Activity Limitations as a Health Outcome: Measurement and Bias.

Rachael Powell

Thesis submitted for the degree of Doctor of Philosophy

School of Psychology

University of St Andrews

November 2002

Candidate's Declarations

I, Rachael Powell, hereby certify that this thesis, which is approximately 87,720 words in length, has been written by me, that is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date: 27.11.02

Signature of candidate:

I was admitted as a research student in September 1999 and as a candidate for the degree of Doctor of Philosophy in August 2000; the higher study for which this is a record was carried out in the University of St Andrews between 1999 and 2002.

Date: 27.11.02

Signature of candidate:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

Date: 27.11.02

Signature of candidate:

Supervisor's Declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: 16/10/02

Signature of supervisor:

For Mum

1946 - 1999

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Marie Johnston, without whom it would never have occurred to me to embark on the PhD process. Her clear thinking and encouragement have been invaluable throughout, as has the laughter when projects turned interesting.

I am very grateful to Derek Johnston for advice on the technological aspects of the thesis, for reading and giving feedback on drafts and for cooking delicious paella. I would also like to thank Debbie Bonetti and Beth Pollard for their time and patience in talking through measurement and statistical problems and ideas. The support from the other members of the Health Psychology group has been invaluable. Thanks also to friends and colleagues in CREATE.

Thanks are definitely due to the School of Psychology's support staff, especially Brian Kirk and Lesley Noble, for their patience and good humour when keeping my computer and measurement devices functioning.

I am particularly grateful to all the participants who donated their time and energy to taking part in these studies.

I have received a lot of support from a number of friends over the last 3 years, for which I am extremely grateful. I would, however, particularly like to thank my good friend and office-mate Stefanie Sonnenberg for being, quite simply, great.

Last but most definitely not least, thanks go to my family – Dad, Rhiannon and Rosalind – for never doubting this thesis would appear.

This work was funded by a studentship from the MRC HSRC.

CONTENTS

Appendices 294

THESIS ABSTRACT

Background

Measuring health status, often operationalised as activity limitations, is important both in research and clinical settings. Methods of assessing activity limitations include self-report, proxy-report, observed performance and continuous monitoring measures. However, different measure types may assess different constructs. Additionally, negative affectivity, social desirability and gender may bias activity measures. Emotion and control cognitions predict activity limitations but their effects may be dependent on the measurement type used and emotion and control cognitions may not have independent effects.

Method

Self-report, proxy-report, observed performance and continuous monitoring measures were cross-sectionally compared within 4 studies, with 3 participant groups: healthy adults, adults with walking difficulties and stroke patients. Effects of potential biasing variables were assessed cross-sectionally. The predictive effects of emotion and control cognitions were longitudinally examined. An experimental study investigating the effects of mood and perceived control manipulations on walking speed was performed with a healthy adult sample.

Results

Correspondence between measure types was not generally high. It appeared that, the closer the activities assessed by 2 measures, the stronger the association. Evidence supporting biases of negative affectivity, social desirability and gender was not found. Overall, emotion and control cognitions most consistently predicted activity on self-report and observed performance measures. Further research should investigate continuous monitoring and proxy-report measures. Within studies, the pattern of prediction by psychological variables differed across measures. Neither mood nor perceived control manipulations were found to affect activity but the perceived control manipulation did not affect reported perceived control levels.

Conclusion

The assessment methods seem to represent different phenomena; they measure different, albeit related, aspects of activity limitations. It cannot be assumed that a finding with one measure will replicate with another. Measures specific to research questions should be selected and using more than one measure type should be considered.

INTRODUCTION

i. Overview

Measuring health status is important both clinically and in research. Clinicians need to be able to assess patients effectively in order to determine appropriate treatments. If research is to find ways to improve recovery from illnesses then valid methods of measuring changes in health status are necessary.

Health outcomes are often operationalised as disability or activity limitations, 'difficulties an individual may have in the performance of activities' (WHO, 1999), activity being defined as 'the performance of a task or action by an individual' (WHO, 1999). A second, not incompatible definition of physical activity is 'any bodily movement produced by skeletal muscles that results in energy expenditure' (Caspersen, Powell, & Christenson, 1985). This is distinct from exercise: 'a subset of physical activity that is planned, structured, and repetitive and has as a final or an intermediate objective the improvement or maintenance of physical fitness' or physical fitness: 'a set of attributes that are either health- or skill-related…the health-related components of fitness are a) cardiorespiratory endurance, b) muscular endurance, c) muscular strength, d) body composition and e) flexibility' (Caspersen et al., 1985). Physical activity is the main outcome of interest in this thesis and 'activity' and 'activity limitations' are regarded as a continuum with those severely limited at one end and exceptional athletes at the other. No individual is capable of limitless activity and so the concept of 'activity limitations' is no more applicable only to individuals with disabling conditions than is the concept 'activity' applicable only to fit, 'healthy' people. Studies were conducted both with participants who were 'healthy' and with participants who had 'health conditions' (stroke patients and people with walking difficulties).

