seldom Never

20. How frequently do you conduct workplace assessments instead of FCE's?

Always Often Sometimes Seldom Never

When you have a choice to use different assessments, how important are the following qualities, criteria or issues in determining the choice you make?

21. Availability at my workplace

Extremely important	Highly important	Moderately important	Slightly important	Not important
5	4	3	2	1

22. How much flexibility it allows within the assessment

Extremely important	Highly important	Moderately important	Slightly important	Not important
5	4	3	2	1

23. If it is standardised

Extremely important	Highly important	Moderately important	Slightly important	Not important
5	4	3	2	1

24. Time effectiveness

Extremely important	Highly important	Moderately important	Slightly important	Not important
5	4	3	2	1

25. Personal knowledge of published literature regarding reliability and validity.

Extremely important	Highly important	Moderately important	Slightly important	Not important
5	4	3	2	1

26. Characteristics of the assessment tasks.

Extremely important	Highly important	Moderately important	Slightly important	Not important
5	4	3	2	1

27. My accreditation / training to conduct the assessment.

Extremely important	Highly important	Moderately important	Slightly important	Not important
5	4	3	2	1

28. How long it would take to complete.

Extremely important	Highly important	Moderately important	Slightly important	Not important
5	4	3	2	1

29. How costly it is for the referrer.

Extremely important	Highly important	Moderately important	Slightly important	Not important
5	4	3	2	1

30. The effort required to administer the assessment.

Extremely important	Highly important	Moderately important	Slightly important	Not important
5	4	3	2	1

Please answer the following questions related to your practice / use of FCE's.

31. Do you have a choice of which FCE you will use?

YES □ **please answer questions 32 – 39.**

NO □ **go to question 40.**

32. How often are you the <u>ONLY</u> person choosing which FCE you will use?

Always 1 2 3 4 5 6 7 Never

33. Is your choice of assessment influenced by the referrer?

NO □ **go to question 34.**

YES □ ↓

How often are you influenced in your choice of assessment by the referrer?

Always 1 2 3 4 5 6 7 Never

34. Is your choice of assessment influenced by the client?

NO □ **go to question 35.**

YES □ ↓

How often are you influenced in your choice of assessment by the client?

Always 1 2 3 4 5 6 7 Never

35. Is your choice of assessment influenced by the doctor?

NO □ **go to question 36.**

YES □ ↓

How often are you influenced in your choice of assessment by the doctor

Always 1 2 3 4 5 6 7 Never

36. Is your choice of assessment influenced by the insurer?

NO □ **go to question 37.**

YES □ ↓

How often are you influenced in your choice of assessment by the insurer

Always 1 2 3 4 5 6 7 Never

37. Is your choice of assessment influenced by the client's employer?

NO □ **go to question 38.**

YES □ ↓

How often are you influenced in your choice of assessment by the client's employer?

Always 1 2 3 4 5 6 7 Never

38. Is your choice of assessment influenced by your employer?

NO □ **go to question 39.**

YES □ ↓

How often are you influenced in your choice of assessment by your employer

Always 1 2 3 4 5 6 7 Never

39. How often do you complete the assessment in it's entirety?

Always 1 2 3 4 5 6 7 Never

Consider for a moment the FCE that you MOST OFTEN USE. For that FCE, how likely are each of the following events?

Circle the appropriate answer.

WHEN I USE THIS ASSESSMENT:					
40. I conduct identical FCE's for all clients					
Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely
41. I adapt specific components of the FCE depending upon the clients' Injury					
Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely
42. Medical reports make me adjust the way I administer this FCE.					
Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely
43. I adapt specific components of the FCE depending upon the clients' job					
Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely
44. The clients' narratives will impact upon the FCE.					
Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely
45. My relationship with the client impacts upon the FCE.					
Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely
46. The chances of successful RTW for the client are improved.					
Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely

WHEN I USE THIS ASSESSMENT:-

47. The outcomes of client retraining are improved.

Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely

48. Case managers have an improved understanding of further rehabilitation involvement needed.

Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely

49. I know what type of work can be attempted, i.e. Manual vs sedentary.

Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely

50. It has the potential to hurt the client.

Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely

51. Some important components may be missed.

Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely

52. A clients functional abilities and limitations are clarified.

Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely

53. RTW for those returning to pre-injury duties can be predicted.

Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely

54. Manual or sedentary work can be predicted for a client.

Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely

55. It predicts return to normal duties.

Extremely Likely	Moderately Likely	A Little Likely	A Little Unlikely	Moderately Unlikely	Extremely Unlikely

Please answer the following questions relating to your performance of conducting FCE's.

