ACARP Project C14045 **Final Report** May 2007

ACARP

Australian Coal Association Research Program

Validating the JobFit System Functional Assessment Method

Researchers:

Jenny Legge
JobFit Systems International

Robin Burgess-Limerick
Burgess-Limerick & Associates

Table of Contents

Primary Contact

Jenny Legge BPhty MErg
Managing Director
JobFit Systems International Pty Ltd
ABN 78 104 491 714
PO Box 8740, Mt Pleasant Q 4740
Australia
P: 07 4954 8652
F: 07 4954 8654
M: 0438 426 477
E: jenny.legge@jobfitsystem.com
www.jobfitsystem.com

Executive Summary

According to the Australian Safety and Compensation Council (formerly known as the National Occupational Health and Safety Commission), workplace injuries are costing the Australian coal mining industry and its communities $410 Million a year. With injuries from sprains and strains accounting for about half of these costs, the industry needs to explore alternative ways of measuring and managing the associated risks in an effort to control these staggering costs.

As the industry evolves and matures, functional capacity testing in the pre-employment or post-offer phase of recruitment is increasing in popularity as a positive injury prevention, wellness and health surveillance tool to supplement other medical assessments in the workplace injury risk management process. These assessments typically consist of a series of tests for mobility, strength, fitness, tolerance to different positions and movements, as well as material handling ability like lifting, carrying, pushing and pulling. Results are often compared to job demands to assist with decisions regarding job placement, task redesign and other risk management strategies such as physical conditioning programs.

Despite the limited published research examining the reliability and validity of functional capacity assessments they are becoming more widely used. With increasing pressure from all stakeholders (legal and health practitioners, workers, insurers and employers) the demand for evidence-based practice is rising. This ACARP study aims to meet those demands by developing a safe, reliable and valid pre-employment functional assessment tool.

All JobFit System Pre-Employment Functional Assessments (PEFAs) consist of a musculoskeletal screen, balance test, aerobic fitness test and job-specific postural tolerances and material handling tasks. The results of each component are compared to the applicant's job demands and an overall PEFA score between 1 and 4 is given with 1 being the better score.

The reliability study and validity study were conducted concurrently. The reliability study examined test-retest, intra-tester and inter-tester reliability of the JobFit System Functional Assessment Method. Overall, good to excellent reliability was found, which was sufficient to be used for comparison with injury data for determining the validity of the assessment. The overall assessment score and material handling tasks had the greatest reliability. Results have been presented at national and international conferences and published in peer-reviewed journals.

The validity study compared the assessment results of 336 records from a Queensland underground and open cut coal mine with their injury records. A predictive relationship was found between PEFA score and the risk of a back/ trunk/ shoulder injury from manual handling (RR 3.56, 95% CI 1.5 to 8.47). An association was also found between PEFA score of 1 and increased length of employment. Lower aerobic fitness test results had an inverse relationship with injury rates. The study found that underground workers, regardless of PEFA score, were more likely to have an injury when compared to other departments. No relationship was found between age and risk of injury. These results confirm the validity of the JobFit System Functional Assessment method.

A number of conclusions and recommendations were drawn from this project. In summary, the reliability and validity of the JobFit System Functional Health Assessment method means that the Australian mining industry will have the confidence and evidence to use this tool and its standardised processes as a component of their risk management activities for preventing sprains and strains in the workplace. Accurate job demands are critical not only to the validity of the functional assessment but also for ergonomic risk assessments and controls. Both job analyses

and worker assessments are snapshots of a moment in time and both need to be reassessed on a regular basis to ensure that the most accurate information is gathered for decision-making purposes. Pre-employment assessments are not foolproof indicators of the occurrence or absence of workplace injury and therefore are only part of the solution. Communication and cooperation between industry stakeholders is vital not only for accurate and comprehensive assessment of data but also if we are to achieve our common goal of reducing the number and costs of sprains and strains in the Australian coal mining industry.

Overview & Objectives

Overview

It is estimated that sprains **and strains are costing the Australian coal mining industry and communities more than $ 7 million every week** in direct costs and indirect costs (NOHSC, 2004). The social costs of workplace injuries also need to be considered. Whilst injury rates are slowly improving, around half continue to be from sprains and strains. With the mining industry facing a competitive and aging labour market, employers are recognising an increased need to better manage the health and wellness of their current workforce to improve both performance and retention.

There have been a number of strategies employed to determine or attempt to minimize a worker's future risk of injury including back X-rays, manual handling training, history of previous pain and medical screenings including strength and endurance and body composition testing but there is limited evidence of their success (Bigos & Battie 1987, Reimer *et al* 1994, Snook 1987, Mooney *et al* 1996, Mostardi *et al* 1992).

As the industry evolves and matures, functional capacity testing in the pre-employment or post-offer phase of recruitment is increasing in popularity as a positive injury prevention, wellness and health surveillance tool to supplement other medical assessments in the workplace injury risk management process.

These assessments typically consist of a series of tests for mobility, strength, fitness, tolerance to different positions and movements, as well as material handling ability like lifting, carrying, pushing and pulling. Results are often compared to job demands to assist with decisions regarding job placement, task redesign and other risk management strategies such as physical conditioning programs.

Despite the limited published research examining the reliability and validity of functional capacity assessments they have become widely used (Legge 2004). With increasing pressure from all stakeholders (legal and health practitioners, workers, insurers and employers) the demand for evidence-based practice is rising. This study which has been funded by ACARP aims to meet those demands.

Based on the National Institute for Occupational Safety and Health (NIOSH) criteria for the development and selection of work-related assessments there are five key attributes of excellence for work-related assessments: safety, reliability, validity, practicality and utility (Innes & Straker 2003). Reliability relates to the level of consistency or repeatability between measurements and validity relates to its predictability and transferability to the workplace. Both of these factors will be subject of this research project.

Reliability is commonly measured three ways. Test-retest reliability is an indicator of the stability of the test. Inter-tester reliability is often described as the objectivity of the test and intra-tester reliability refers to the consistency of the test. Practitioners using a reliable testing method can be confident that changes in performance on two different occasions can be attributed to change in the participant (eg. effort, motivation, conditioning) rather than as a result of variations in testing procedures and interpretation. Reliability of a testing procedure needs to be determined before validity can be tested.

There is limited published research investigating the reliability of functional assessments, particularly in healthy workers which is the assumption in a pre-employment situation. Of those that were reviewed there appears to be some consistency between lower reliability scores for above shoulder lifts and tolerance to reaching forward and squatting, however there is a degree of variation between different functional capacity assessment methods (Gross & Battie 2002, Reneman et al 2004, Reneman et al 2002, Tuckwell 2002, Durand et al 2004).

The reliability study results for this project were previously reported in the September 2005 progress report and will be presented in detail in the 'Results – Secondary Study' section of this report. The results have also been published in 'Work', an international peer-reviewed journal. Overall, good to excellent reliability was achieved.

Validity of pre-employment assessment tests can be measured by comparing test performance to injury rates, types of injuries, costs and duration of injuries, turnover rates and productivity. Of course, many factors influence these measures so whilst they are not the perfect indicators, they can provide reasonable feedback.

Preliminary results for the validity study were presented in the August 2006 Interim report. Although promising, the main conclusion of the study at that time was that more data was required. As a result, injury statistics and assessment results were collected for an additional year until January 2007. Final results for the validity study will be presented in detail in the 'Results – Primary Study' section of this report. A statistically significant relationship was identified between PEFA score and injury risk and PEFA score and length of employment.

Objectives

There were three key objectives of this research project:

1. Develop a commercially viable functional health assessment method to assist in determining the suitability of current and prospective personnel for particular work functions throughout the coal industry.
2. Reduce the incidence and severity and the direct and indirect costs of work-related musculoskeletal disorders through the use of a functional health assessment method in conjunction with the "JobFit System" software program.
3. Determine if there is a predictive relationship between performance in a Pre-Employment Functional Assessment, as a whole or in parts, and workplace musculoskeletal injury rates.

Program & Work Schedule

The project has been broken up into two concurrent studies. The primary study (validity study) and the secondary study (reliability study). The work schedule and progress for each is tabulated below (Table 1).

Table 1
Project Schedule & Progress

	Schedule & Progress	*Status*
Secondary Study		
Step One Collection of data	o Study conducted onsite at Newlands Coal Mine with 28 current employees	Complete
Step Two Analyse, report and presentation of results	o Analysis complete o Results presented (poster) at Qld Mining Industry annual safety conference Aug 05 o Results reported and disseminated to interested parties including Qld NMA, Coal Services, CFMEU and industry representatives o Paper (in press) published in international peer-reviewed journal 'Work' o Results presented at formal and informal Australian physiotherapy, occupational therapy and ergonomics forums	Complete
Primary Study		
Step One Collection of PEFA test results	o PEFAs conducted at Newlands Coal Mine since Dec 02 to Dec 06 o Data from Dec 02 to Dec 06 collected	Complete
Step Two Collection of injury statistics	o Stats from pre-test period (Jan 02 to Dec 02) collected o Stats from Jan 02 to Dec 06 collected o Available data categorised	Complete
Step Three Analysis, report and presentation of results	o Analysis complete o Results presented at NSW Minerals Council annual safety conference June 06 o Results presented at Qld Mining Industry annual safety conference in Aug 06 with application for Aug 07 o Results will be disseminated through ACARP, email and web channels to interested parties upon approval of final report o Results presented at author's expense at other industry and professional conferences	Complete

Major Milestones

- Reliability research reported
    - ACARP C14045 Progress Report
      September 2005

- Reliability research presented
    - Poster Queensland Mining Industry Annual Safety Conference
      August 2005
    - Publication in 'Work'
      (in press)

- Validity research reported
    - ACARP C14045 Progress Report
      August 2006
    - ACARP C14045 Final Report
      April 2007
    - Submissions for publication will also be made to peer-reviewed international journals

- Validity research presented (interim results)
    - NSW Minerals Council Annual Safety Conference
      June 2006
    - Queensland Mining Industry Annual Safety Conference
      August 2006
    - Brisbane Safety Conference
      June 2006
    - Sydney Safety Conference
      October 2006
    - Human Factors & Ergonomics Society of Australia Annual Conference
      November 2006
    - Applied Ergonomics Conference (Dallas, USA)
      March 2007
    - Human Factors In Safety Conference
      March 2007

- Validity research planned* future presentations (final results)
    - Human Factors: Reliability and Safety Conference
      June 2007
    - Queensland Mining Industry Annual Safety Conference
      August 2007
    - Sydney Safety Conference
      October 2007
    - Australian Physiotherapy Association Annual Conference
      October 2007
    - Human Factors & Ergonomics Society of Australia Annual Conference
      November 2007

* Pending acceptance of abstract submission

Introduction

Work-related musculoskeletal injuries cost companies millions of dollars every year in the form of reduced productivity, replacement wages, medical costs, lump sum payments and performance-based workers compensation premiums. According to the Australian Safety and Compensation Council's most recent report (2006), in 2003-04 there were 2105 new claims in the Australian mining industry. The industry's incidence rate is 24 per 1000 employees which is well above the national average of 16. The frequency rate of 10.7 per million hours worked is close to the national average of 9.9. Labourers, tradespersons and related occupations accounted for 44.8% of all claims with the incidence rate for labourers of 45.2 and tradespersons 28.8 well above the national average.

Consistent with previous years, sprains and strains accounted for around half of all claims (45.6%) with the back being the most commonly injured body part (24.3%). Muscular stress while lifting, carrying or putting down a load accounted for 19.8% of all new claims, and when combined with all 'muscular stress' causes, the contribution increased to 41.6% of all new claims.

The National Occupational Health and Safety Commission in their 2004 report on 'The Cost of Work-related Injury and Illness for Australian Employers, Workers and the Community" cited the mining industry as having the highest average total direct and indirect cost of $194 800 per case. When multiplied by the 2105 reported new claims in 2003-04 this totals $410 Million for the year.

The Queensland Mines and Quarries Safety Performance and Health Report 2005-06, reported that in the 2005-6 year, the Queensland coal mining sector incurred 207 lost time injuries which is an increase on last year's 177. The average lost time injury frequency rate (4.5) and severity rate (105) both improved however the underground coal mining sector's rates increased. The most recent workers' compensation data from 2004-05 recorded 782 claims costing $6 million in workers' compensation costs only. Sprains and strains were the most common cause of injury accounting for 62% of all new claims which is above the average for all Australian industries.

The Queensland Workplace Health and Safety Act 1995 Section 28 (1) states "An employer has an obligation to ensure the workplace health and safety of each of the employer's workers at work". Other jurisdictions within Australia have equal requirements. In relation to manual tasks, this is typically achieved by modifying tasks and equipment in an effort to match the task to the human. Sometimes, due to technical or cost considerations, this approach becomes impractical and the shift then changes to matching the worker to the task. A review of the peer-reviewed literature finds that there have been a number of studies on the effectiveness and validity of pre-employment and pre-placement assessments. Many of these have focused on only one or two aspects of the assessment, such as fitness, strength or material handling (Legge, 2004). None were identified that focused on all aspects that were included in this research project nor allocated an overall performance score for comparative purposes. The general consensus from all of the reviewed articles is that much research still needs to be done in the area of pre-employment and pre-placement assessments.

The literature review highlighted that there have been a number of strategies employed to determine or attempt to minimise a worker's future risk of injury including back X-rays, manual handling training, history of previous pain and medical screenings including strength and endurance and body composition testing but there is limited evidence of their success (Bigos & Battié, 1987; Reimer et al, 1994; Snook, 1987). These items typically make up a large portion of the musculoskeletal component of the current Coal Board Medical. A more recent approach in employee assessment is the use of pre-employment or post-offer functional assessments with the majority centred around the format of Functional Capacity Evaluations. A Pre-Employment Functional Assessment (PEFA) is a series of tests that provide objective information about a worker's functional performance in relation to the job for which they are applying.

Based on the National Institute for Occupational Safety and Health (NIOSH) criteria for the development and selection of work-related assessments, Innes and Straker (2003) summarise the key attributes of an assessment as: safety, reliability, validity, practicality and utility (Table 2). This study after controlling for safety, will examine the reliability and validity of Pre-Employment Functional Assessments.

Table 2
Key Attributes of Work-Related Assessments

Safety	Is the test safe to administer?
Reliability	Are the test results reproducible on any occasion between evaluators (inter-rater) and participants (test-retest)?
Validity	Does the test measure what it reports to measure and is it predictive of performance?
Practicality	Is the test easy to administer with reasonable / minimal cost?
Utility	Does the functional test relate to job performance and does it meet the needs of the involved parties?

This table is based on Randolph (2000), Innes & Straker (2003), King et al (1998)

A key distinction needs to be made when evaluating the effectiveness of pre-employment functional assessments (PEFAs) in controlling work-related musculoskeletal disorders. That is, whether the assessment is being used to assess an individual's current safe working capacity or whether it is being used as a predictor of injury. Anderson (1999) believes that the emphasis of a PEFA should be on objective information such as an individual's ability to perform the job rather than speculative conclusions such as risk of injury that may occur in the future. This approach would also be most consistent with current anti-discrimination legislation. Whilst it may seem that these are essentially the same thing, a review of the literature indicates that they may need to be analysed as two separate issues to obtain accurate data on their effectiveness. The author believes that much of the confusion in the literature and thus for consumers of these products is because these two issues are not clearly identified (Legge, 2004).

Another reason for the inconclusiveness of the available literature on the effectiveness of these tools is that in the vast majority of cases the focus of these studies has been on back injuries and back function to the exclusion of the rest of the body which accounts for just over half of the remaining work-related musculoskeletal injuries.

The design of a PEFA, as was the case in this research project, typically consists of the following activities:

- physical and musculoskeletal screen
- fitness test
- postural tolerances and dynamic activities
- manual handling tasks

Physical screening is used to identify any conditions such as elevated blood pressure, cardiovascular or respiratory abnormalities or restricted limb movement which may prevent the worker from safely participating in the required functional tasks. They can also be used to screen for any current injuries or injuries common to the job for which they are applying (Scott, 2002).