Activity limitations are commonly assessed with self-report or observed performance measures. However, both these measure types have associated problems and it is unclear to what extent activity limitations are assessed rather than the confounding variables. Proxy-report measures are also used, sometimes being used as a substitute for self-report where participants are unable to respond for themselves. Self-report, proxy-report and observed performance measures and the benefits and disadvantages of each are discussed below, and the potential for automated measures of activity to act as validation standards for such measures is considered.

Negative affectivity, social desirability and gender are factors that have been suggested to bias self-report measures of activity. Whether such biases are a concern in assessing physical activity is investigated.

Emotion and perceived control have been reported to affect activity limitations and recovery, but it is unclear whether their effects depend on the measurement type used to assess the health outcome. It is also not clear whether perceived control and emotion have independent effects on activity or whether one of these variables might mediate the other.

These issues are discussed in detail below and lead to the design of four studies. The first two studies (Chapters 2 and 4) involve small samples of students (Chapter 2) and people with walking difficulties (Chapter 4) in the comparison of self-report, performance and ambulatory monitoring assessments of activity. The intervening chapter, Chapter 3, discusses instrument calibration issues. Chapter 5 reports on a larger-scale study using self-report, proxy-report and performance measures with stroke patients. Finally, an experimental study was conducted with a student population, examining the effects of manipulating emotion and perceived control on walking speed (Chapter 6).

Part 1: Measurement of Activity Limitations

ii. Self-report Measures.

Self-report measures of activity involve asking participants about their activity patterns. Scales are usually administered either by having participants complete pencil-and-paper questionnaires, or with an interviewer reading out questions and recording answers.

Content of Assessments

The activities assessed in a physical activity questionnaire depend on the population being studied. Whilst it would be reasonable to ask a healthy population for time spent performing sport and number of miles walked each day, the same questions put to a population with functional limitations would lead to large floor effects. Hence questionnaires for chronically ill populations tend to focus on less demanding activities or activities of daily living (ADLs), rather than energy expenditure. The Functional Limitations Profile (FLP) is one such measure, including the statement 'I am not doing any of my usual physical recreation or more active pastimes' (Patrick & Peach, 1989).

ADLs are basic activities required in a community setting such as dressing, bathing and mobility (Ostir, Carlson, Black, Rudkin, Goodwin & Markides, 1999) and are frequently used to assess independence and disability. Such items are only useful in assessing the activity of very disabled individuals (Kovar & Lawton, 1994). Some measures include instrumental activities of daily living (IADLs) such as using the telephone, walking half a mile, handling money; these typically require a higher level of cognitive functioning and are not purely physical tests (Ostir et al., 1999). In institutional settings, individuals are often not required to perform IADL tasks, and contextual factors can also be problematic with ADL items. For example, in many residential homes for the elderly, assistance with bathing is mandatory, so a questionnaire item recording requiring assistance as a disability could be vulnerable to bias, particularly amongst individuals with middle levels of limitation compared with the very able or the very severely affected (Kovar & Lawton, 1994).

In ADL scales, it is debatable whether it is best to ask people whether they 'do' or 'can' perform tasks. Asking what people 'do' fails to account for the effect of environmental or social constraints, or personal preferences in activity performance (Kovar & Lawton, 1994). For example, someone could decide not to go outside because of poor weather, rather than because of inability to walk down the path. It could be that an activity is only performed with assistance because the carer wishes to help, rather than because the patient really can not manage. However, asking participants what they 'can' do will be an assessment of what participants think they can do, which could be an over- or under-estimate of their true ability. The FLP (Patrick & Peach, 1989) gives participants statements such as 'I do not walk at all'. Participants are asked whether they agree with the statements, and whether the agreement is because of the state of their health. Hence situational factors are limited without asking participants to estimate what they could do.

It is usual for physical activity questionnaires given to non-impaired populations to ask respondents about the activity they usually perform or what has been done over a particular time period. It would seem illogical to ask someone whether they were able to run a mile as, for the many individuals who do not run, responses could reflect self-confidence rather than actual ability. One difference between ADL measures and questionnaires designed for healthy populations is that ADL scales are frequently used as dependent variables, as measures of health outcome whereas physical activity questionnaires tend to be used as independent variables in detecting, for example, whether those who exercise more suffer less heart disease.