56. I feel highly skilled to perform this FCE.

Strongly agree 1 2 3 4 5 6 7 Strongly disagree

57. I have adequate knowledge of body mechanics and anatomy to conduct this FCE.

Strongly agree 1 2 3 4 5 6 7 Strongly disagree

58. I am competent to perform this FCE with a range of clients and injury types.

Strongly agree 1 2 3 4 5 6 7 Strongly disagree

59. My undergraduate education provided me with adequate skills to perform this FCE.

Strongly agree 1 2 3 4 5 6 7 Strongly disagree

60. I have had to complete further training/ study to be able to complete this FCE.

Strongly agree 1 2 3 4 5 6 7 Strongly disagree

61. I need to continually update my knowledge to be able to successfully conduct this FCE.

Strongly agree 1 2 3 4 5 6 7 Strongly disagree

62. Further comments.
Please use this space to record any additional comments you may have related to FCE's.

Thank you ☺

for taking the time to complete this questionnaire. Your participation is very much appreciated.

Appendix 4: Content validity online survey instrument - Application of an FCE.

1. Introduction

Thank you for completing this survey which is being conducted by Carole James, as part of her doctoral studies under the supervision of Dr Lynette Mackenzie (conjoint lecturer) and Professor Mike Capra (conjoint professor) from the School of Health Sciences at the University of Newcastle.
The purpose of the research is to evaluate content validity specifically of the WorkHab FCE, however the information presented is relevant for many FCE's as it links to the physical demands as detailed in the Dictionary of Occupatonal Titles, a copy of these definitions was attached to the email sent to you. You may find it useful to have access to this for question 14 & 15.
Any professional who currently performs FCE's is invited to participate in this study.
This project will add to the body of knowledge and build evidence for practice, so ensuring that best practice is achieved in the area of occupational rehabilitation. It will improve understanding of the assessment tool and its validity in the occupational rehabilitation setting.
If you have any questions, please contact Carole James at Carole.James@newcastle.edu.au.

Completion of this survey implies consent and all survey information is provided anonymously.

1. Please indicate your profession.

- [ ] Physiotherapist
- [ ] Occupational Therapist
- [ ] Occupational Health Nurse
- [ ] Exercise Physiologist
- [ ] Other

Other (please specify)

2. Do you work in Australia?

	ACT	NSW	NT	QLD	SA	TAS	VIC	WA
Yes	☐	☐	☐	☐	☐	☐	☐	☐

If NO please indicate the State and Country in which you work.

3. How many years clinical experience do you have?

4. How many years experience do you have conducting FCE's?

5. Do you have Post graduate qualifications?

- ◯ Yes
- ◯ No

If Yes please specify

2. Relevance of FCE components to work.

The following three questions look at item relevance, in relation to validity, for each of the common components within an FCE. You are asked to rate the common components of FCE's listed in relation to the different work types indicating how relevant you believe these different FCE components are to the different work types identified.

6. Please identify the relevance of the following FCE components to SEDENTARY work (work involving sitting most of the time, occasional walking and standing and occasional use of force to lift or move objects).

	Always relevant	Often relevant	Sometimes relevant	Rarely relevant	Never relevant	Not Applicable
Grip strength	◯	◯	◯	◯	◯	◯
Lifting: floor to bench	◯	◯	◯	◯	◯	◯
Lifting: bench to shoulder	◯	◯	◯	◯	◯	◯
Lifting: bench to overhead	◯	◯	◯	◯	◯	◯
Carrying	◯	◯	◯	◯	◯	◯
Pushing/ Pulling	◯	◯	◯	◯	◯	◯
Reaching	◯	◯	◯	◯	◯	◯
Standing	◯	◯	◯	◯	◯	◯
Sitting	◯	◯	◯	◯	◯	◯
Walking	◯	◯	◯	◯	◯	◯
Climbing	◯	◯	◯	◯	◯	◯
Stooping	◯	◯	◯	◯	◯	◯
Kneeling	◯	◯	◯	◯	◯	◯
Balancing	◯	◯	◯	◯	◯	◯
Crouching	◯	◯	◯	◯	◯	◯
Squatting	◯	◯	◯	◯	◯	◯
Crawling	◯	◯	◯	◯	◯	◯
Job simulation tasks	◯	◯	◯	◯	◯	◯