In the past, and unfortunately in many cases still today, trunk mobility and muscle strength are also tested in an effort to predict worker performance. There are numerous publications that refute the inclusion of these tests for this purpose alone. Mooney et al (1996) in a study of 152 shipyard workers found no evidence that isometric (static) strength testing of back extensors would predict workplace back injury. Isokinetic (constant speed) back strength testing of a group of 171 nurses, as well as past history of reported pain, were also found to be poor indicators of low back pain or injury in work-related manual tasks (Mostardi et al, 1992). However, there may be some bias to this study in that only volunteers, and thus those with confidence in their performance, were tested (Legge, 2004). These results are not surprising considering that neither isometric nor isokinetic strength are functional measures of lifting performance. In addition, a designer of any physical screening test needs to consider that isolated muscle strength tests are not job-specific and may not be justifiable under current anti-discrimination legal requirements.

Fitness tests are designed primarily to determine whether the worker has the aerobic capacity to perform the required tasks based on aerobic requirements identified in the initial task analysis. The aerobic requirements of the tasks were not been identified for this research project and thus aerobic fitness testing will only be included as a potential predictor of injury.

Aerobic physical fitness is not infrequently included as a predictor of physical injury. Numerous studies including a study of a group of 1652 firefighters by Cady et al (1979) have indicated that there is a graded protective effect for added levels of fitness against the incidence and cost of back injuries. Cady et al's measure of fitness was based on a total score from five items, including three of cardiovascular fitness, and one each for isometric back and leg strength and flexibility. They suggested that future studies may be able to determine if different components could be weighted separately to give more accurate predictions. This research project will attempt to identify these relationships. Based on the previously discussed limited evidence to support isometric strength and flexibility testing, it appears that aerobic fitness may be a clearer indicator. A preliminary retrospective study by Bigos and Battié (1987) also indicated that low cardiovascular fitness level is a risk factor for chronic back pain disability.

Postural tolerance and dynamic tolerance tests include activities such as reaching forward, squatting, stooping, climbing, walking and balancing. Again, their inclusion should be based on the job analysis. Procedures for assessing these tasks are extremely varied and their reliability depends greatly on standardized procedures for assessment. Information directly related to these tasks was scarce in the peer-reviewed published literature and could only be identified in product training manuals. Much work is needed in the research of the reliability and validity of postural tolerance and dynamic tolerance testing which is why these factors were included in this study.

There is a wealth of published information, and subsequent debate, about the methodology for, and validity of lifting assessments. There are two main topics of debate. Firstly, what comprises safe lifting? Secondly, which is a more accurate predictor of performance – isometric, isokinetic, kinesiophysical, functional, or isoinertial tests? In consideration of comments previously made about including assessment tasks consistent with actual work tasks, the tests which preserve ecological validity as far as possible would seem to be the most obvious choice. Battié et al (1989) in a four-year follow up study of 3020 voluntary aircraft manufacture workers failed to demonstrate that isometric lifting strength in either a torso, arm or leg lift position was indicative of an ability to predict that an individual was at risk of industrial back problems. Interestingly, partway through the initial testing phase, the torso lift (straight legs and bent forward position) was discontinued following a number of participant injuries. It is common knowledge that the power lift or a modified leg lift is the current preferred method for lifting. This method was employed in this research project.

The validity of the lifting component of the PEFA also relates to an ability to translate information obtained during the assessment relating to a participant's occasional lifting capacity (up to 33% of a workday) to that of a frequent lifting capacity (33% to 67%) or more. Saunders et al (1997) concluded that estimates of frequent lifting capacity can be made from occasional lifting capacity but that the usefulness of these estimates is questionable and such estimates should be used with caution. When these lifting assessments are transferred for application in an industrial environment there are additional limitations that need to be acknowledged. Whilst it was not specified, these estimates are typically based on an 8-hour working day and as such may not be as easily transferred to a 12-hour working day which is becoming more common in labour-intensive industries such as mining. The additional demands of awkwardness of loads, positions, team lifting and harsh environments have also not been taken into account.

Secondary (Reliability) Study

Introduction

The purpose of the secondary study was to determine the reliability of pre-employment functional assessments (PEFA) as a whole, or in parts, as a precursor for a validity study investigating the relationship between PEFA results and workplace injury rates and severity. Reliability refers to the level of consistency or repeatability between the measurements recorded for a test on different occasions (test-retest, intra-rater), and between different assessors (inter-rater). Clinically, this typically refers to obtaining the same results rather than proportional and consistent change (Innes & Straker, 1999).

There are four major sources of potential error affecting reliability:

1. Participant – fatigue and health, motivation and attitude, practice and memory, experience and knowledge
2. Testing – clarity of instruction and adherence to procedure
3. Scoring – suitability of scoring method, experience, competence, familiarity and accuracy of scorers
4. Instrumentation – calibration and setup of equipment, suitability of assessment tools (Thomas & Nelson, 2001)

The factors cited above as affecting the participant could also be applied to the assessor. These human sources of error, that is the participant and assessor, could also be influenced by environmental factors such as time of day, temperature and humidity, noise, visibility and other distractions.

Test-retest reliability

Test-retest reliability is an indicator of the stability of a test. That is, the ability to produce the same results on two different occasions on the assumption that the measure being scored does not change over time. The time between the two testing occasions varies and is a balance between the need for rest, the desire to reduce memory or avoid changes in the conditions, in the case of this study, changes in health and fitness of the participant. Sources of error in test-retest reliability could be from all four listed above, but in comparison to inter- and intra-tester reliability, it is assumed that participant and instrumentation errors would be expected to be higher.

Inter-tester reliability

Thomas & Nelson (2001) describe inter-tester reliability as "objectivity – the degree to which different testers can obtain the same scores on the same participants", or conversely is a measure of the variation between testers. Testing and scoring would be the main sources of error with this measure of reliability. To address these sources of error the majority of commercially available functional capacity testing tools have detailed procedures with which practitioners must become competent before they become 'certified' assessors.

Intra-tester reliability

Intra-tester reliability measures the consistency of scoring for an individual assessor on two different occasions. It is a form of test-retest reliability, however errors are influenced more by testing and scoring rather than participant and instrumentation sources.

Intra- and inter-rater reliability are considered to be particularly important when using subjective observations as is often the case when using work-related assessments. Reliability in work-related assessments is critically important so that any changes recorded in a worker's performance can be attributed to actual changes in their level of physical function and not simply an error in measurement (Innes & Straker, 1999). These changes could also be attributed to the worker's effort and motivation however it is assumed that in pre-employment testing processes that the candidate is putting forth their best effort. Standardization of the procedures and scoring systems is the key to reducing the 'subjectivity' and improving the objectivity (reliability) of the assessments.

Methods of measuring reliability

The degree of reliability, or consistency between two sets of scores, is typically expressed as a correlation coefficient. As the degree of variance between two sets of the same variable are being compared, intraclass correlation is the appropriate method (Thomas & Nelson, 2001). The intraclass correlation coefficient (ICC) is a number between 0 and 1. The closer to one, the higher the stability. However, the range of scores and sample size also need to be considered when interpreting the results.

Whilst there a number of different measures for interpreting each form of reliability, for simplicity and to facilitate interpretation of the results, a single respected measure, ICC, will be used. Where questions arise as to the potential suitability of this measure, percentage agreement in the raw data will also be examined and findings discussed. A review of the literature indicates that whilst there is no definitive source, it appears to be accepted, that an ICC score of < 0.75 is poor to moderate and > 0.75 is good. Portney and Watkins (in Innes & Straker, 1999) suggest that a score above or equal to 0.90 is required for clinical application to ensure valid interpretation of the findings. Gross & Battie (2002) and Reneman et al (2002) go one step further, rating an ICC > 0.90 as excellent.

Reliability literature

Despite the wide use of FCEs, there is limited published literature on the inter-, intra- and test-retest reliability of functional capacity evaluations. Of that which is available, the results indicate good reliability. Test-retest and intra-rater reliability are the most widely published.

Method

Subjects

A Queensland Coal Mine agreed to participate in the study. A total of 28 workers participated in a generic PEFA. Twenty of the participants participated in a second trial between one week and three months later. Demographic data including age and their usual role were collected. Before testing, each participant was required to sign a written consent form outlining: (i) the components of the assessment (ii) the risks and expectations of submaximal physical testing and the precautions that would be taken (iii) the purpose of the assessment and the use and disclosure of the collected information (iv) the opportunity to discontinue testing at any time. The consent form was designed to meet relevant medico legal and privacy law requirements. The study was approved by the Ethics Officer of the School of Human Movement Studies, University of Queensland. Participants were screened for exclusion factors prior to commencement of the assessment. Exclusion factors included current injury, significant injury or surgery in the last six months, elevated blood pressure (resting systolic >160mmHg or resting diastolic >95mmHg) or specific medical advice.

Experimental Design

Assessment Process

The Pre-employment Functional Assessments (PEFAs) were generic assessments representative of those used for coal miners in labour-intensive roles as identified with the JobFit System. The JobFit System is a database program that contains the key physical requirements of jobs and the physical capabilities of workers in a same-value format for immediate and objective comparison. Each task has been analysed by a physiotherapist and the following information recorded: task overview; frequency and duration; working posture requirements; material handling requirements; and any other relevant information such as environmental considerations. Working posture requirements are described as 'Never', 'Occasional', 'Frequent' or 'Continuous' as per the widely recognized US Department of Labor's Dictionary of Occupational Titles. This data is entered into the JobFit System. A Job Summary is then formulated by the JobFit System for a job based on the combined requirements of the tasks required for that job. Postural requirements for each task that were considered to be a high risk for work-related musculoskeletal disorders and the key requirements for the job were identified for inclusion in the PEFA. Material handling requirements were also identified.

Each PEFA contained the following components and was delivered in the same sequence:

1. musculoskeletal screen
2. balance test
3. aerobic fitness test (3-minute Step Test)
4. postural tolerances (Reaching forward, Reaching overhead, Stooping, Squatting, Climbing)
5. material handling tasks (Floor to bench, Bench to shoulder, Bench to Overhead, Bilateral Carry)

The procedures for each task were fully explained to the participants prior to the commencement of each task. The procedures for the aerobic fitness test, postural tolerances and material handling tasks were conducted as per the WorkHab FCE training manual (Bradbury & Roberts, 1998) and the JobFit System PEFA training module.

PEFA Score: A PEFA Score is the overall score for the worker's performance in comparison with the physical requirements of the job for which they are applying. A worker can be scored one, two, three or four. Table 3 defines each score. The overall PEFA score was determined with the use of the JobFit System. The worker's capabilities (postural tolerances and material handling tolerances only) were determined by the above assessment processes and entered into the JobFit System. Each postural and material handling requirement was assessed against the requirement of a task. If all requirements were met, the indicators and scores were green. If not, then they were yellow. If a worker's analysis displayed all green indicators, then they obtained a PEFA score of one. If they displayed one or more yellow indicators, then their record was analysed further. They scored a two if their material handling capacity was within 15% of the requirement and / or they had a single minor postural tolerance limitation. A minor postural limitation is defined as a single level difference (eg frequent to occasional). They scored a three if their material handling capacity was more than 15% of the requirement and / or they had more than one minor or one moderate postural tolerance limitation. They scored a four if a gross mismatch was present. Fitness and balance test results had no direct bearing on the PEFA score but were collected to determine their reliability prior to being used in the subsequent validity study.

Table 3
Definition of PEFA Scores

Score	*Definition*
One	Has demonstrated the functional capacity to perform the proposed position as described with no restrictions
Two	Has demonstrated the functional capacity to perform the proposed position as described with minimal restrictions (specified)
Three	Has demonstrated the functional capacity to perform the proposed position as described with moderate restrictions (specified)
Four	Has not demonstrated the functional capacity to meet the inherent requirements of the proposed position as described

Trial Groups

All twenty-eight participants completed the first trial. Twenty completed a second trial. Selection for the second trial was based on participant availability amongst those of whom one week had lapsed since their initial assessment and who had volunteered to participate in the second trial. Each live assessment was videotaped and conducted by the primary assessor. Each first and second trial video was watched by the second assessor and scored allowing a minimum one week period between watching the first and second trial videos. After a minimum period of one week had lapsed, the primary assessor also watched the videos and rescored the assessments.

Assessors

The primary assessor was a registered physiotherapist with six years experience in conducting functional capacity evaluations, five years as a registered WorkHab FCE provider and a JobFit System functional assessment trainer. The second assessor was a registered occupational therapist with one year experience in conducting functional capacity evaluations and a registered WorkHab FCE provider who had participated in the JobFit System functional assessment training program.

Data Analysis

Intraclass correlation coefficient (ICC) and percentage agreement were used to measure test-retest, intra- and inter-rater reliability. ICC scores greater than 0.75 were interpreted as good and scores greater than 0.90 were interpreted as excellent (Gross & Battie, 2002, Innes & Straker, 1999, Reneman et al, 2002a). Where disagreements occurred, raw data was examined in an effort to offer explanations for the variations.

Results

Subjects

The group consisted of 28 males aged 19 to 55 years (Mean: 35.5yrs). Half were currently employed in an office / professional role and the other 50% were employed in a labour-intensive role, the majority of which were underground coal miners. No subjects were excluded based on the musculoskeletal screen however one had temporary limitations identified in the lower limb due to pain from a recent tattoo.

PEFA Score

The JobFit System PEFA score is determined by comparing a worker's capabilities to the job demands. The worker's material handling capacity is the primary factor. The second most influential factor is their postural tolerances. Fitness and balance test results do not have a significant effect on the overall score. The results for the various test components will thus be described in this order of influence rather than the order of data collection.

The PEFA scores for all participants by department are illustrated in Figure 1. PEFA scores range from 1 to 4, with 1 being the better score. It is interesting to note that despite the huge variation in physical demands of their usual roles, on average, each group scored equally on the overall PEFA score. There were twice as many scoring 3 (moderate limitations) as there were scoring 1 or 2.

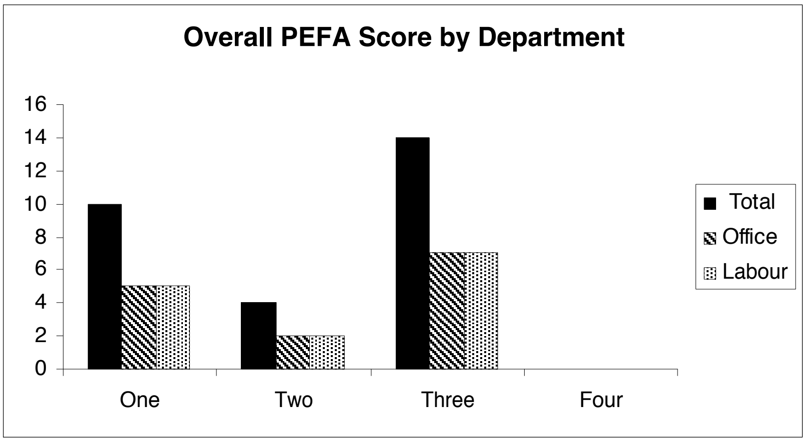

Figure 1. Overall PEFA Score by Department

Table 4
Intraclass Correlation Coefficients (ICC)
and Confidence Intervals for Overall PEFA Scores

Comparison	*ICC*	*Lower limit*	*Upper limit*
Intra-rater reliability [live vs. video (n=48)]	0.94	0.90	0.96
Inter-rater reliability [video vs. video (n=48)]	0.83	0.74	0.89
Inter-rater reliability [live vs. video (n=48)	0.84	0.75	0.90
Test-retest reliability [trial 1 vs. trial 2 (n=20)]	0.78	0.57	0.89

ICC scores indicate good to excellent reliability in determining the overall PEFA score (Table 4). One of the limitations of the ICC is that when only a small sample and small range of scores is used, a single change can have a dramatic result and can provide an inaccurate representation of the data. For this reason, actual values are discussed in the following paragraphs.

Test-retest: Twenty participants completed two trials. Sixteen (80%) of these showed consistency between trials. Three improved and one declined in performance. These are identified as participants 7, 10, 23 and 3 in figure 2 below.