In questionnaires designed for more active populations, estimates of activity level are often obtained through either summing the time spent in activity or through summing the time spent in activity weighted by estimated intensity (Kriska & Caspersen, 1997). Many measures give energy expenditure or exercise intensity in METs. A MET is a measure of activity intensity, where one

MET 'represents the metabolic rate of an individual at rest and is set at 3.5 ml oxygen consumed per kilogram body mass per minute' (1 kcal/kg/hr) (Kriska & Caspersen, 1997). Therefore, if an activity has a score of 4 METs, it requires four times as much energy as resting. Such estimates make no assumptions about the association between energy expenditure and body mass (Jacobs, Ainsworth, Hartman, & Leon, 1993). If required, estimates in MET-hours can be converted to kilocalories by multiplying by body weight (Kriska & Caspersen, 1997).

Many physical activity questionnaires only assess leisure activities as occupational physical activity levels have decreased (Kriska & Caspersen, 1997). However, whilst, as Kriska and Caspersen suggest, this method is unlikely to be valid for older, disabled populations, it is also likely to be problematic for younger, healthier populations where, for some, travelling to work could be a major part of their regular exercise and, although physical activity at work may have decreased, many jobs still require much activity. Activities of light intensity have been found to be less reliably assessed than more vigorous activities but do appear to have importance when assessing the impact of external events such as seasons on activity (Levin, Jacobs, Ainsworth, Richardson, & Leon, 1999, see below).

Time Frames

Some questionnaires ask about activity over a long time period (e.g. the Minnesota Leisure-Time Physical Activity Questionnaire) whilst others refer to activity in the last week or day (e.g. the KIHD 24-Hour Total Physical Activity Record, see Kriska and Caspersen, 1997). Those assessing short time frames have the advantages of being less vulnerable to recall bias and more easily validated with objective measures (Kriska & Caspersen, 1997). Blair, Haskell, Ho, Paffenbarger, Vranizan, Farquhar and Wood (1985) proposed using recall over a week, stating it to be more accurate than general recall of 'usual' activity.

On the other hand, short assessment periods are less likely to reflect the behaviour usually performed; activity could vary according to seasons, illness and time available (Kriska & Caspersen, 1997). Levin et al. (1999) took monthly records of activity, using both self-reports and an automated measure of activity. It was found that higher levels of physical activity were recorded in the warmer months of the year for group mean scores, but, for many individuals, activity scores did not vary over the year. Month of the year reduced intra-individual variance by 7% in a self-reported 4-week measure, but only by 5.2% for automated records and 4.2% for a self-reported measure taken concurrently with wearing the activity monitor for 48 hours (participants reported activity as frequently as was convenient, generally every 4 hours). Levin et al. (1999) suggested that these results were due to the type of questions asked: whilst both 48-hour measures included activities of daily living, the four-week histories did not. It is likely that the type of activity used will affect the variation caused by factors such as seasons and time available. The optimal time frame will probably depend on the purpose of the study.

Hence a range of activity questionnaires has been devised to assess activity according to population and research question. Advantages and disadvantages of self-report measures are discussed below.

Self-report Measures: Benefits and Problems.

An advantage of self-report measures is that a wide range of activities can be assessed (Harding, Williams, Richardson, Nicholas, Jackson, Richardson, & Pither, 1994); the researcher is not restricted to activities that can practicably be completed in front of the researcher as are performance measures. Activities undertaken from the participant's perspective are measured; participant reports have the potential to more accurately reflect actual activity than the reports of others, such as a caregiver or physiotherapist, as proxy reporters can only report on what they observe – they can not assess activity of which they are unaware.

Questionnaires are also useful because they are non-reactive in the sense that they do not alter the behaviour examined (Kriska & Caspersen, 1997) as respondents are asked about activity already performed. In contrast, where laboratory tests of activity are conducted, individuals could be performing activities outside of their normal behaviour range. Self-report measures could, however, be reactive in that respondents could give the answers they perceive the researchers to expect. This issue of social desirability bias is considered below. Kriska and Caspersen (1997) assert that questionnaires can be engineered to suit a given population and are practical for large scale studies whereas performance and direct assessments of activity and energy expenditure are less feasible in such work because of high costs.