7. Please identify the relevance of the following FCE components to MANUAL work (physically demanding work).

	Always relevant	Often relevant	Sometimes relevant	Rarely relevant	Never relevant	Not Applicable
Grip strength	◯	◯	◯	◯	◯	◯
Lifting: floor to bench	◯	◯	◯	◯	◯	◯
Lifting: bench to shoulder	◯	◯	◯	◯	◯	◯
Lifting: bench to overhead	◯	◯	◯	◯	◯	◯
Carrying	◯	◯	◯	◯	◯	◯
Pushing/ Pulling	◯	◯	◯	◯	◯	◯
Reaching	◯	◯	◯	◯	◯	◯
Standing	◯	◯	◯	◯	◯	◯
Sitting	◯	◯	◯	◯	◯	◯
Walking	◯	◯	◯	◯	◯	◯
Climbing	◯	◯	◯	◯	◯	◯
Stooping	◯	◯	◯	◯	◯	◯
Kneeling	◯	◯	◯	◯	◯	◯
Balancing	◯	◯	◯	◯	◯	◯
Crouching	◯	◯	◯	◯	◯	◯
Squatting	◯	◯	◯	◯	◯	◯
Crawling	◯	◯	◯	◯	◯	◯
Job simulation tasks	◯	◯	◯	◯	◯	◯

8. Please identify the relevance of the following FCE components to VOCATIONAL RETRAINING.

	Always relevant	Often relevant	Sometimes relevant	Rarely relevant	Never relevant	Not Applicable
Grip strength	◯	◯	◯	◯	◯	◯
Lifting: floor to bench	◯	◯	◯	◯	◯	◯
Lifting: bench to shoulder	◯	◯	◯	◯	◯	◯
Lifting: bench to overhead	◯	◯	◯	◯	◯	◯
Carrying	◯	◯	◯	◯	◯	◯
Pushing/ Pulling	◯	◯	◯	◯	◯	◯
Reaching	◯	◯	◯	◯	◯	◯
Standing	◯	◯	◯	◯	◯	◯
Sitting	◯	◯	◯	◯	◯	◯
Walking	◯	◯	◯	◯	◯	◯
Climbing	◯	◯	◯	◯	◯	◯
Stooping	◯	◯	◯	◯	◯	◯
Kneeling	◯	◯	◯	◯	◯	◯
Balancing	◯	◯	◯	◯	◯	◯
Crouching	◯	◯	◯	◯	◯	◯
Squatting	◯	◯	◯	◯	◯	◯
Crawling	◯	◯	◯	◯	◯	◯
Job simulation tasks	◯	◯	◯	◯	◯	◯

9. Do you have any other comments to add to your responses to the relevance of FCE components section?

3. Difficulty of FCE components

The next three questions look at item difficulty in relation to validity.
When considering a typical client with the identified injury type, please indicate how difficult you believe the different commonly used FCE components are for this client type to complete. Base this on maximum capacity.

10. When considering a client with a LOWER BACK INJURY - how difficult are the following FCE components?

	Extremely difficult	Very difficult	Moderately difficult	Slightly difficult	Not difficult	N/A
Grip strength	◯	◯	◯	◯	◯	◯
Lifting: floor to bench	◯	◯	◯	◯	◯	◯
Lifting: bench to shoulder	◯	◯	◯	◯	◯	◯
Lifting: bench to overhead	◯	◯	◯	◯	◯	◯
Carrying	◯	◯	◯	◯	◯	◯
Pushing/ Pulling	◯	◯	◯	◯	◯	◯
Reaching	◯	◯	◯	◯	◯	◯
Standing	◯	◯	◯	◯	◯	◯
Sitting	◯	◯	◯	◯	◯	◯
Walking	◯	◯	◯	◯	◯	◯
Climbing	◯	◯	◯	◯	◯	◯
Stooping	◯	◯	◯	◯	◯	◯
Kneeling	◯	◯	◯	◯	◯	◯
Balancing	◯	◯	◯	◯	◯	◯
Crouching	◯	◯	◯	◯	◯	◯
Squatting	◯	◯	◯	◯	◯	◯
Crawling	◯	◯	◯	◯	◯	◯
Job simulation tasks	◯	◯	◯	◯	◯	◯

11. When considering a typical client with a UPPER LIMB INJURY(including shoulder, elbow, arm, hand injury) how difficult are the following FCE components?