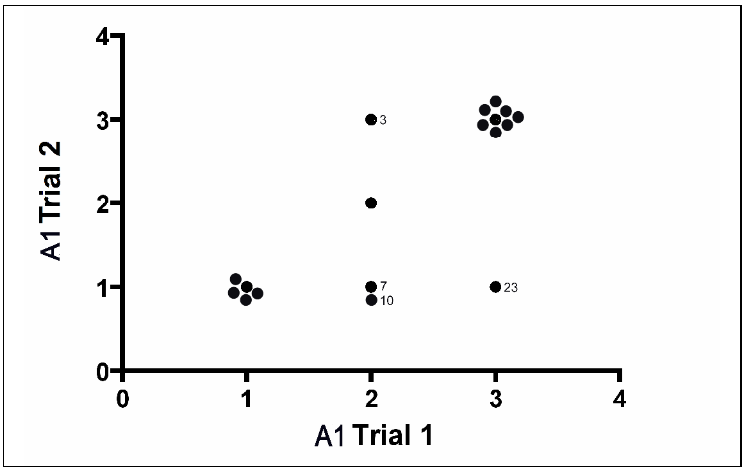

Figure 2. Test-retest Reliability for Overall Score

Participant seven improved from a PEFA score of two to one. This was a direct result of increasing his overhead lifting capacity from 30.5kg to 35kg. It was noted, that the assessing therapist stopped the participant at 30.5kg in the first trial, as it was determined that their safe lifting tolerance had been reached. The second assessor, when watching the video scored both trial one and trial two less at 23kg and 30.5kg respectively, again an improvement between trials albeit a more conservative score. Participant seven attributed his improvement to rugby training.

Participant ten also improved from a PEFA score of two to one also as a result of increasing their overhead lifting capacity from 30.5kg to 35kg. The result achieved in trial one was due to the participant stopping the test due to complaints of wrist discomfort. The second assessor did not agree with the improvement in trial two.

Participant twenty-three had the biggest improvement from three to one increasing his shoulder lift from 28kg to 35kg and his overhead lift from 23kg to 35kg. No reason was documented for these improvements. Motivation, or fatigue in the first trial, is expected to be the main contributing factor as only two weeks had passed between trials thus making a training effect unlikely. Both assessors agreed on the original and revised scores.

Participant three who declined in his performance lowered his overhead lifting capacity from 30.5kg to 28kg. His shoulder lifting capacity also decreased from 33kg to 30.5kg but this would not have affected his overall score. Both assessors agreed on the change in results. There was no reason documented for his decline in performance between trials.

When looking at the scatter plots below (figures 3 and 4), two clear trends appear:
1. the second assessor was consistently more conservative, and
2. video assessments were typically scored more conservatively than live assessments.

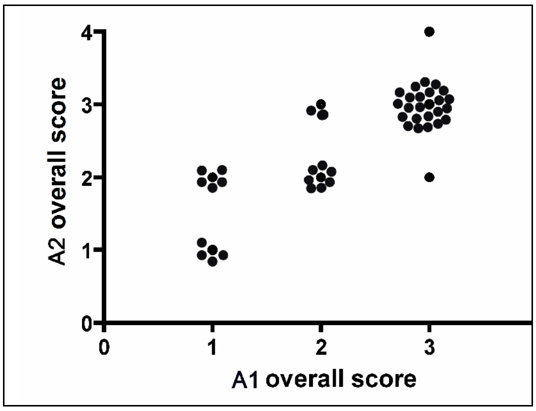

Figure 3. Inter-rater Reliability for Overall PEFA Score

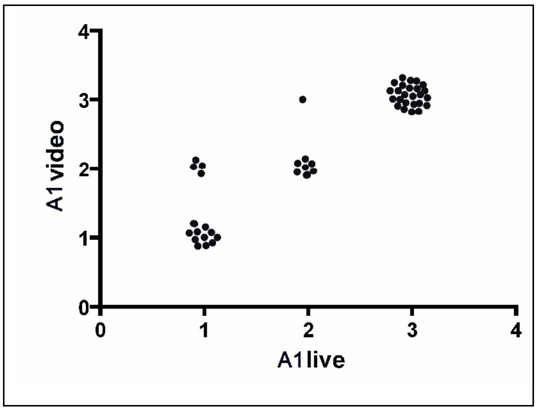

Figure 4. Intra-rater Reliability for Overall PEFA Score

Inter-rater: Eleven of the forty-eight trials (23%) varied between assessors. The main differences between the assessors were years of experience and different disciplines. As both are looking for the same signs of safe maximal lifting and it is expected that each discipline would have equivalent observational skills, it is reasonable to assume that the main contributing factor would be confidence based perhaps on years of experience or personality differences. Reneman et al (2002b) investigating the reliability of determining effort level of lifting and carrying in a functional capacity evaluation compared the inter-rater reliability of three physical therapists and two occupational therapists, four of which had only minimal experience. Reliability was expressed as a percentage and ranged from 87% to 96% which is a fair representation of the results achieved in this study. The variations between the different disciplines and the experience levels were not published and so could not be compared.

Intra-rater: In the few cases that varied between live and video scores, the video scores were typically rated lower. Three explanations are offered:

1. in the live scenario, the assessor can receive feedback from the participant when the decision to increase or stop is uncertain;
2. in the live scenario, the assessor can alter their observation point to obtain more information;
3. in the video situation, the assessor can pause for more time or rewind the tape if uncertain of the participant's performance.

Only five of the forty-eight trials (10%) varied for the first assessor.

Material Handling Tests

Four different material handling tests were conducted – floor to bench lift, bench to shoulder lift, bench to overhead lift and bilateral carry. Combining both trials, the average, high and low results for each are tabulated below (figure 5).

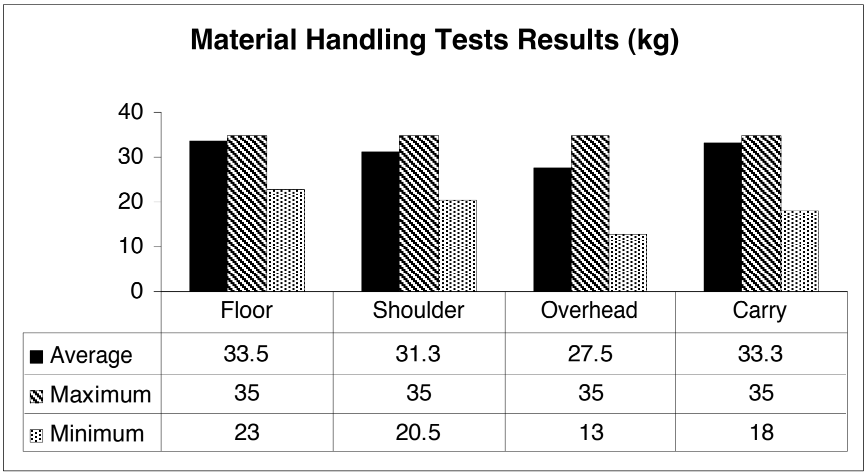

Figure 5. Material Handling Tests Results

Inter-rater ICC values ranged from 0.81 to 0.98 (good to excellent), and intra-rater ICC values ranged from 0.86 to 1 (good to excellent). Whilst the range of available scores with the material handling was larger than that of the postural tolerances and the confidence intervals overall much narrower, the use of the ICC for determining inter- and intra-rater reliability is still questionable (Table 5). The largest variation in these measures of reliability was with the bench to shoulder lifts. This could be due to the difficulty in observing the onset of compensatory movements and loss of postural control with this task in comparison to the others.

Table 5
Intra-class Correlation Coefficient (ICC) scores and Confidence Intervals for Material Handling Tests

Test	*Inter-rater [live vs. video (n=48)]*	*Inter-rater [video vs. video (n=48)]*	*Intra-rater [live vs. video (n=48)]*
Floor to bench	0.96 (0.93 – 0.98)	0.98 (0.96 – 0.99)	0.98 (0.96 – 0.99)
Bench to shoulder	0.92 (0.87 – 0.95)	0.81 (0.70 – 0.88)	0.86 (0.78 – 0.91)
Bench to overhead	0.89 (0.83 – 0.93)	0.91 (0.85 – 0.94)	0.95 (0.93 – 0.97)
Bilateral carry	0.96 (0.94 – 0.98)	0.96 (0.94 – 0.98)	1.0

Test-retest ICC values ranged from 0.56 to 0.88 (poor to good) The sample size for the test-retest (n=19 to 20) and the narrow range of results for the floor to bench and bench to shoulder lifts further weakened the value of determining the ICC for this group. These results have been included (Table 6) simply to illustrate this point. Discussion of the results in the following paragraphs will give a more accurate representation of the test-retest reliability and the implications that this would have on the participant's overall PEFA score.

Table 6
Test-retest Intra-class Correlation Coefficient (ICC) Scores and Confidence Intervals for Material Handling Tests

Test	*ICC*	*Lower limit*	*Upper limit*
Floor to bench (n=19)	0.56	0.22	0.78
Bench to shoulder (n=20)	0.64	0.34	0.81
Bench to overhead (n=20)	0.82	0.63	0.91
Bilateral carry (n=20)	0.88	0.74	0.94

Floor to Bench

Test-retest: Only nineteen floor to bench trials were included, as one participant could not comfortably squat during the first trial due to discomfort from a recent tattoo. Only four scores (21%) varied between trials. The variation is illustrated in figure 6. Two improved and two declined in performance, both due to self-limiting behaviour. That is, the worker stopped the test prematurely with complaints of lower back pain for one, and feeling 'heady with sinus' by the other. The worker with lower back pain declined in performance from 30kg to 22kg. This is a positive indicator of the validity of this assessment methodology. Both results would have lowered their overall score. In both cases, the intra-rater and inter-rater scores were 100% consistent. Conversely, the two participants that improved would have increased their score and similarly the intra-rater and inter-rater scores were in agreement.
Inter-rater: Of forty-seven trials, there were two variations in scores demonstrating excellent reliability.
Intra-rater: There was only variation in scoring. This variation was agreed upon by both assessors watching the video.

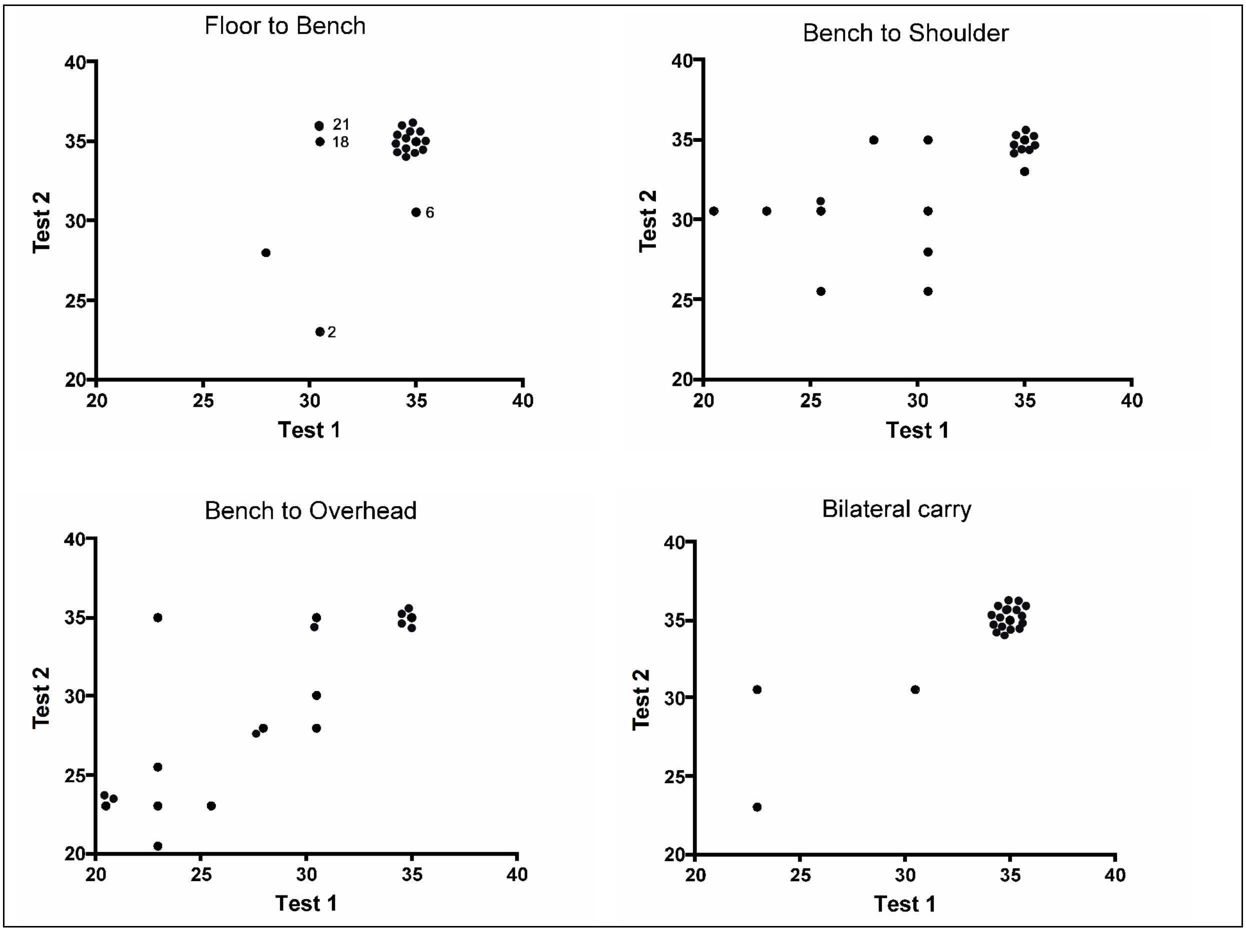

Figure 6. Test-retest Reliability of Material Handling Tests

Bench to Shoulder

Test-retest: As indicated by the confidence intervals, the variation in bench to shoulder lifts was larger. Of the twenty trials, eight (40%) varied between trials. Only three declined in their performance. One of these was self-limiting, the other two were based on the assessors' decision. The second two only declined in performance by 2kg. Two of the three would have achieved a lower overall score. The other five variations were improvements in performance, ranging from 5 to 7kg. All of these would have achieved a higher overall score. It is suspected that motivation was a major contributing factor to this change.

Inter-rater: A quarter of the 48 trials recorded variation between the assessors, with the second assessor typically more conservative.

Intra-rater: 14.5% of the trials recorded an intra-rater variation with the video score typically more conservative than the live score.

Bench to Overhead

Test-retest: Again, there was significant variation amongst the two trials for the bench to overhead lift. However, only three declined in performance with the results of only one affecting their overall score. As with seven of the ten variations, the change was only 2-2.5kg which was one increment in the progressive weight protocol. It is worth noting that one participant improved from 23kg to 35kg which would have improved their score from a three to a one. The reason for this dramatic improvement is not known however, it was noted that they improved on all aspects of their test, excluding fitness.

Inter-rater: As predicted by the confidence intervals, the variation in scores between assessors was higher (16 out of 48, 33%) for the bench to overhead lift with the live assessor again giving higher scores
Intra-rater: The intra-rater variation (7 out of 48) was the same as for the bench to shoulder lift with no identifiable trend to lower scores on video or live.

Bilateral Carry
Test-retest: Out of twenty participants, only one varied between trials. His improvement of 7kg was directly as a result of self-limiting behaviour in the first trial. That is, he stopped the test prior to the assessor determining that his safe maximal lift had been reached. The improvement would have resulted in him achieving a higher PEFA score.
Inter-rater: Of the forty-eight trials, there were three occasions where the second assessor would have scored the participant one increment lower on the bilateral carry task.
Intra-rater: No variation recorded.

Postural and Dynamic Tolerances Tests

As discussed previously, one of the limiting factors of using the ICC as a measure of reliability is that when there is only a small range in the values it loses some of its sensitivity. In these cases, such as the postural tolerances results below, reporting of individual scores and explanation of the variation from the raw data can provide more useful information. This limitation is magnified when a small sample size (n=20 for test-retest) is also used. The ICC results for the inter-rater and intra-rater reliability for the postural tolerances are tabulated below (Table 7) with more detailed explanations in the following paragraphs. No consistent trend between video vs. video and live vs. video was identified and so it can be assumed that the medium did not make a significant difference to the result in the postural tolerances tasks.