Recall Biases

A benefit of using self-report measures is that retrospective data can be collected (Craig, Prkachin, & Grunau, 1992). Unfortunately, there is also the problem of recall bias. Different types of activity are not necessarily remembered with the same ease: Wareham and Rennie (1998) claim that factors improving an activity's recall are its being discrete, time limited and requiring a purposeful decision to perform. Hence an activity such as playing football is apparently likely to be remembered more accurately than walking around the workplace during the day. Blair, Dowda, Pate, Kronenfeld, Howe, Parker, Blair and Fridinger (1991) and Sallis, Haskell, Wood, Fortmann, Rogers, Blair and Paffenbarger (1985) report the recall of vigorous activity to be more accurate than the recall of less intensive activity. Nisbet and Ross (1980) assert there is evidence that recent or salient memories are over-relied upon when estimating the frequency of event. It is proposed that self-report diaries, requiring participants to fill out activity undertaken at regular intervals during the day, have the potential to reduce retrospective bias (Kamarck, Shiffman, Smithline, Goodie, Thompson, Ituarte, Jong, Pro, Paty, Kassel, Gnys & Perz,1998) – all memories recorded will be recent, and activity may seem more salient as fewer events would have occurred since the previous report.

Culture, Language, Education Level and Individual Variation in Activity Report

Self-report measures are unable to access data from all sections of the population. Individuals with low-level cognitive functioning or communication difficulties may be screened out of a study. Culture, education level and language could affect responses to a questionnaire (Guralnik, Branch, Cummings, & Curb, 1989), and Kelly-Hayes, Jette, Wolf, D'Agostino & Odell (1992) found disparities between reported disability and a performance measure of functioning to be greater amongst participants with cognitive impairment.

Participants affected in similar ways by an illness could respond differently to the same questions (Walker, Heslop, Plummer, Essex, & Chandler, 1997) and questionnaires must manage the problem that individuals can vary in the way they report activity. Such variation can be caused by individual differences in experience (for example, walking speeds perceived to be 'fast' or 'slow' could be determined by an individual's normal walking speed) or by ambiguity of questionnaire items. Ainsworth, Haskell, Leon, Jacobs, Montoye, Sallis, and Paffenbarger (1993) suggest reducing variation by giving detailed instructions to respondents, for example stating that 3mph is moderate walking and having trained interviewers review the data with participants. This latter strategy seems likely to increase the risk of social desirability bias however: participants are likely to alter their reports in the direction seeming favourable to the researcher. Guralnik et al. (1989) favour giving precise definitions of activity as both validity and reliability can otherwise be reduced: it is unclear exactly what participants intend by responses and difficult for participants to give the same response repeatedly where activity definitions are not clear.

Apart from individual differences in responding to questions, especially ambiguous ones, individual variability in activity is also likely to affect the intensity ratings of activity. Compendiums are available from which MET values for various activities can be calculated e.g. Ainsworth, Haskell et al. (1993). However, it is assumed that the MET value is representative of the manner in which the activity is performed and activities can be performed at a range of levels (Kriska & Caspersen, 1997). Weighting activities by intensity also assumes that body weight is proportional to resting metabolic rate, and that the increase in METs is constant between individuals of any body weight (Kriska & Caspersen, 1997).

Gender Differences in Activity Patterns

A potential difficulty with questionnaires is that the traditionally different activity patterns of men and women may lead to similar overall activity levels being represented differently. Whilst women may undertake housework and childcare activities, men may perform more intensive physical activity (Kriska & Caspersen, 1997). As discussed earlier, different activity types are differentially recalled and general exercise around the house is likely to be difficult to gauge. In a different way, ADL measures may differentially represent men and women with the same level of disability. For example, the FLP (Patrick & Peach, 1989) includes statements such as 'I do not do any of the daily household chores that I would usually do'. If a participant has never been the one to do daily household chores, the item would not be agreed with whatever the level of functional limitations because a negative response to the question 'is it due to the state of your health?' would be given. Williams, Johnston, Willis and Bennett (1976) found that, in a comparison of disabled women and men who kept house with a view to assess a Guttman scale model of disability, women would accept limited mobility before accepting help with cooking whereas men would take cooking help before reducing mobility. Thus, depending on the questionnaire's design, male or female activity could be over- or under-represented.

Negative Affectivity Bias and the Effect of Mood on Recall.

Negative affectivity has been described as 'a general dimension of subjective distress…a broad range of mood states including anger, disgust, scorn, guilt, fearfulness, and depression' (Watson & Pennebaker, 1989). From an extensive literature review and testing 6 samples of their own, Watson and Pennebaker (1989) found self-report health measures to be related to negative affectivity. As negative affectivity was not consistently associated with objective health indicators such as evidence of risk, pathology or mortality, it would seem to be a confounding variable. Both trait and state negative affectivity were found to correlate with symptom reporting although trait negative affectivity showed the strongest correlations. Their findings supported the symptom perception hypothesis, that individuals high in negative affectivity are hypervigilant in their attention to symptoms and ambiguous stimuli are interpreted negatively, rather than the psychosomatic hypothesis (high negative affectivity causes health problems) or the disability hypothesis (health problems cause high negative affectivity) as relationships between negative affectivity and health problems seemed to be limited to the self-report domain.