	Extremely difficult	Very difficult	Moderately difficult	Slightly difficult	Not difficult	N/A
Grip strength	◯	◯	◯	◯	◯	◯
Lifting: floor to bench	◯	◯	◯	◯	◯	◯
Lifting: bench to shoulder	◯	◯	◯	◯	◯	◯
Lifting: bench to overhead	◯	◯	◯	◯	◯	◯
Carrying	◯	◯	◯	◯	◯	◯
Pushing/ Pulling	◯	◯	◯	◯	◯	◯
Reaching	◯	◯	◯	◯	◯	◯
Standing	◯	◯	◯	◯	◯	◯
Sitting	◯	◯	◯	◯	◯	◯
Walking	◯	◯	◯	◯	◯	◯
Climbing	◯	◯	◯	◯	◯	◯
Stooping	◯	◯	◯	◯	◯	◯
Kneeling	◯	◯	◯	◯	◯	◯
Balancing	◯	◯	◯	◯	◯	◯
Crouching	◯	◯	◯	◯	◯	◯
Squatting	◯	◯	◯	◯	◯	◯
Crawling	◯	◯	◯	◯	◯	◯
Job simulation tasks	◯	◯	◯	◯	◯	◯

12. When considering a client with a LOWER LIMB INJURY (including hip, knee, ankle and foot injury) how difficult are the following FCE components?

	Extremely difficult	Very difficult	Moderately difficult	Slightly difficult	Not difficult	N/A
Grip strength	◯	◯	◯	◯	◯	◯
Lifting: floor to bench	◯	◯	◯	◯	◯	◯
Lifting: bench to shoulder	◯	◯	◯	◯	◯	◯
Lifting: bench to overhead	◯	◯	◯	◯	◯	◯
Carrying	◯	◯	◯	◯	◯	◯
Pushing/ Pulling	◯	◯	◯	◯	◯	◯
Reaching	◯	◯	◯	◯	◯	◯
Standing	◯	◯	◯	◯	◯	◯
Sitting	◯	◯	◯	◯	◯	◯
Walking	◯	◯	◯	◯	◯	◯
Climbing	◯	◯	◯	◯	◯	◯
Stooping	◯	◯	◯	◯	◯	◯
Kneeling	◯	◯	◯	◯	◯	◯
Balancing	◯	◯	◯	◯	◯	◯
Crouching	◯	◯	◯	◯	◯	◯
Squatting	◯	◯	◯	◯	◯	◯
Crawling	◯	◯	◯	◯	◯	◯
Job simulation tasks	◯	◯	◯	◯	◯	◯

13. Do you have any other comments to add to your responses to the difficulty of FCE components section?

4. FCE and the Dictionary of Occupational Titles physical demands

Question 14 & 15 investigate the relationship between common FCE components and the Physical demands as outlined in the Dictionary of Occupational Titles. You may wish to refer to the definitions of the DOT physical demands sent with the email.

14. Please indicate where the common components of an FCE (on the vertical axes) link with the DOT physical demands (along the horizontal axes) of this matrix.

For example:
in the case of lifting from floor to bench you may choose to fill out the matrix indicating a link between the following items in the DOT:
Lift; carry; reach; stand; walk; stoop; crouch; balance; handle; finger; see.

A 'Maybe' response should be indicated as a 'NO'.

	Lift	Carry	Push	Pull	Reach	Stand	Sit	Walk	Climb	Stoop	Kneel	Crouch	Crawl	Balance	Handle	Finger	Feel	Talk	See	Hear
Grip Strength	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Lifting: floor to bench	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Lifting: bench to shoulder	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Lifting: bench to overhead	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Carrying	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pushing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pulling	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Reaching	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Standing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

15. Please indicate where the common components of an FCE (on the vertical axes) link with the DOT physical demands (along the horizontal axes) of this matrix as in question 14.

A 'Maybe' response should be indicated as a 'NO'.

	Lift	Carry	Push	Pull	Reach	Stand	Sit	Walk	Climb	Stoop	Kneel	Crouch	Crawl	Balance	Handle	Finger	Feel	Talk	See
Sitting	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Walking	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Climbing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Stooping	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Kneeling	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Balancing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Crouching/Squatting	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Crawling	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Job Simulation	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

16. Do you have any other comments about using FCE's in practice?