Table 7
Intra-class Correlation Coefficient (ICC) Scores
and Confidence Intervals for Postural Tolerances Tasks

Test	*Inter-rater (live vs. video)*	*Inter-rater (video vs. video)*	*Intra-rater (live vs. video)*
Reach Forward	0.87 (0.79 – 0.92)	0.93 (0.89 – 0.96)	0.93 (0.89 – 0.96)
Reach Overhead	0.86 (0.78 – 0.91)	0.75 (0.62 – 0.84)	0.60 (0.41 – 0.73)
Stoop	0.84 (0.75 – 0.90)	0.72 (0.57 – 0.82)	0.81 (0.70 – 0.88)
Squat	0.68 (0.53 – 0.80)	0.82 (0.72 – 0.89)	0.67 (0.51 – 0.78)
Climbing	1	1	1

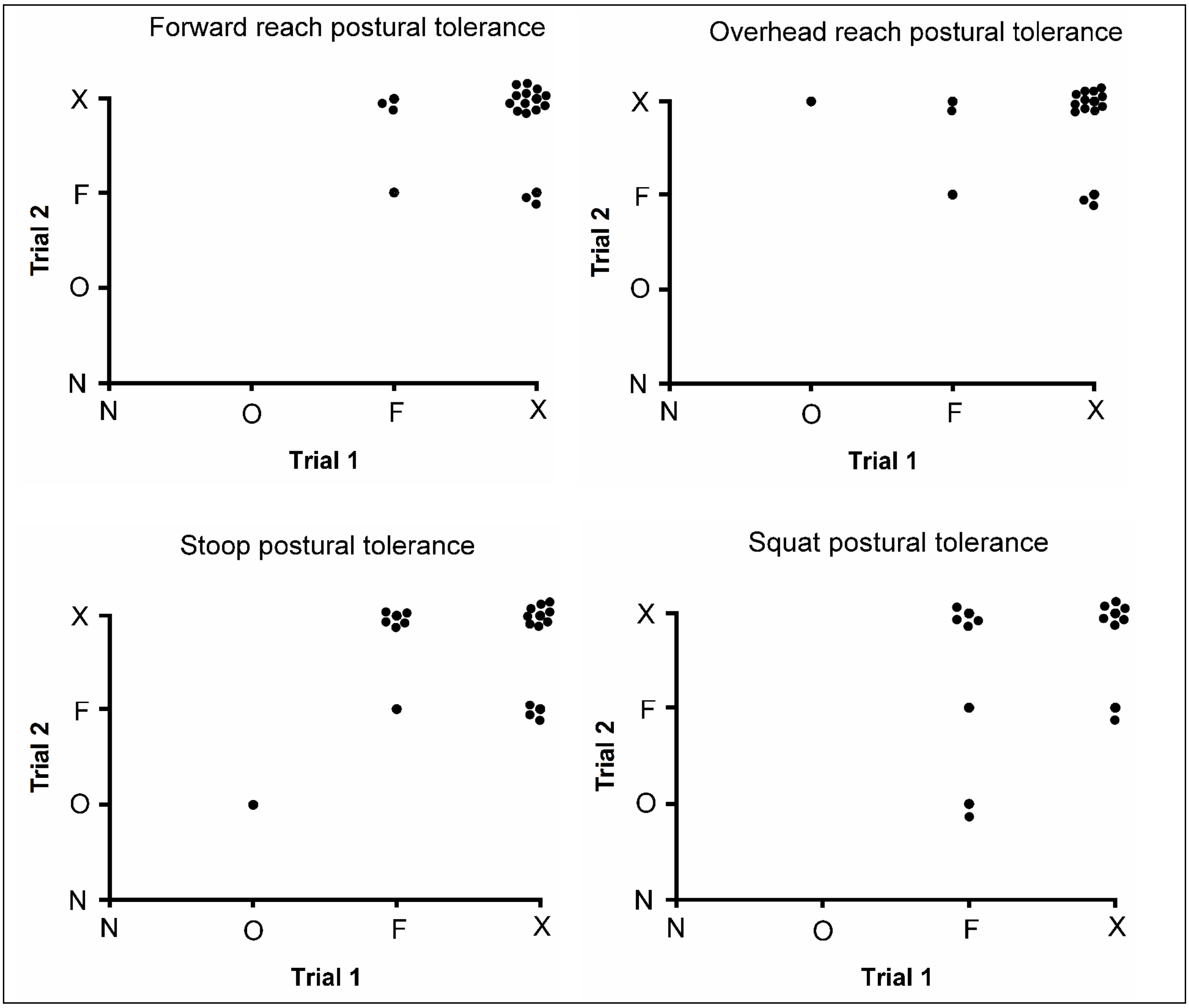

Figure 7. Test-retest Reliability of Postural Tolerances Tests

Forward Reach

Test-retest: Six of the twenty participants varied between trial one and trial two. Three improved from 'F' to 'X' and three decreased from 'X' to 'F'. Of those that decreased, two reported feeling unwell. The third's result was based solely on heart rates changes and was also scored inconsistently between the raters. These changes would not have changed their overall score.

Inter-rater: Of the forty-eight trials, there were only two variations. There was a 50/ 50 split between variation of live vs. video and video vs. video.

Intra-rater: There was only variation of the forty-eight trials which is indicative of excellent reliability.

Overhead reach

Test-retest: Again, there were six variations between trials one and two. Three also improved, this time, two from 'F' to 'X' which would not have altered their overall score, but one from 'O' to 'X' which would have increased their score. Evaluation of the raw data demonstrated this participant did not complete the task in the first trial. This variation is therefore a positive indicator toward the validity of the data. The three participants whose score reduced from 'X' to 'F' all reported arm fatigue with corresponding changes in their heart rates. The workers reported no explanation for their change in performance. These scores would not have changed their overall rating but would indicate a referral for behaviour modification such as avoiding repetitive or sustained overhead reaching.

Inter-rater: Variation in this task was double that of forward reach (8% versus 4%) but still low.
Intra-rater: There were six variations amongst the forty-eight trials (12.5%). Although this is higher than the forward reach, this still indicates good reliability despite a moderate score in the ICC value (0.60).

Stoop
Test-retest: There was a higher rate of variation for the stooping task. Half of the results varied between trials but only four worsened. Three out of the four participants reported discomfort, two from football training the night before. Changes in heart rate coincided with three of the changes. Only one had disagreement between assessors. None of the changes would have affected the participant's overall score.
Inter-rater: Variation was the same as the overhead reach task (four of the forty-eight trials)
Intra-rater: Intra-rater variation was also the same as the overhead reach task, again scoring good. The ICC value in this case however was 0.81.

Squat
Test-retest: Eight of the nineteen participants (42%) varied between trials of squatting tolerance but with only three showing a decline in performance, one of which was due to self-limiting behaviour (i.e. stopped test prematurely). The other two decreased from an 'X' to an 'F'. Evaluation of the raw data shows that this was based on heart rate change alone and these scores did not show intra- or inter-rater reliability. These results did not affect the participant's overall scores.
Inter-rater: Variation was highest in the squatting task. Six (12.5%) of the forty-eight trials varied.
Intra-rater: Intra-rater variability was also the highest at 14.5% (seven trials). These higher rates of variation could be contributed to less clear definition of compensatory behaviour. It could also indicate that heart rate changes during this task may not be as strong an indicator of discomfort or effort as large muscle groups are not being used and the task is performed lower to the ground thus decreasing the work of the heart.

Climbing
There was no variation in the climbing scores with the test-retest, inter-rater or intra-rater comparisons.

Fitness test

The results of the aerobic fitness test are illustrated in Figure 8. Two participants did not complete the test within their 85% MHR and thus rated 'poor'. Nineteen fitness test results were recorded for both trials. Ten participants scored the same result in both trials (five fair, three average and two good). Three declined in their rating and four improved. It is worth noting, that whilst the two departments scored equally on the overall PEFA score, those employed in the labour-intensive roles, on average demonstrated higher levels of aerobic fitness by an increased number with a rating of 'good' (six versus two).

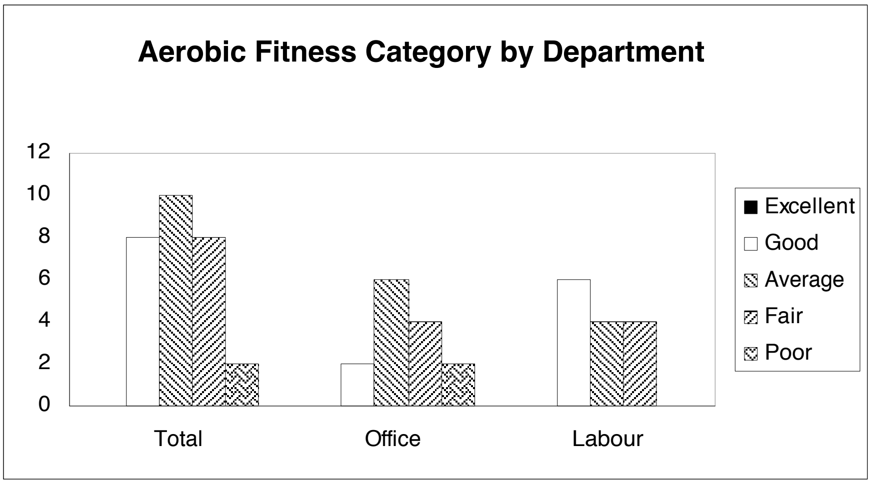

Figure 8. Aerobic Fitness Category by Department

Test-retest scores for the fitness tests are illustrated below (figure 9). Due to the variation in results between trials one and two of the fitness test, recovery heart rates were also compared in an attempt to account for the variation. No clear and consistent explanation can be offered for these results. Factors influencing heart rates include, but are not limited to: emotional state, physical fitness, prior activity, caffeine, tobacco, prescription and non-prescription drugs and fatigue. Whilst this extreme variation in fitness test results does not have any direct implications on the overall PEFA score it may influence the conclusions that can be drawn from the subsequent validity study.

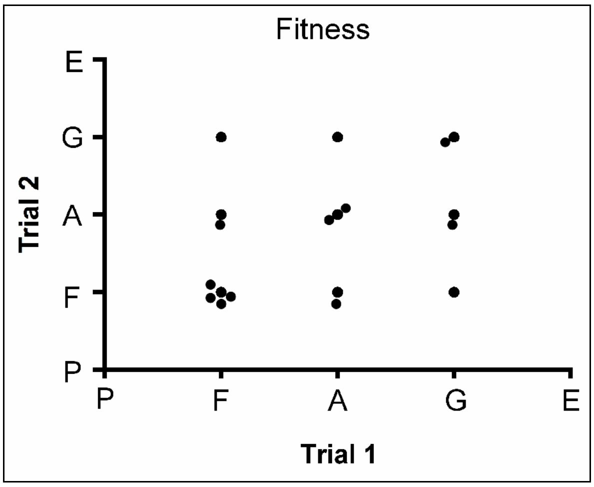

Figure 9. Test-retest Reliability for the Fitness Test

Balance Test

Nineteen balance test results were recorded. Between the two trials, twelve participants consistently scored 'unlimited'. Two consistently scored 'limited'. Five scored a 'limited' result in trial one but improved to 'unlimited' in trial two. No reason for this improvement was documented nor reported by the participants. It is reasonable to assume that there is a positive practice and motivational component to the second trial results in these five participants.

Discussion

Reliability encompasses test-retest, intra- and inter-rater reliability. Reliability of a measure needs to be determined prior to addressing the validity of a test. In consideration of the ICC values, confidence intervals and raw data, the reliability ratings for each test are tabulated below (Table 8).

Table 8
Reliability Ratings for PEFA Score and all Tests

Test	*Test-retest*	*Inter-rater*	*Intra-rater*
PEFA score	Good	Good	Excellent
Floor to bench lift	Moderate	Excellent	Excellent
Bench to shoulder lift	Moderate	Good	Good
Bench to overhead lift	Good	Good	Excellent
Bilateral carry	Good	Excellent	Excellent
Reach forward	Moderate	Good	Good
Reach overhead	Moderate	Good	Moderate
Stooping	Poor to moderate	Good	Good
Squatting	Poor to moderate	Moderate	Moderate
Climbing	Excellent	Excellent	Excellent
Fitness	Poor	NT	NT
Balance	Moderate	NT	NT

As discussed previously, the ICC as a measure of reliability is not necessarily sensitive enough to account for the small ranges of values used in the components of this test. Therapists when interpreting these results for clinical use would be better informed by taking note of the actual values and reason for change between them rather than looking at the ICC alone.

Despite the variation in some of the scores, it was only a small number of cases where the changes would have affected the participant's overall score (six negatively, eight positively). The overall score, is not meant to pass or fail potential job candidates but rather give the worker and the employer an indication of the level of risk of injury to that worker performing that role at that time. The individual test results are designed to offer both parties useful information on how the job can be modified or appropriate steps that the worker can take to minimise their risk of injury from manual handling injuries at work.

Conclusion

The individual test components of the JobFit System Pre-employment Functional Assessment demonstrated sufficient reliability to be used for comparison with injury data for determining the validity of the assessment.

More published research is required on test-retest reliability of all components, particularly postural tolerances. Once the validity of the postural tolerance components and the impact of their results on workplace injuries has been identified, the need for additional reliability data on these measures may vary.

When tests with small data ranges, particularly if coupled with small sample sizes, are used it is difficult to accurately measure reliability using conventional statistical methods.

The overall PEFA score, climbing task and all four material handling tasks (floor to bench lift, bench to shoulder lift, bench to overhead lift and bilateral carry) demonstrated sufficient reliability for their inclusion in the subsequent validity study. The remaining tasks were included but results have been interpreted with caution and weighted according to the reliability study findings.

Primary (Validity) Study

Introduction

The validity of a measurement refers to the extent to which it measures what it is intended to measure. In this case the JobFit PEFA aims to provide a measure of the likelihood of a person sustaining a musculoskeletal injury performing the duties of the job against which the person's physical capabilities were assessed. The primary objective of this study was to determine whether the PEFA score was predictive of subsequent musculoskeletal injury rates and thus also predictive of performance.

Method

Subjects

Xstrata Newlands Coal Mine agreed to participate in the study. The mine employs around 500 workers and 900 contractors. They have surface and underground operations with CHPP facilities. Pre-employment functional assessments were conducted on male and female applicants from all operations as well as office and professional roles as part of the usual recruitment process of the operation. Xstrata Newlands Coal have been using the JobFit System software and assessment method since December 2002.

Demographic data was collected. Before testing, each participant was required to sign a written consent form outlining: (i) the components of the assessment (ii) the risks and expectations of submaximal physical testing and the precautions that would be taken (iii) the purpose of the assessment and the use and disclosure of the collected information(iv) the opportunity to discontinue testing at any time. The consent form was designed to meet relevant medico legal and privacy law requirements. Information was collected, de-identified, used and disclosed in accordance with the National Privacy Principles. The study was approved by the Ethics Officer of the School of Human Movement Studies, University of Queensland. Participants were screened for exclusion factors prior to commencement of the assessment. Exclusion factors included current injury, significant injury or surgery in the last six months, elevated blood pressure (resting systolic >160mmHg or resting diastolic >95mmHg) or specific medical advice.

Experimental Design

Applicants participated in PEFAs specific to the job for which they were applying. Job demands were retrieved from the JobFit System task database. The task database contains task descriptions and itemised functional demands for individual tasks. This data had been collected on a previous occasion by a physiotherapist. The functional demands for all the tasks for a job were collated and summarised. The high risk components of those jobs were recorded and included in the job-specific PEFAs.

When an applicant was applying for more than one job, then the assessment for the job that had the highest functional demands was used.

Each PEFA contained the following essential components:

- musculoskeletal screen (posture, mobility, strength)
- balance test
- aerobic fitness test

Each PEFA contained at least one of the following job-specific components:

- reaching forward
- reaching overhead

- stooping
- squatting
- safe maximal lifts at floor, bench, shoulder and overhead heights
- bilateral carry

A breakdown of the most common job-specific assessment components at the time of the study are tabulated below (Table 9)*:*

Table 9
Job-specific PEFA components

Job	*Reach Fwd*	*Reach O'head*	*Stoop*	*Squat*	*Climb*	*Lift floor to bench*	*Lift bench to sh'lder*	*Lift bench to o'head*	*Bilat. carry*
Plant Op.	F	F*				35kg*			
Dragline Op.	F		F	F	F	35kg			11kg
OC Fitter	F	F	F	F	F	35kg	35kg	25kg	35kg
OC B'maker	F	F	F	F	F	35kg	35kg	35kg	35kg
OC Elec.	F	F	F	F	F	35kg	35kg	25kg	25kg
Office / Prof.	F					8kg	8kg	8kg	8kg
UG Op.	F	F	F	F		35kg	35kg	35kg	35kg
UG Elec.	F	F	F	F		35kg	35kg	35kg	35kg
CHPP Op.	F	F	F	F	F	35kg	35kg	35kg	35kg

* Job criteria have changed in recent months

All PEFAs were conducted in the same order:

1. PEFA consent signed
2. Demographic data collected
3. Medical history questionnaire completed
4. Musculoskeletal screen
5. Aerobic fitness test
6. Balance test
7. Reaching forward task
8. Reaching overhead task
9. Stooping task
10. Squatting task
11. Floor to bench lift
12. Bench to shoulder lift
13. Bench to overhead lift
14. Bilateral carry

At the conclusion of the assessment, the worker was given feedback on their performance so that they were more aware of their possible injury risks. They were also presented with a list of recommendations for voluntary injury prevention programs such as strengthening and flexibility.