The symptom perception hypothesis has been supported by McCrae, Bartone and Costa (1976) and Tessler and Mechanic (1978). McCrae et al found negative affectivity to be correlated with self-reported symptoms but not physician-reported health. Tessler and Mechanic found distress measures to be significantly associated with perceived health status when entered into multiple regression equations alongside more objective 'physical health status' measures in 4 population samples.

Watson and Clark (1984) claim people with high negative affectivity are more introspective and ruminative than those with low negative affectivity. Watson and Pennebaker (1989) suggest this introspective style could enhance attention to normal or minor somatic symptoms. However, Watson and Clark (1984) suggest people high in negative affectivity are 'more emotionally self-aware and honest with themselves' which could imply that their worse self-reported symptoms could actually be an accurate reflection of sensation; it could be that a positive bias of self-reported symptoms is seen in the data of those without high negative affectivity. Mora, Robitaille, Leventhal, Swigar and Leventhal (2002) argued that the vigilance associated with trait negative affect (a construct including anxiety and depression, two key aspects of negative affectivity) would be a motivation source in acquiring information about symptoms enabling a more detailed account of usual symptoms than someone low in negative affect, predicting negative affect to be associated with everyday, background symptoms but unrelated to the reporting of symptoms in acute illness episodes. This hypothesis was supported.

It is possible that negative affectivity's effects on symptom reporting operate by affecting recall of positive or negative events; an associative network theory is described by Bower (1981). Teasdale and Fogarty (1979) induced happy or depressed moods in participants using the Velten mood induction procedure (briefly described in the Experimental Study, Chapter 6). Participants were then asked to associate past pleasant or unpleasant memories with a list of stimulus words. It was found that the time to retrieve pleasant memories, relative to the time required to retrieve unpleasant memories, was significantly longer for participants in the 'depressed' condition than for those in the 'happy' condition. It was suggested that mood had an effect on the accessibility of different types of memory. However, judges' ratings of the pleasantness of events recalled were similar for the 2 conditions so, although speed of recall differed, content of recollection did not differ significantly. It could be that a more significant difference would be seen with a larger sample size (here, $n = 16$). Whilst a depressed mood is not an identical construct to negative affectivity, it is a component of negative affectivity (Watson & Pennebaker, 1989).

Nevertheless, Bower (1981) did see content of recollection varying according to mood at recall. Participants recorded emotional events in a daily diary for a week and rated the events' pleasantness on a 10-point scale. After a 1-week interval, participants were asked to recall the incidents from their diaries whilst in an induced mood. Participants were hypnotised and asked to get into the required mood by imagining a scene where they had been either happy or sad. They were then asked to recall every incident they could from their diaries. It was found that those in a pleasant mood recalled a greater percentage of pleasant than unpleasant experiences whereas participants in an unpleasant mood recalled a higher percentage of unpleasant than pleasant memories. This study is unfortunately limited by its small sample size: as almost half the participants did not record enough diary incidents the sample was only 14 participants and the statistical significance of results is not reported.

In a following experiment (Bower, Gilligan, & Monteiro, 1981), participants were asked to recall childhood incidents whilst in happy or sad induced moods. The next day, whilst in 'neutral' moods, participants were asked to categorise each incident as pleasant, unpleasant or neutral. It was found that 'happy' participants retrieved many more pleasant than unpleasant memories (a 92% bias) whereas sad participants retrieved slightly more unpleasant than pleasant memories (a 52% bias). Again, no significance levels are reported. Thus it would seem that, when recalling real-life incidents, there could be an effect of mood at recall on events reported. However, as the mood induction techniques used in these experiments involved instructing participants to obtain the required mood, demand effects could be playing a role: it could be that participants are, to some extent, second-guessing the experimenters' hypothesis.

It is not only at recall that mood effects can have a role: it is possible that mood can affect the encoding of events. Bower and Mayer (1989) asked participants to learn different word lists in different moods (happy and sad) induced through the use of hypnotic trances as described above. They were then asked to recall the lists in *either* happy *or* sad induced moods. Whilst mood dependent recall was not clearly demonstrated, mood-congruent learning was seen: 'happy' participants learned 72% and 55% of pleasant and unpleasant words respectively whilst 'sad' participants learned 69% and 60 % unpleasant and pleasant words. This experiment involved learning in an artificial environment and it is not clear to what extent these findings would generalise to recall of real-life personal experiences but it nevertheless demonstrates another mechanism through which the negative affectivity bias could take effect. Teasdale and Fogarty's (1979) results were not replicated but Teasdale was looking at recall of personal memories rather than the more artificial retrieval of words learnt in the laboratory.

Mood congruity at encoding was also demonstrated by Bower et al. (1981). Sixteen undergraduate participants from a pool of 'highly hypnotisable' people had happy or sad moods induced whilst they read a story about 2 college men, one of whom was happy, the other sad. Participants recorded significantly more facts about the character with whom their mood had identified: 80% of the facts recalled by the sad readers were about the sad character and 55% of the facts recalled by the happy readers were about the happy character. Participants were in a neutral mood state during recall. A similar effect was found when participants were given simulated psychiatric interviews involving descriptions of happy and sad events to read. Participants again had happy or sad mood states induced whilst reading the material. 'Happy' readers recalled about one and a half as many happy incidents as sad, whereas 'sad' readers recalled about one and a third as many sad events as happy. Similar results were found when the story consisted of a single person recounting happy and sad events, suggesting that participants were not merely identifying with the mood-relevant person in the text. Bower et al. (1981) used a 2 x 2 factorial design (happy/sad mood on encoding (reading) x happy/sad mood on retrieval) to investigate whether mood at retrieval also influenced recall. A significant effect of mood on retrieval was not found. However, participants were also asked to estimate the proportion of sad to happy events in the story. It was found that, in this context, a significant effect of mood during testing was found but neither a main effect of mood at encoding nor an interaction effect was seen. This has implications in the area of activity reports: it could be that someone who is depressed at the time of being asked about activity might have similar recall of active events to someone who is not depressed but, if asked to estimate, for example, the overall length of time of activity or speed of walking, their estimate might be lower than that of someone who has performed similar activity but who is not depressed.

A general limitation in Bower's work is the sample used: with typically only 8 participants per experimental group, it is not clear whether the lack of significance of some results is because there is no effect or because the design lacked statistical power. Also, Bower et al. (1981) used only highly hypnotisable individuals because of the nature of the mood induction. This reduces potential participants to 20-25% of the general population (Bower, 1981) and so their results may not be highly generalisable.

In contrast, Parrott and Sabini (1990) found evidence for mood incongruent recall. The memories of students who performed well on a mid-term exam were compared with the memories of students who did less well. After their graded papers had been returned, they were asked to recall three events from their high school years. It was found that, for the first memory recalled, students whose grades were worse than expected recalled memories with more positive affect; no difference was seen for second and third memories. These findings are supported by further experiments using a musical mood induction procedure but mood incongruent effects were only seen when participants were unaware that the induction was designed to affect their mood. When participants were asked to help by maintaining the required mood, mood congruent recall was found. In all these studies, however, the mood was only recently induced and its effects seemed quite transient in only affecting the first memory recalled on each occasion. It is not clear to what extent this effect might be found with long-term negative mood or affectivity. It was suggested that participants might have been performing self-regulation of their moods.

On asking participants to recall target words from sentences presented for memorisation, Kwiatkowski and Parkinson (1994) found mood-congruent recall effects with naturally depressed participants compared with non-depressed controls but not with participants whose moods had been induced using a Velten-type procedure (described in the Experimental Study, Chapter 6).

Rholes, Riskind and Lane (1987) compared the effects of self-evaluative and somatic (describing somatic states that accompany mood states) Velten-type statements on recall latencies for life experiences. It was found that, although the effect of the two types of statement had equally strong effects on mood states, the self-evaluative statements had a stronger impact on recall latencies (negative events being recalled quicker by participants in negative moods and positive events being recalled quicker by participants in positive moods). The effects of the somatic statements seemed to be dependent on the mood state achieved whereas self-evaluative statements' effects appeared to be independent of this. Thus it would seem that the cognitions associated with an affective state can be important in determining recall of life events and not only the mood per se. As negative cognitions are associated with negative affectivity (see above), this could be another mechanism through which negative affectivity could influence patient symptom report.

The literature is not completely consistent however: Riskind, Rholes and Eggers (1982) found changes in mood following the Velten mood induction procedure to be significant for negative recall in the opposite direction to that predicted: the more positive the mood, the faster the recall ($r = -.24$). The correlation for positive recall was not significant.

Evidence has been reviewed suggesting that negative affectivity and mood may affect recall of a range of stimuli: self-reported symptoms, life events, and experimental materials. It seems likely, therefore, that negative affectivity could be expected to influence the reporting of physical activity. Individuals high in negative affectivity could perceive themselves to have performed less activity than individuals of low negative affectivity when actual activity levels are identical, although the generation of this hypothesis from the above literature does assume that to have been more active is considered positive and inactivity negative.

On the other hand, rather than confounding activity assessments, negative affectivity could be of use in the prediction of actual activity: it could be that individuals high in negative affectivity are less active than other people. The impact of negative affectivity on activity as assessed using different activity measures would, therefore, be worth investigating: is negative affectivity only associated with self-reported activity or disability, or is it also related to activity performed both in the laboratory under controlled conditions and in the outside world?