The worker's results were compared to the job demands and they were given an overall PEFA score between 1 and 4:

- *Score 1* - Has demonstrated the functional capacity to perform the proposed position as described with no restrictions.
- *Score 2* - Has demonstrated the functional capacity to perform the proposed position as described with minimal restrictions.
- *Score 3* - Has demonstrated the functional capacity to perform the proposed position as described with moderate restrictions
- *Score 4* - Has not demonstrated the functional capacity to meet the inherent requirements of the proposed position as described.

At the commencement of the study, a scoring system of 1,2,3 was used. During the study, it was determined that additional detail was required in the scoring system. Score 1 remained unchanged. Score 2 was divided into Score 2 and 3. Score 3 became Score 4. Records using the initial scoring system were re-scored to reflect the revised version.

All PEFA records were collected at the time of assessment. Only records for employed subjects and only those that had completed an assessment for the position in which they were employed were used for the study. The PEFA results were not used in the hiring decision-making process.

Injury statistics were collected retrospectively at the end of the initial period (December 2002 to December 2006) and subsequent reporting period (January 2006 to December 2006). Injury statistics supplied by the Health & Safety team at Xstrata Newlands Coal included date of injury, type of injury, body location, cause of injury and task performed at time of injury. Only sprains and strains data was used.

Other data collection provided by the Xstrata Human Resources Department included start and finish dates to measure length of employment and to ensure that the record was applicable to the study period. Workers who were employed for less than ninety days at the time of the analysis were not included in the analysis.

Unfortunately, Xstrata Newlands Coal did not have direct access to the cost or duration of individual claims. Citing privacy regulations, the insurer would not release the required information. Attempts were made to contact the workers directly to ask for a written consent to release the information but with insufficient response rates to be included in the analysis.

Data Analysis

Demographic data was summarised. Overall PEFA scores were compared to worker age, department and length of employment. Overall PEFA scores and individual test components were compared to the following sprains and strains injury data:

- injury rates
- mechanism of injury
- body location of injury

Results

Subjects

Four hundred and ninety PEFA records were available from assessments undertaken between December 2002 and January 2007 of which 385 workers were employed. Only 368 were employed in the position for which they applied. Retaining only those employed for more than

90 days to end December 2006, 336 records were available for analysis with 243 from the surface operations (open cut, workshop and CHPP), 53 from the underground and 40 in office / professional roles. The average employment duration was 797 days (2.18 years).

Age

The average age of the subjects at the start of employment was 39.3 years and the oldest 63 years.

Age and Length of Tenure

For those 112 who ended their employment before 31 December 2006, there was no association between age at the start of employment and duration of employment (p=0.706) (Figure 10).

Similarly, the difference between the average age of the employees who left employment before 31 December 2006 (mean 38.0 years„ median= 37.2 years) and the average of those still employed at 31 December 2006 (mean 39.7 years, median=38.9 years) was not significant (Mann-Whitney U=11800, p=0.187) (Figure 11).

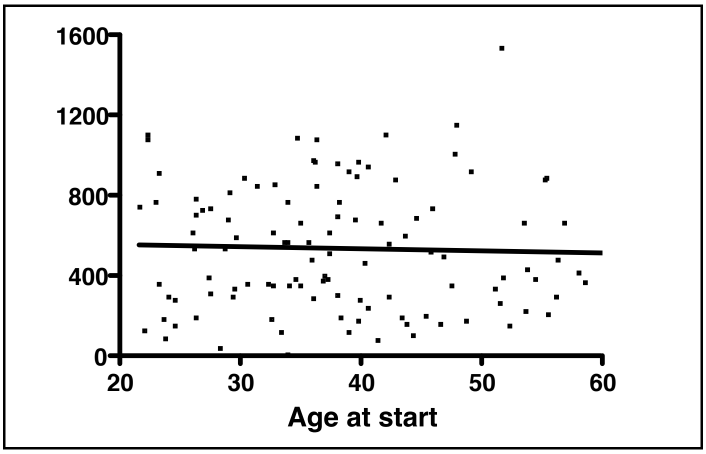

Figure 10. Duration of Employment by Age

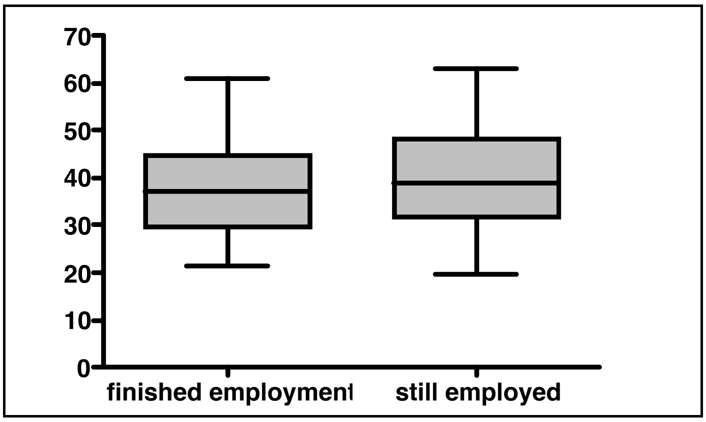

Figure 11. Average Age of Workers by Employment Status

Age and Department

Age groups by department are illustrated in Figure 12. The average age for workers employed in the workshop was 36.7 years, open cut operations (inc CHPP) was 41.4 years, underground operations 35.3 years and office / professional roles 40.8 years.

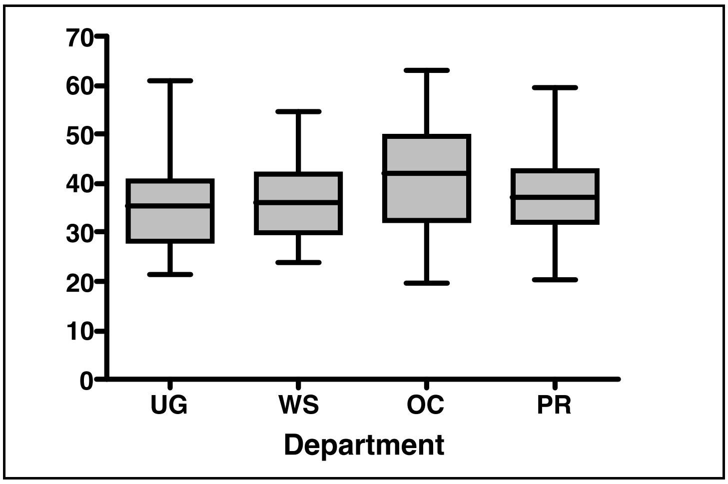

Figure 12. Average Age by Department

A significant difference exists for age at the start of employment as a function of department (Kruskal-Wallis statistic= 18.55, p<0.001), however Dunn's multiple comparison test determined that the difference is restricted to between UG and OC departments (p<0.01). All other comparisons are not significant. This analysis indicates that jobs in higher risk departments (UG and WS) are more likely to be occupied by younger employees.

Age and PEFA score

Age groups by PEFA score are illustrated in Figure 13a. The average age for PEFA score 1 was 38.6 years, score 2 was 41.6 years and score 3 was 41 years.

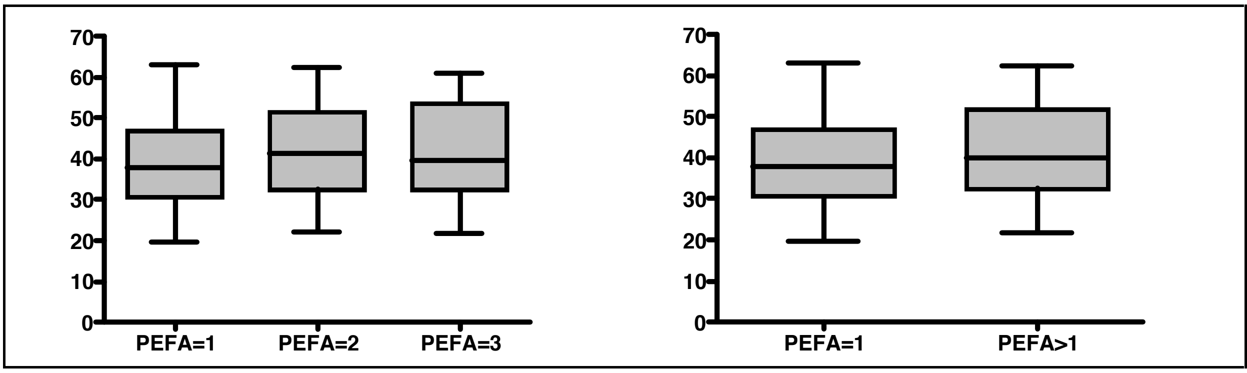

Figure 13. Average Age by PEFA Score

A Kruskal-Wallis test indicated no significant difference in age across the PEFA groups (Kruskal-Wallace = 4.053, p=0.132). However, if PEFA scores of 2 and 3 are combined (Figure 13b), a significant difference is evident (Mann-Whitney U = 8881, p=0.0451). Workers who scored 1 were generally younger (median 37.8 years) than those that did not (median 40.0 years).

Age and Injury Reports

The mean age at start of employment for those who reported an injury was 38.6 years (median= 37.7), while the mean age for those who did not report an injury was 39.4 years (median= 38.1). This difference is not significant (Man-Whitney U=8153, p=0.62), indicating no overall association between age and likelihood of injury (Figure 14).

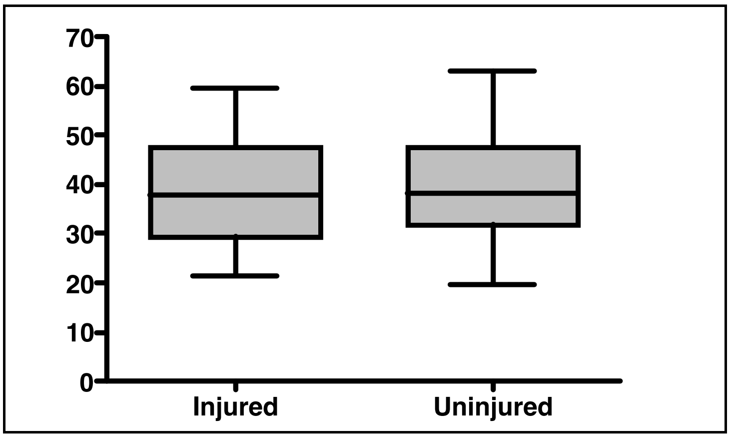

Figure 14. Average Age of Injured and Uninjured Workers

Sex

Of the 490 records available, twenty-eight belonged to female participants. Of these, twenty-two were suitable for analysis. At the start of employment, the average female age was 31.2 years (min 19.7 years, max 47.6 years). Half were employed as Plant Operators in the Surface Operations and the remainder were employed in office / professional roles. None were employed in the workshop or underground operations. Two of the eleven plant operators had received a single back injury during the study. The numbers in this group are too small for a segregated detailed statistical analysis and have been included in the overall subject groups.

PEFA score

Of the 336 suitable subjects, 254 were allocated a PEFA score of 1, fifty-three PEFA score 2 and twenty-nine PEFA score 3. For analysis, these were combined into PEFA=1 (n=254) and PEFA>1 (n=82) as PEFA=2 and PEFA=3 numbers were insufficient to analyse independently.

PEFA score and length of tenure

The average duration of employment was 805 days (2.21 years) for PEFA 1, and 774 days (2.12 years) for PEFA > 1.

Retention Rates

During the trial period, 113 (29%) of the participants that were employed left employment. Of those, ninety-one were PEFA 1 and twenty-two were PEFA>1. The average duration of employment was 562 days (1.54 years) for PEFA=1 and 399 days (1.09 years) for PEFA>1. This difference is significant (Mann-Whitney U=688.5, $p=0.0237$) suggesting that turnover of PEFA=1 employees was less than PEFA>1 over the period studied (Figure 15).

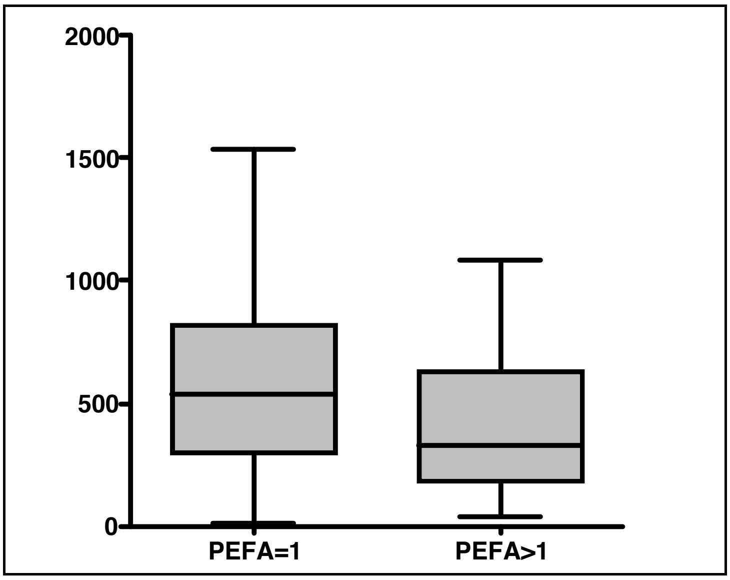

Figure 15. Employment Duration by PEFA Score

PEFA score and sprain and strain injury rates

Ninety-seven injury records were available for sprain and strain injuries reported by sixty-seven employees. Of these, six injuries were sustained by four people who were not tested for the correct job role and one who was employed for less than ninety days and therefore these records could not be used. This left ninety-one injuries to sixty-two employees for whom a PEFA score was available.

Thirty-eight employees reported a single injury, seventeen reported two injuries, five reported three injuries and one person sustained four injuries during the study period.

Injury rates and exposure

The average duration of employment for those not reporting an injury was 743 days (2.04 years), compared to 1040 days (2.85 years) for those reporting an injury. This difference is statistically significant (Mann-Whitney U=5470 $p<0.0001$). Not surprisingly, the overall probability of sustaining an injury is significantly related to exposure.

Injury rates for body location and cause of injury

Injury records were coded by body location and cause of injury. Injury rate per person year was calculated using the following formula:

Injuries / # Employees / Average duration of employment = Injuries per person year

Injury rates per person year by body location and cause are tabulated below (Table 10).

Table 10
Sprain and strain injuries reported by all employees by body location and cause

	Slip/ Trip/Fall	*Access/ Egress*	*Manual Handling*	*Rough Road*	*Other*	*Subtotal*	*Total Injury Rate per person year by Body Location*
Hand/Wrist			5	1	2	8	0.011
Head/Neck				4	6	10	0.014
Shoulder		5	10		1	16	0.022
Back/Trunk		2	18	5	7	32	0.044
Lower Limb	9	10	3	1		23	0.030
Multiple/ Unknown	1				1	2	0.002
Subtotal	10	17	36	11	17	91	
Total injury rate per person year by cause	0.014	0.023	0.049	0.015	0.023	0.12	

Manual handling had the highest rate of strain and sprain injuries per person year (0.049) which was more than double that for injuries associated with accessing and egressing equipment (0.023) and more than triple the rate of injuries associated with slip, trips and falls (0.014) or travelling on rough roads (0.015). These ratios are displayed in Figure 16 below.