Social Desirability Bias.

'Social desirability bias' is the tendency of participants to report in such a way as to look good to the researcher. For example, this phenomenon has been investigated in the area of pain communication research. Craig and Patrick (1985) compared facial behaviour assessed using Ekman and Friesen's (1978) Facial Action Coding System (FACS) with verbal reports of pain using the cold pressor test (immersion of a hand and wrist in cold water) of 72 volunteers. FACS is a method of coding facial movement developed to reduce observer bias. Participants performed the experiment in the presence of a tolerant, intolerant or non-participatory model. It was found that models influenced verbal reports of pain and pain tolerance level (again, self-report), but did not affect facial expressions. It would, therefore, seem that, whilst self-reports were vulnerable to the wish to give the 'right' impression to other people, the more objective measure was unaffected. Similarly, Patrick, Craig and Prkachin (1986) compared self-report and facial behaviour in undergraduates given electric shocks in the presence of tolerant, intolerant or neutral models. Tolerant model participants showed higher pain thresholds and endurance levels and reported equivalent pain to other groups when shock intensity level was randomly ordered despite being given higher intensity shocks (shock levels were determined by individual pain thresholds and endurance limits). Using both FACS and untrained observers, however, tolerant model participants were rated as being in more pain than those with intolerant models.

In the domain of mental health, Phillips and Clancy (1970) found participants' evaluation of the social desirability of questionnaire items (psychological, psychophysiological and physiological indicators of mental disorder) to be related to whether or not they reported having experienced the symptoms.

Hebert, Clemow, Pbert, Ockene and Ockene (1995) tested social desirability biases by comparing nutrient scores derived from seven 24-hour diet with two 7-day diet recalls. Social desirability score was found to produce a bias in nutrient estimation in the 7-day recall compared with the 24-hour recall such that lower reported food intake was related to social desirability score. However, the social desirability measure was taken 2 years after the other measures without evidence being presented to suggest that social desirability levels remain unchanged over such long time periods. It is not known whether such a finding would generalise to the outcome variable physical activity.

Studies examining the effects of social desirability on the report of activity seem to be sparse. However, Klassen, Hornstra and Anderson (1976) found social desirability to be negatively associated with the reporting of depression, aggression, satisfaction and general happiness (higher social desirability being associated with less reporting of pathology) but did not find significant correlations between social desirability and either 'disability days' (the number of times in the week when activities were reduced because of not feeling good) or general functioning (the number of times various activities were engaged in over the past week).

It seems probable, therefore, that self-report measures of activity could be affected by social desirability bias to a greater extent than more objective assessments such as automated monitoring of every-day activity or performance measures, although performance measures are not without risk of social desirability bias as will be discussed below.

Gender Bias

Hakala, Nieminen and Manelius (1994) found no significant difference between males and females on an observed functioning test (the Keitel Function Test, KFT) in a sample of patients with rheumatoid arthritis. However, on two self-report scales, the Arthritis Impact Measurement Scale and the Health Assessment Questionnaire, males' functioning scores were higher than the females'. Unfortunately, Hakala et al. (1994) did not examine gender differences in scores between specific self-report and performance tasks. The self-report measures would have given an assessment of general health status whereas the KFT gives a specific indication of joint functioning and women seem to have more chronic illnesses than men, although the conditions are less severe (see Verbrugge & Wingard, 1987). Estlander, Vanharanta, Moneta, and Kaivanto, (1994) found self-reported pain severity and disability to have a negative effect on the isokinetic trunk muscle test of male low back pain patients but to have no effect on female patients' performance. Such an effect could be caused either by women over-reporting symptoms or by women persevering despite suffering symptoms.

Merrill, Seeman, Kasl and Berkman (1997) compared the scores of older males and females on 3 specific self-report and performance tasks out of a battery of self-report ADL and functional limitations questions and a range of observed tests. For example, reported ability to walk across a room was compared with observed ability to walk 8 feet and back. Overall performance scores explained all of the gender difference in ADL disability and most of the functional limitation measure difference and self-reported and performed ability were found to match in most specific items. However, where differences on items were seen, more men than women under-reported and more women than men over-reported disability (as compared with observed performance), supporting Hakala et al's results. Nevertheless, Ferrer, Lamarca, Orfila and Alonso (1999) also compared self-report and performance scores on specific tasks to find that disagreement between the measures was not influenced by participants' gender. It is possible that culture is influencing responding; Ferrer et al. (1999) researched an elderly Spanish population whereas Merrill et al. (1997) are based in USA and Hakala et al. in Finland.