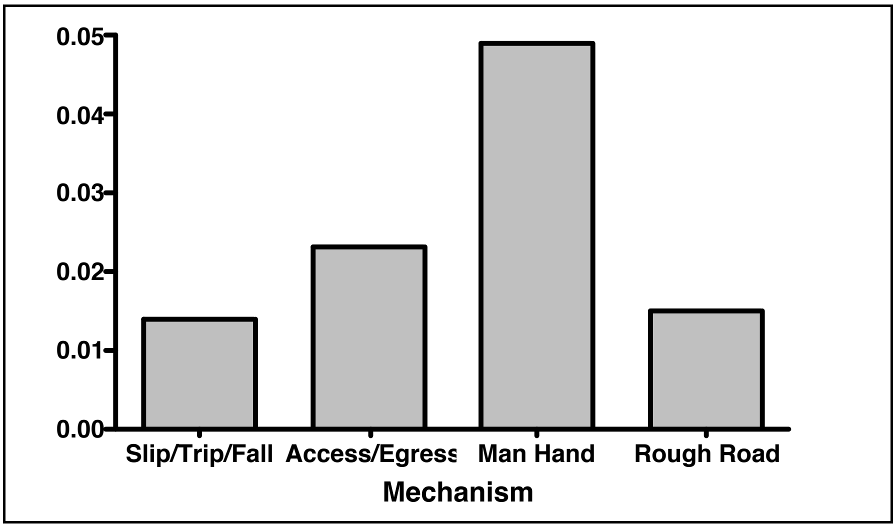

Figure 16. Injury Rate by Mechanism of Injury

Figure 17 illustrates the breakdown of injuries by body location. On their own, back and trunk injuries had the highest incidence (0.044), followed lower limb injuries (0.03) and shoulder injuries (0.022). Back, trunk and shoulder injuries were the most common body locations affected by manual handling injuries. The combined back, trunk and shoulder injury rate was 0.066 injuries per person year. When looking at only those associated with manual handling activities, the injury rate is 0.038 manual handling related sprain and strain injuries to the shoulder, back or trunk each person year.

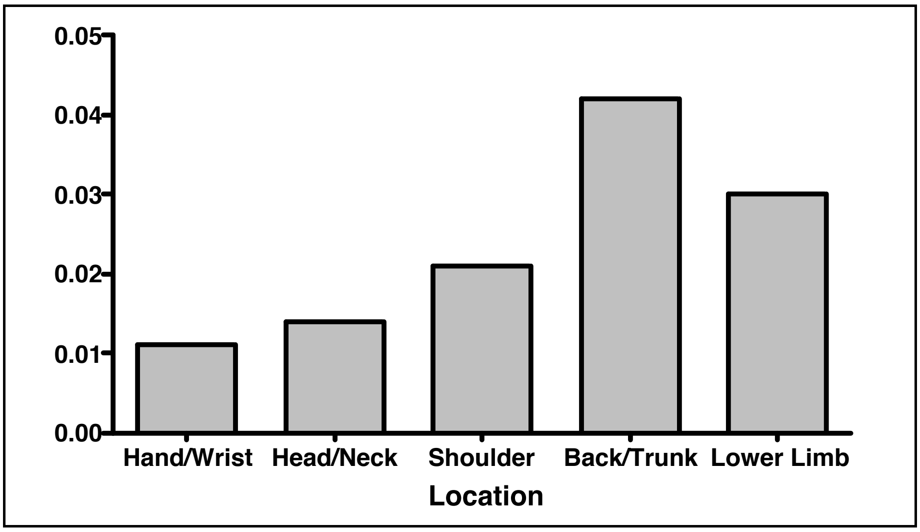

Figure 17. Injury Rates by Body Location

Injury rates for body location and cause of injury by PEFA score
In the PEFA score 1 group (n=254), fifty-four injuries were reported by thirty-nine workers. Using the same formula as above, the total injury rate was 0.096 injuries per person year. Again, manual handling was the most common cause of injury and injuries to the back / trunk had the highest rate of injury in this group (Table 11).

Table 11
Sprain and strain injuries reported by employees with PEFA 1 by body location and cause

	Slip/Trip/ Fall	*Access/ Egress*	*Manual Handling*	*Rough Road*	*Other*	*Subtotal*	*Total Injury Rate by Body Location*
Hand/Wrist			4		1	5	0.009
Head/Neck				4	4	8	0.014
Shoulder		3	3		1	7	0.012
Back/Trunk		1	10	4	5	20	0.036
Lower Limb	4	7	2			13	0.023
Multiple/ Unknown					1	1	0.001
Subtotal	4	11	19	8	12	54	
Total injury rate by cause	0.007	0.02	0.034	0.014	0.021	0.096	

Back / trunk and shoulder injuries were combined resulting in an injury rate of 0.048 injuries per person year and the total back and shoulder injury rate with manual handling as the cause was 0.023 per person year.

In the PEFA score >1 group (n=82), thirty-seven injuries were reported by twenty-three workers with a total injury rate of 0.213 injuries per person year compared to 0.096 in the PEFA score 1 group (Table 12). Again, manual handling was the most common cause of injury and injuries to the back / trunk had the highest rate of injury in this group although the rates were much higher (0.098 compared to 0.034 and 0.069 compared to 0.036 respectively).

Table 12
Injuries reported by employees with PEFA > 1
by body location and cause

	Slip/Trip/ Fall	*Access/ Egress*	*Manual Handling*	*Rough Road*	*Other*	*Subtotal*	*Total Injury Rate by Body Location*
Hand/Wrist			1	1	1	3	0.017
Head/Neck					2	2	0.012
Shoulder		2	7			9	0.052
Back/Trunk		1	8	1	2	12	0.069
Lower Limb	5	3	1	1		10	0.058
Multiple/ Unknown	1					1	
Subtotal	6	6	17	3	5	37	
Total injury rate by cause	0.035	0.035	0.098	0.017	0.029	0.213	

As for the PEFA score 1 group, back / trunk and shoulder injuries were combined this time resulting in an injury rate of 0.12 compared to 0.048 injuries per person year. The total back and shoulder injury rate with manual handling as the cause was measured at 0.086 which was 3.7 times the rate of the PEFA score 1 group of 0.023 per person year.

The differences in injury rate between PEFA score 1 group and PEFA score >1 group by body location and cause of injury are illustrated below (Figures 18 and 19).

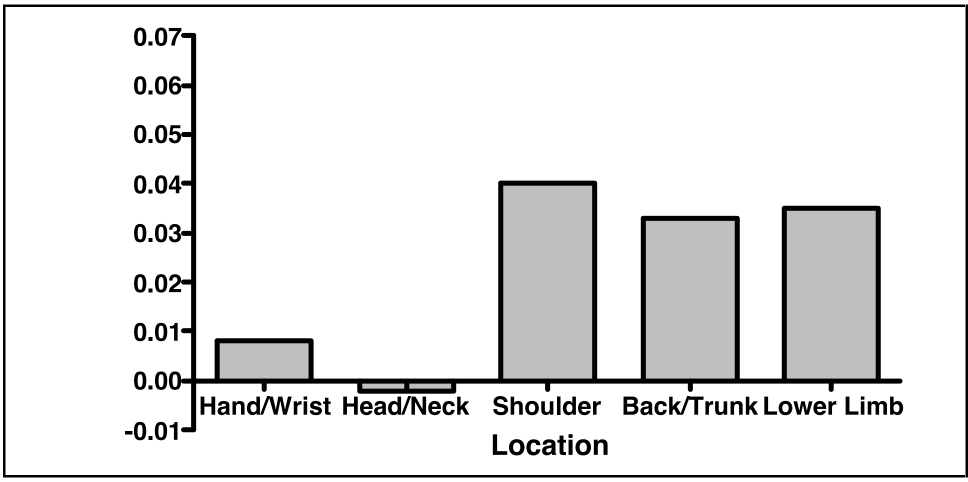

Figure 18. Differences in Injury Rate between PEFA Scores by Body Location

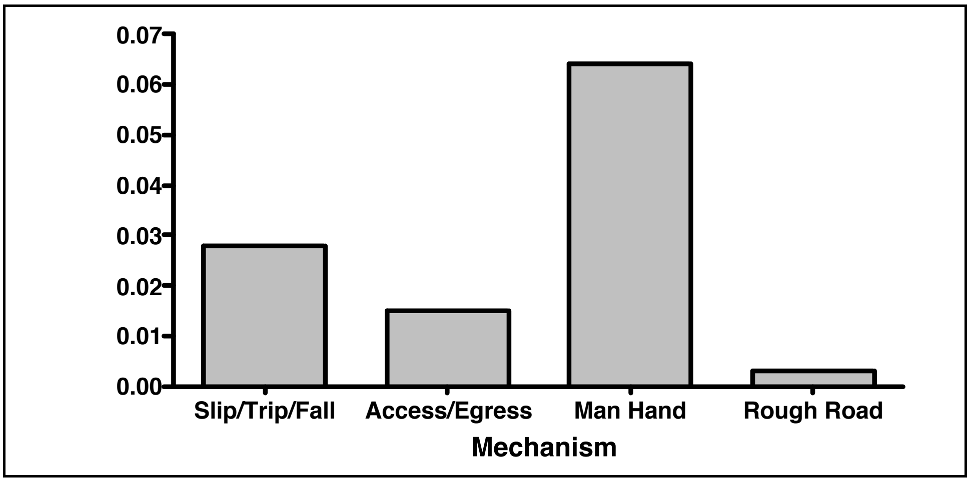

Figure 19. Differences in Injury Rate between PEFA Scores by Mechanism of Injury

Relative injury rates by PEFA score

When examining the relative risk of any injury for PEFA score >1 workers compared to PEFA score 1 workers, the first group has more than double (2.12) the injury rate. Slip, trips and falls had the largest difference (5.0) with manual handling rating second at almost three times the injury rate (2.9). Access and egress was 1.75 and injury rate from driving on rough roads was 1.21. These ratios are clearly displayed below in Figure 20.

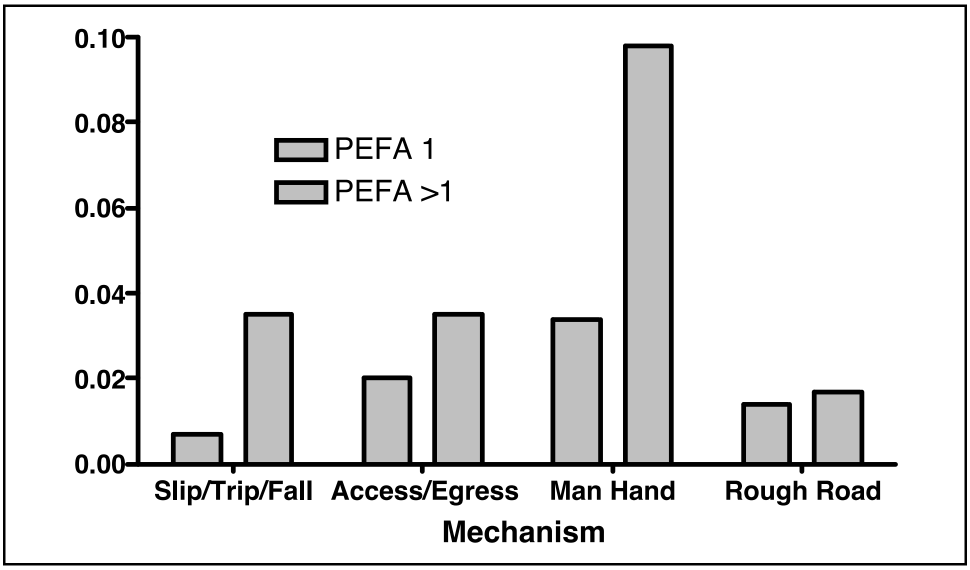

Figure 20. Relative Injury Rates between PEFA Groups by Mechanism of Injury

By body location of injury, the difference is still visible although not quite as dramatic (Figure 21). The relative risk of injury for PEFA score >1 workers compared to PEFA score 1 workers was again overall higher when investigating by body part with the exception of head and neck injuries. Workers who scored PEFA >1 were four times more likely to have a shoulder injury (4.33), 2.52 times more likely to have a lower limb injury and almost twice as likely (1.92) to have a back / trunk injury.

If we combine the back and shoulder injuries, the relative injury rate is 2.5 times higher for the PEFA>1 group. However, if we only look at those back and shoulder injuries associated with manual handling, the rate of injury in the PEFA score >1 group is 3.75. times the rate of shoulder, back & trunk injuries in the PEFA=1 group.

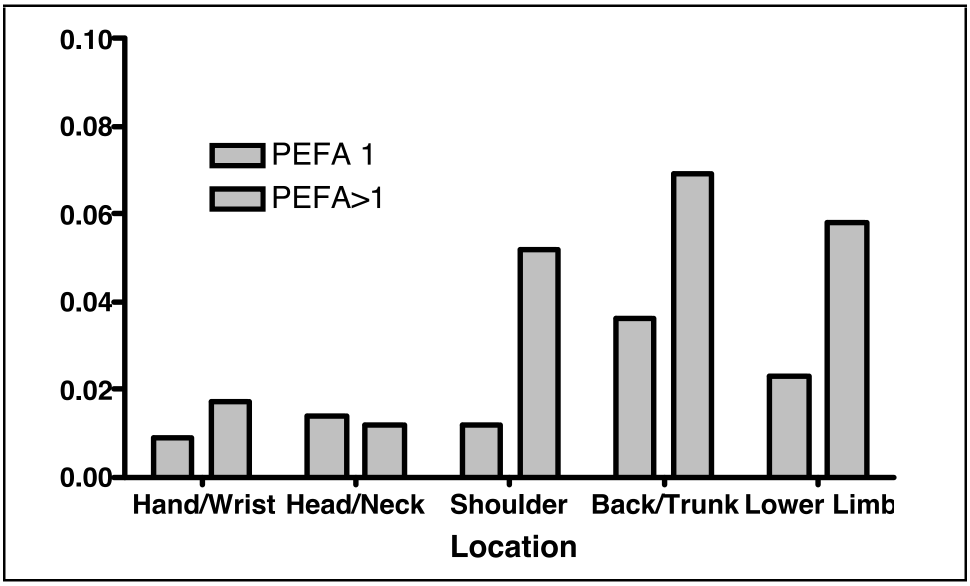

Figure 21. Relative Injury Rates between PEFA Groups by Body Location of Injury

PEFA scores and Relative Risk with 95% confidence intervals

Overall injury Relative Risk

Thirty-nine of the 254 PEFA score 1 employees, reported at least one injury over the course of employment (avg 2.21 years). This equates to 15% or 6.8% per year. Of the eighty-two PEFA score >1 workers, twenty-three reported at least one injury over their average of 2.12 years (28% or 13.2% per year). The difference between the two PEFA groups in the frequency due to this mechanism is 6.4% per year. The relative risk of a person rated a PEFA score >1 reporting at least one injury over the average 2.18 years is calculated as 1.83 with a 95% confidence interval of 1.16 to 2.87. This provides evidence of an association between PEFA score and the likelihood of reporting at least one injury.

Multiple injury Relative Risk

Of the 254 PEFA score 1 employees, eleven reported more than one injury over an average of 2.21 years of employment (4.3% or 1.9% per year). Nine workers from the PEFA score >1 group (n=82) reported more than one injury over an average of 2.12 years employment (11% or 5.2% per year). The difference between PEFA groups in the frequency due to this mechanism is 3.3% per year. The relative risk of a person rated a PEFA score >1 reporting more than one injury over the average 2.18 years is calculated as 2.53 (95%CI =1.09 to 5.9). This provides evidence of an association between PEFA status and the likelihood of reporting multiple injuries.

Mechanism of injury Relative Risks

Table 13
Comparison of Injury Risk between PEFA score and Mechanism of Injury

Mechanism	*Injury Frequency per Year*		*Difference (per year)*	*Relative Risk (over 2.18 years avg)*	*95% CI*
	PEFA 1 (n=254, 2.21yrs)	*PEFA >1 (n=82; 2.12yrs)*			
Slip/Trip/Fall	0.7%	2.8%	2.7%	3.87	1.06 – 14.1
Access / Egress	1.9%	3.4%	1.5%	1.69	0.64 – 4.26
Manual Handling	2.5%	7.1%	4.6%	2.66	1.28 – 5.51
Rough Road	1.4%	1.7%	0.3%	1.16	0.32 – 4.28

Table 13 above summarises the relative risk of injury between PEFA score groups for various mechanisms of injury. As can be seen from the 95% confidence intervals, there is a statistically significant relationship between PEFA score and manual handling injuries. Workers with a PEFA score > 1 are 2.66 times more likely to sustain at least one injury from manual handling tasks.

There was no statistically significant relationship between PEFA score and relative risk of injury from access / egress incidents nor rough road incidents. Although the confidence interval for slip / trip and fall injuries does not include a one, the range is too large to draw any firm conclusions about the relationship between PEFA score and injuries from slip / trip / fall incidents. The relative risks and 95% CI are illustrated in Figure 22 below.

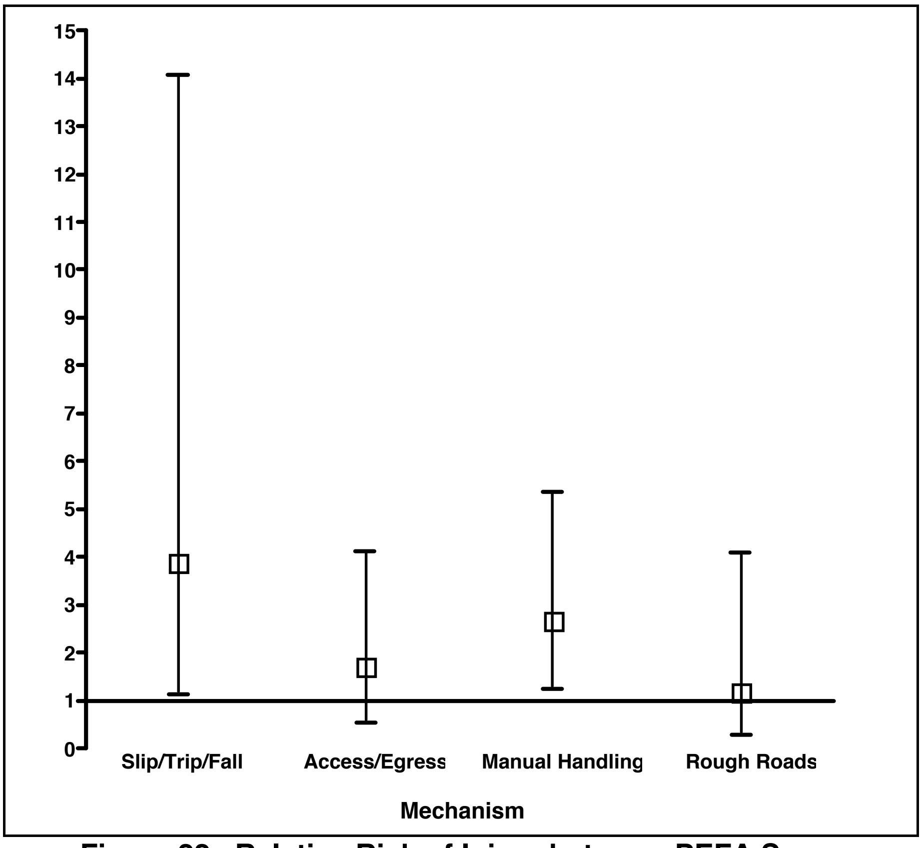

Figure 22. Relative Risk of Injury between PEFA Score and Mechanism of Injury

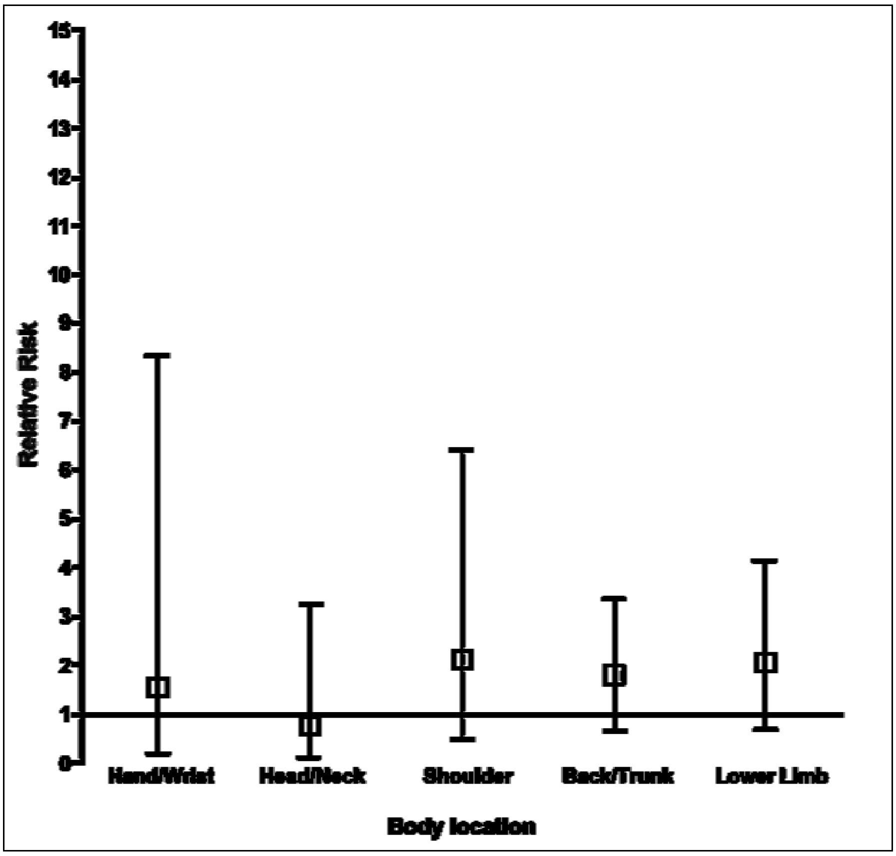

Figure 23. Relative Risk of Injury between PEFA Score and Location of Injury

Body Location Relative Risks

Table 14
Comparison of Injury Risk
between PEFA score and Body Location

	Injury Frequency per Year				
Body Location	*PEFA 1 (n=254, 2.21yrs)*	*PEFA >1 (n=82; 2.12yrs)*	*Difference*	*Relative Risk*	*95% CI*
Hand / Wrist / Elbow	0.7%	1.1%	0.4%	1.55	0.29 – 8.3
Head / Neck	1.4%	1.1%	- 0.3%	0.77	0.17 – 3.57
Shoulder	1.3%	2.9%	1.6%	2.21	0.72 – 6.78
Back / Trunk	3.0%	5.8%	2.8%	1.82	0.87 – 3.8
Lower Limbs	2.3%	5.1%	2.8%	2.06	0.91 – 4.64

As can be seen in Table 14 and Figure 23 above if we consider each body location separately, all confidence intervals regardless of body location include a 1. This means that there does not appear to be a significant relationship between PEFA score and a single upper limb, head / neck, shoulder, back / trunk or lower limb injury.

Combined Relative Risks for Shoulder, back or trunk injuries +/- associated with manual tasks
Twenty-four of the 254 PEFA=1 employees reported at least one shoulder, back or trunk injury (9.4% or 4.3 %/ year) and fifteen of the eighty-two PEFA>1 employees reported at least one shoulder, back or trunk injury (18% or 8.5%/ year). The difference between PEFA groups in the frequency of shoulder, back or trunk injuries is 4.2% per year. The relative risk of reporting at least one back, shoulder or trunk injury is 1.94 (1.07 to 3.51).

Nine of the 254 PEFA=1 employees reported at least one shoulder, back or trunk injury associated with performing a manual handling task (3.7% or 1.7%/ year). Ten of the eighty-two PEFA>1 employees reported at least one shoulder, back or trunk injury associated with a manual handling task (12.2% or 5.8%/ year) which is a difference of 4.1% per year. The relative risk of a single back, trunk or shoulder injury related to a manual handling task is 3.56 (1.5 to 8.47). The relative risks of injury associated with these factors are illustrated in Figure 24 below.

These results could be under-estimated as the bias in difference in average days (1103 and 993) worked between the injured worker groups (PEFA=1 and PEFA>1 respectively) acts to reduce effect

These results suggest that a statistically significant relationship exists between PEFA score and the risk of reporting an injury related to manual handling. This is particularly true for a shoulder, back or trunk injury associated with a manual handling task.

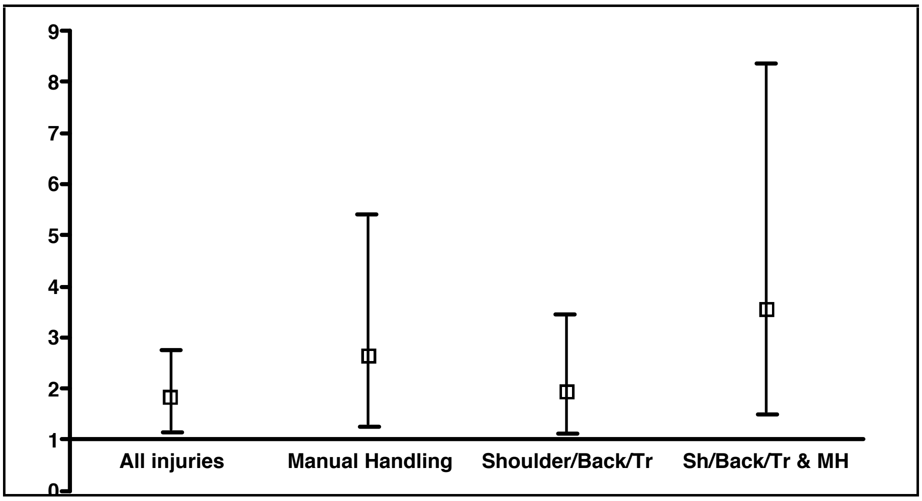

Figure 24. Relative Risks for Shoulder, Back and Trunk Injuries Associated with Manual Handling

PEFA score and Surface vs Underground workers
The breakdown of PEFA score by department is illustrated below (Table 15). A relationship exists between the department and likelihood of PEFA scores greater than 1. 79% of open cut employees were considered PEFA =1, while this was only true for 43% of underground employees (RR=2.69, 95%CI 1.92 to 3.79). This is likely to reflect the different physical requirements of the tasks involved in each area as the job demands for the underground workers are on average higher and thus a lower score is more likely.

Table 15
PEFA Score and Numbers by Department

Department	*PEFA = 1*	*PEFA = 2*	*PEFA = 3*	*Total*
OC	192	34	17	243
UG	23	18	12	53
PR	39	1	0	40

Regardless of PEFA score, the risks of musculoskeletal injury are elevated for underground workers compared with open cut workers (excluding professional staff in both areas). The relative risk of a single back, shoulder or trunk injury for underground employees was 1.8 (95%CI 0.96 to 3.39), the relative risk of a manual handling injury was 2.69 (1.3 to 5.55), and the relative risk of a back, shoulder or trunk injury associated with a manual handling task was 3.17 (1.43 to 7.04).

The overall relationship found between PEFA score and injury risk however was reflected in the experience of surface workers in that the relative risk of a single back, shoulder or trunk injury for the 243 Open Cut workers was 1.88 (0.9 to 3.98), the relative risk of a manual handling injury was 2.63 (1.05 to 6.58) and the relative risk of a single back, shoulder or trunk injury related to a manual handling task was 3.22 (1.13 to 9.18). This relationship was not found in the data gathered from the fifty-three underground workers (RR of 0.92 [0.32 to 2.64], 0.76 [0.25 to 2.33], and 0.61 [0.18 to 2.03] respectively). While the confidence intervals do not allow any strong conclusion to be drawn, it may be that the task demands of the underground environment require further examination.

PEFA Components

In addition to identifying a relationship between overall PEFA score and risk of injury, there are insufficient data to explore the contributions made by the various components of the PEFA on their own or in combination to the elevated risk associated with scores greater than PEFA=1.

Material Handling Tests

The job-specific nature of the PEFAs potentially placed artificial limits on the material handling data collected during the study and as such the raw data alone has limited significance in predicting injury risk associated with lifting capacity alone. For example, if the PEFA set a limit of 30kg for an overhead lift and in the course of employment, the worker was exposed to a lift of 35kg and they are injured, it cannot be determined from this data set alone if that worker was working above their capacity as their full capacity may not have been assessed.

Material handling capacity was the major determining factor in the PEFA score however and as a result its impact has been analysed to a degree in the earlier analysis.

Postural Tolerances and Other Tests

Various comparisons were made between postural tolerances tests and injury types. Even though the percentages in one group were often higher than the other, there were no statistically significant relationships identified. The comparisons are summarised in Tables 16 and 17 below. While the relative risk of a lower limb injury approaches twice the rate (RR=1.9), the 95% confidence interval (0.68 to 5.3) prevents any conclusion being drawn.

Table 16
Comparison of Injury Risk
by PEFA score, injury location and postural tolerances

	Percentage of Injured Group members			
Comparison of Postural Tolerance and injury location	*PEFA 1*	*PEFA >1*	*Relative Risk*	*95% CI*
Reach forward or overhead deficit vs shoulder / neck injury	5%	8%	1.4	0.34 – 5.74
Stoop or reach forward deficit vs back injury	8%	11%	1.4	0.5 – 3.8
Squat or balance deficit vs lower limb injury	6%	11%	1.9	0.68 – 5.3

Table 17
Relative Risk by Postural Tolerance Scores

	Percentage of Injured Group Members			
Postural Tolerance task	*Score X*	*Score O or F*	*Relative Risk*	*95% CI*
Reach Forward	18%	14.5%	0.81	0.40 – 1.64
Reach Overhead	21%	21%	0.99	0.39 – 2.49
Stoop	20.5%	25%	1.21	0.65 – 2.26
Squat	20.6%	25%	1.21	0.49 – 2.98
Climb	19%	18%	0.94	0.6 – 1.48

Balance.
One hundred and thirteen people were assessed as having balance limitations. The remaining 223 did not have balance limitations. Fifteen of the persons with balance limitation reported injuries (13%) compared with forty-seven (21%) of the remainder. This relative risk of 0.63 is not significant (95%CI 0.37 to 1.08).

Fitness.
Seventy-nine (23%) workers were unable to complete the aerobic fitness test as a result of their heart rates reaching 85% of their maximum heart rate before the conclusion of the three minutes. This is an indicator of a low level of fitness. Five people were assessed as having “poor” fitness, 108 as “fair”, seventy-six as “average”. Of this group totalling 268, fifty-seven reported sprain or strain injuries. Sixty-six workers recorded a “good” fitness level, and two were “excellent”. Five of the good to excellent group recorded sprain or strain injuries. The relative risk of sustaining a sprain or strain injury for those that scored lower than good is 2.89 (95%CI 1.21 to 6.94) (Figure 25). This is statistically significant.

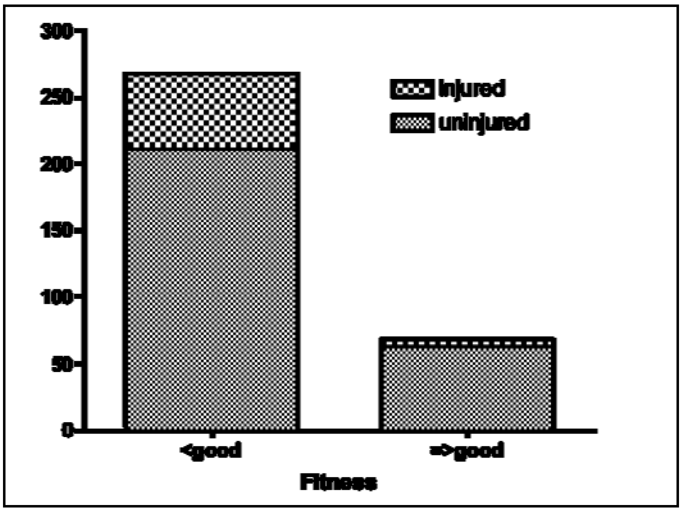

Figure 25. Number of Injured Workers by Aerobic Fitness Category

Reasons for not allocating PEFA=1
Of the reasons for not allocating PEFA=1, for twenty-two of the eighty-two PEFA>1 workers, the reasons were because of postural tolerance deficits only. All but two of these included climbing test deficits. Six of this group of twenty-two (27%) sustained injuries. If these are removed from the overall injury relative risk calculation the relative risk reduces slightly from 1.83 (1.16 to 2.87) to 1.73 (1.07 to 2.83).

Whether this increase in sensitivity is sufficient to justify inclusion in the PEFA is not an easy question to answer. It does seem that with the exception of the climbing test, the other postural tolerance tests as they are currently used may not add much to the sensitivity of the PEFA. Whether they might, if weighted more highly, is impossible to determine conclusively from the current data, however given the results above, it seems unlikely.

This conclusion is consistent with the outcomes of the reliability study which suggested that the material handling and climb tests were the most reliable measures.

Discussion

PEFA scores were determined by comparing the workers performance to the job demands. If the workers performance matched the job demands, they scored 1. If there was a mismatch, they scored a 2, 3 or 4 (PEFA >1). PEFA scores were compared to injury statistics as a measure of the validity of the assessment tool. There are however a number of external factors that could also affect injury rates which should be considered when interpreting these results. These include: age, exposure and external factors such as the environment.

According to the national workers compensation statistics (ACC, 2006), the proportion of new claims within each age group generally increases with age. Considering industry concerns about the aging workforce addressing age and the likelihood of injury was considered to be of interest. This study found no relationship between age and likelihood of injury.

One possible explanation for this finding is that younger workers tended to be employed in the higher risk jobs (underground and workshop) and older workers tended to be employed in the lower risk jobs (open cut operations). This higher / lower risk combination could have accounted for a 'cancelling out' effect of the protective younger age factor by the harmful higher exposure factor. An alternative explanation could be that older workers are less likely to take risks than younger workers.

Regardless of age however and as expected, there was an increased probability of sustaining an injury related to exposure. Exposure is a known risk factor for sprains and strains in the workplace. The 'duration of exposure' is in part determined by a worker's age under the assumption that the older they are, the longer they have been working and thus the more exposure to risk factors that they have had. Analysis of the demographic data at this site indicated that age made no difference to length of employment during the study period and thus could not be considered as a contributing factor.

PEFA scores however did have an effect on the length of employment such that turnover rates were higher in the PEFA>1 group. Higher injury rates in this group despite the reduced exposure, indicate that the results may be underestimated and that the injury risk in the PEFA>1 group may have been higher than what was recorded in this study if the exposure for the two groups was the same.

Examination of the injury statistics in this study demonstrated that sprains and strains related to manual handling tasks were the most common and that injuries to the back were the most frequent. These rankings are consistent with the Australian mining data and thus the sample set was considered to be reflective of the current industry situation.

A notable difference between the two groups (PEFA=1, PEFA>1) which is also related to exposure was the departmental mix. The study found that underground workers, regardless of PEFA score, were more likely to have an injury when compared to the other departments. The first and logical response would be to explain this by exposure. That is, underground workers are more likely to be exposed to higher job demands in a less predictable environment with greater variability in tasks and thus are presented with a higher risk of sprain or strain injuries. The argument for the job-specific PEFAs is that the assessment criteria (based on actual job demands) are designed to account for this variation. This is typically the case. In the case of the study site however, there was a management decision prior to the commencement of the project, to lower the job demands in the assessments as the criteria were determined to be 'too high'. Based on the author's experience, this is not uncommon practice in this and other industries as employers strive to find a balance between practicality, accuracy and safety. This could have resulted in more workers scoring 1 who if compared to actual job demands, may have scored lower. Examination of the raw data (job demands of the task at time of injury) for those underground workers who scored PEFA=1 and sustained a strain injury from manual handling indicated that this may have been the case in some (but not all) of the cases.

One major assumption used in this study was that the PEFA score at pre-employment was a reasonable representation of the worker's capabilities at the time of injury. This information will never be accurately known since a PEFA cannot be reassessed at the time of injury. However, this does raise the argument for regular reassessment. Transfer between jobs or when job demands change or post injury / extended absence would be a natural point of reassessment but for those workers who remain in the same job, a periodic schedule would need to be established. Exposure would be a reasonable determinant as increased exposure theoretically results in increased risk. Based on this argument, we could reasonably propose that reassessments would be more frequent for higher risk jobs (eg underground or workshop) or higher risk workers (eg older or PEFA>1). From this data set alone however, we are unable to determine how often PEFAs should be redone to maintain a current record of a worker's capabilities in comparison to their demands of their job.

One of the objectives of this study was to determine if the PEFA as a whole or in parts could be predictive of injury. We have established that those workers with a PEFA>1 are 3.56 times (95% CI 1.5 to8.47) more likely to sustain a back, trunk, or shoulder injury related to manual handling. The overall PEFA score was determined primarily by the manual handling (lifting and carrying) results and thus we can conclude that lifting and carrying performance in relation to job demands is a valid indicator of injury risk.
There was insufficient data to make meaningful comparisons between different postural tolerance tests and injury rates. A number of comparisons were made based on assumptions such as: reduced tolerance to reaching forward or overhead lifting and shoulder or neck injuries; reduced tolerance to stooping or reaching forward and back injuries; and, reduced tolerance to squatting or balance deficits and lower limb injuries. Although the percentage of these injuries in the PEFA>1 group was higher than the PEFA=1 group, the data volume was insufficient to draw firm conclusions.

Fitness test results in the validity study indicate an increased risk of injury for those workers scoring less than 'good' aerobic fitness. Although the test-retest reliability of the 3 minute step test in the reliability study was poor, the broad grouping of workers into 'less than good' or 'good or better', would account for the variation and still allow conclusions to be drawn. These findings are consistent with previously mentioned studies citing an increased risk of musculoskeletal injury based on aerobic fitness

Underlying all of this is the methodology for allocating the applicable PEFA score. In the study, manual handling played the biggest role in determining the overall score. Postural tolerances had a lower ranking and only reduced an injured worker's score in just over a quarter of the cases. Removing these workers from the study group did not make a significant difference to the overall results. Based on the available data it is difficult to determine if their inclusion in the PEFA makes a significant difference to their predictability of injury. Conversely, there is also insufficient data to rule for their exclusion.

For the purposes of this study, it was disappointing not to be able to compare PEFA scores with injury statistics such as duration and cost. Governed by the requirements of the National Privacy Principles which allow for the use of health related information for research purposes, improved information sharing between employers, insurers and researchers will be the only way to accurately measure this relationship. A multipartite commitment to this research topic is required for this to occur.

Conclusion

A number of conclusions can be drawn from the validity study:

- The JobFit System PEFA is a valid predictor of workplace injuries with workers scoring PEFA >1 being 3.5 times more likely to sustain a back, trunk or shoulder injury from tasks involving manual handling.
- The JobFit System PEFA is a valid risk management tool for identifying workers at an increased risk of manual handling injuries to the back, trunk or shoulder so that steps can be taken to reduce those risks (including task redesign, job rotation, team mix and conditioning programs).
- Based on the results obtained from this data set, the PEFA in its current form is not as sensitive in an underground environment. However their accuracy is still questionable as the assessment criteria undervalued the actual job demands and so were likely to have underestimated results. Job demands should be reevaluated and assessment criteria reassessed.
- Manual handling had the biggest impact on overall PEFA score and thus potentially is the strongest indicator of risk

- If the reliability of the fitness test can be improved, it may also be a valid indicator of injury risk
- The JobFit System PEFA score was also associated with workforce retention rates.

Project Conclusions and Industry Recommendations

There are a number of conclusions that can be drawn from this study. These have been tabulated with the resultant recommendations below (Table 18):

Table 18
Project Conclusions and Industry Recommendations

Conclusion	*Recommendation*
The JobFit System PEFA is a reliable tool for measuring the functional performance of workers	o The mining industry has the confidence to use this tool using its standardised processes and suitably qualified assessors
The JobFit System PEFA is a valid tool for predicting the risk of back, trunk or shoulder injuries from manual handling	o The mining industry has the evidence to use this tool for risk management activities associated with preventing sprains and strains in the workplace
The manual handling component of the assessment is the strongest indicator of risk	o More research is required to determined the impact of postural tolerances and aerobic fitness testing
Accurate job demands are critical to ensure accuracy of results and indications of risk	o Employers need to acknowledge the high job demands of some roles based on task analysis and if considered to be 'excessive' then work on changing the job demands rather than artificially lowering them by changing the assessment criteria. o Build job demand reviews into change management processes so that task demands are updated as conditions (eg equipment and processes) change
The JobFit System PEFA predicts injury risk and is not an absolute indicator of the occurrence or absence of injury	o Tool should be used as part of a risk management process. o Employers need to continue to address items higher in the hierarchy of control such as task and equipment redesign. o Worker conditioning programs (eg strengthening and fitness) would logically benefit those at higher risk but are still applicable to the whole workforce as PEFA = 1 does not 'protect' a worker from injury
A PEFA is a 'snapshot in time' and is likely to be most valuable when repeated during the course of employment	o Employers need to establish review processes based on job risk (ie job demands), worker risk (ie PEFA score) and exposure risk (ie time) so that assessment results remain current
Full value of the data has not been realised due to limited access to injury data	o The industry needs to improve cooperation and communication between stakeholders (eg employers, insurers, workers, unions and healthcare providers) to obtain the full picture.

Acknowledgements

This project number C14045 titled 'Validating the JobFit System Functional Assessment Method' was supported by a grant from the Australian Coal Association Research Program (ACARP)

Thanks goes to the ACARP team (Anne Mobardi, Roger Wischusen and Richard Jennings) and industry monitors Cameron Pocknee (former Safety Manager, Xstrata NCA Project) and Michael Solomon (Manager Site Services, Xstrata NCA Project) for their support and advice during this project.

Thanks also to the Human Resources and Safety teams from Xstrata Newlands Coal mine who assisted in the data collection processes: Adam Austin (Safety & Training Superintendent, NUG), Michelle Gomersall (Senior Administration Officer, NUG), Kylie Cervai (Administration Officer, NUG), Kris Riemer (HR & Payroll Officer, NOC), Simone Whelan (Senior Administration Officer, NOC) and Mary Whitney (Executive Assistant, NCA Project) and Sarah Nausbaum (HSEC Officer, NCA).

A special thanks also goes to the volunteers from Xstrata Newlands Underground Coal who participated in the PEFAs for the reliability study – particularly those that did it twice!

Your time, contribution and support are very much appreciated.

References

Anderson, C.K. (1999). Chapter 70: Physical Ability Testing for Employment Decision Purposes. In Karwowski, W. & Marras, W.S. (Eds). *The Occupational Ergonomics Handbook*. Boca Raton: CRC Press LLC. pp 1291-1298

Australian Government: National Occupational Health and Safety Commission (2004) The Cost of Work-related Injury and Illness for Australian Employers, Workers and the Community. http://www.ascc.gov.au/NR/rdonlyres/C8FC4DE9-8786-4DD6-BAFC-063F5EFC8C64/0/CostsOfWRID.pdf Australian Government

Australian Government: Australian Safety and Compensation Council (2006) Compendium of Workers Compensation Statistics Australia 2003-04. http://www.ascc.gov.au/ascc/AboutUs/Publications/StatReports/CompendiumofWorkersCompensationStatistics2003to2004.htm Australian Government

Battié, M.C., Bigos, S.J., Fisher, L.D., Hansson, T.H., Jones, M.E. & Wortley, M.D. (1989) Isometric Lifting Strength as a Predictor of Industrial Back Pain Reports. *Spine 14(8),* pp 851-856

Bigos, S.J. & Battie, M.C., (1987) Preplacement Worker Testing and Selection Considerations, *Ergonomics 30(2)*, pp 249-251.

Bradbury, S. & Roberts, D., (1998) *WorkHab Australia Functional Capacity Profiling Procedure Manual*. Australia: WorkHab Australia

Cady, L.D., Bischoff, M.P.H., O'Connell, M.S., Thomas, P.C. & Allan, J.H. (1979) Strength and Fitness and Subsequent Back Injuries in Firefighters. *Journal of Occupational Medicine 21(4),* pp 269-272

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

Gross, D.P. & Battie, M.C., (2002) Reliability of Safe Maximum Lifting Determinations of a Functional Capacity Evaluation, *Physical Therapy 82(4)*, pp 364-371.

E. Innes & L. Straker, Reliability of work-related assessments. *Work **13*** (1999), 107-124

Innes, E. & Straker, L. (2003) Attributes of Excellence in Work-related Assessments. *Work 20*, pp 63-76.

King, P.M., Tuckwell, N. & Barrett, T.E., (1998) A Critical Review of Functional Capacity Evaluations. *Physical Therapy 78(8)*, pp 852-867

Legge, J., (2004) Pre-Employment Functional Assessments as an Effective Tool for Controlling Work-Related Musculoskeletal Disorders: A review. *Ergonomics Australia 18(2),* pp 27-30.

Legge, J. & Burgess-Limerick, R., (in press) Reliability of the JobFit System Pre-Employment Functional Assessment Tool. *Work*

Mooney, V., Kenney, K., Leggett, S. & Holmes, B., (1996) Relationship of Lumbar Strength in Shipyard Workers to Workplace Injury Claims. *Spine 21 (17),* pp 2001-2005

Mostardi, R.A., Noe, D.A., Kovacik, M.W. & Porterfield J.A., (1992) Isokinetic Lifting Strength and Occupational Injury: A Prospective Study. *Spine 17(2),* pp 189-193

Queensland Government: Department of Mines and Energy (2006) Queensland Mines and Quarries Safety Performance and Health Report. 1 July 2005 to 30 June 2006. http://www.nrw.qld.gov.au/mines/publications/safety_health/mqsafe06.html Queensland Government

Queensland Government: Department of Industrial Relations (1995) Queensland Workplace Health and Safety Act 1995. http://www.whs.qld.gov.au/whsact/act95v1.pdf: Queensland Government

Randolph, D.C. (2000) Use of Functional Employment Testing to Facilitate Safe Job Placement. *Occupational Medicine: State of the Art Reviews 15(4),* pp 813-821

Reimer, D.S., Halbrook, B.D., Dreyfuss, P.H. & Tibiletti, C., (1994) A Novel Approach to Preemployment Worker Fitness Evaluations in a Material-Handling Industry. *Spine 19(18)*, pp 2026-2032.

Reneman, M.F., Dijkstra, P.U., Westmaas, M. & Goeken, L.N.H. (2002) Test-Retest Reliability of Lifting and Carrying in a 2-day Functional Capacity Evaluation. *Journal of Occupational Rehabilitation 12(4)*, pp 269-275.

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

Saunders, R.L., Beissner, K.L. & McManis, B.G. (1997) Estimates of Weight that Subjects can Lift Frequently in Functional Capacity Evaluations. *Physical Therapy 77(12),* pp 1717-1728

Scott, L.R. (2002) Post Offer Screening. *American Association of Occupational Health Nurses Journal 50(12),* pp 559-563

Snook, S.H., (1987) Approaches to preplacement testing and selection of workers. *Ergonomics 30(2)*, pp 241-247

Thomas, J.R. & Nelson, J.K., (2001) *Research Methods in Physical Activity. 4th Edition*. USA: Human Kinetics

Tuckwell, N.L., Straker, L., Barrett, T.E. (2002) Test-retest reliability on nine tasks of the Physical Work Performance Evaluation. *Work 19***,** pp 243-253.

List of Publications, Presentations and Reports

Publications

Legge, J. & Burgess-Limerick, R.., (in press) Reliability of the JobFit System Pre-Employment Functional Assessment Tool *Work*

Legge, J. & Burgess-Limerick, R.., (2006) Pre-Employment and Periodic Functional Testing: A review of the evidence. *Conference Proceedings 2006 Queensland Mining Industry Annual Safety Conference*

Legge, J. & Burgess-Limerick, R.., (2006) Pre-Employment and Periodic Functional Testing: A review of the evidence. *Conference Proceedings 2006 Human Factors & Ergonomics Society Annual Conference*

Presentations

'*Reliability of the JobFit System Functional Assessment Method*'; Queensland Mining Industry Annual Safety Conference 2005 (Poster)

'Pre-employment Assessments: Do they do what we think they do?'; NSW Minerals Industry OHS Conference 2006

'Pre-employment Assessments: Do they do what we think they do?'; Queensland Mining Industry Annual Safety Conference 2006

'Pre-Employment Assessments: The First Step to Workplace Wellness'; Queensland Safety Forum 2006

'Pre-Employment Assessments: A Review of the Evidence'; The Safety Conference 2006

'Pre-Employment & Periodic Functional Testing: A Review of the Evidence'; Human Factors & Ergonomics Society Annual Conference 2006

'*The Case for Pre-Placement Functional Testing: The Evidence*'; Annual Applied Ergonomics Conference 2007

'Is Functional Testing Functional?'; Human Factors in Safety Conference 2007

Reports

ACARP Progress Report (2005)
ACARP Interim Report (2006)

List of Figures and Tables

List of Figures

List of Tables