It has been suggested that men and women could differ in their perception of the presence and severity of bodily symptoms (Hibbard & Pope, 1983; Wingard, 1984). Indeed, Hibbard and Pope (1983) found females to report more symptoms than males (their sample was restricted to adults who had rated their health as 'good' or 'excellent' in an attempt to control for actual health status). Mechanic (1976) reviewed a study carried out by the National Center for Health Statistics, which suggested women report more illness symptoms than men whilst a clinical examination suggested no difference in chronic illness. No such discrepancy was found with obvious symptoms such as nosebleeds, however. It was suggested that women could therefore be more sensitive to body cues, and more likely to report them as symptoms than men. Similarly, Verbrugge and Wingard (1987) cited studies finding no gender differences in reporting of events such as hospitalisation and doctor visits. They point out that interviewers in health studies are usually female and women may enjoy the opportunity to discuss morbidity with another woman. Conversely, men may be unwilling to admit illness, given society's emphasis on men being 'tough' as well as being reluctant to appear unhealthy to a stranger of the opposite sex. Thus, social desirability could differentially bias male and female self-reporting.

Hebert et al. (1995) found social desirability to predict a downward bias in nutrient estimation in a study examining the effect of social desirability on dietary report. This social desirability bias was approximately twice as large for women as for men. In the area of activity assessment, it could be that social desirability functions differently for men and women, with reporting high levels of activity being more important to men than to women.

Prkachin (1992) found male participants reported higher pain tolerance than females in a series of electric shocks and females showed more intense expressions of strong pain than males. It is possible that males have been socially conditioned to express pain through verbal and non-verbal modalities less than females. Unfortunately, FACS was not used here. Craig, Hyde and Patrick (1991) found independent judges to rate female back pain patients as more expressive than male patients whilst FACS detected no gender difference. Either observers were biased by gender stereotypes or FACS was insensitive to the gender differences. It is possible that interviewer expectation could influence both participant self-report and bias the interviewer's interpretation of information presented by the participant.

A factor proposed by Hibbard and Pope (1983) to influence symptom report was that women could be more interested in and concerned with health. This hypothesis was supported: women scored higher than men in health interest and concern. Women with a high interest in health reported more symptoms than women with low interest, although no such relationship was found among male subjects. As this study was cross-sectional, it could be that perceiving more symptoms increases an individual's concern with health. Farmer, Locke, Moscicki, Dannenberg, Larson and Radloff (1988) compared self-reported physical activity measures with physiological assessments (pulse, blood pressure, body weight, height and cholesterol). Among women, inactivity was associated with greater body mass index, blood pressure and heart rate whereas, for men, no relationship was found. It is possible that the two physical activity questions were too vague. One question assessed recreational activity, the other non-recreational activity – 'In your usual day aside from recreation, are you physically very active, moderately active or quite active?'. The questions rely on participants having similar understandings of how active being 'very', 'moderately' and 'quite' active entail.

However, gender effects have been found in studies assessing reported activity that seems less likely to be related to symptoms than those cited above. Bond and Clark (2000) compared wives' and husbands' reports of their own and their spouse's performance of domestic chores. Although the intraclass correlation coefficient of husbands' and wives' reports of husband's activity was 'good' at .69, the ratings were significantly different with wives rating husbands' domestic chores lower than the husbands rated themselves. It is unfortunately not possible to determine which partner was the most accurate. With an ingenious study design, Klesges, Hanson, Eck, Mellon, Fulliton, and Somes (1990) compared the activity ratings of participants who were unaware of observation with those of observers. Participants were instructed to engage in any activity they wished for 1 hour in a public track and fitness area. On their return to the laboratory, they were required to recall their activity for the previous hour using the same scales as a trained observer. Males were found to significantly overestimate their activity levels (mean (actual – reported activity) = 14.08) relative to female participants (mean = -6.15, p = 009).

As described earlier, Watson and Pennebaker (1989) argue that self-report health measures are strongly influenced by negative affectivity. Given that depression is suffered by more women than men (see Nolen-Hoeksema (1987) for review) it is plausible that, compared with men, women perceive their physical status less positively. 'Depression' is not synonymous with 'negative affectivity', but it is recognised as a mood contributing to negative affectivity (Watson and Pennebaker, 1989). Depressed individuals could be less active than non-depressed people; they could also have a lower sense of self-efficacy with respect to every-day tasks and so complete fewer tasks day-to-day. Fisher and Johnston (1996a; 1996b) found manipulations lowering mood and perceived control lessened performance on a lifting task. Depressives' task performance could, therefore, be limited by their psychological state. Thus negative affectivity and or depression could produce 2 routes through which women report higher levels of activity limitation than men: both perceptions of activities performed and actual performance of activity could be affected.

The evidence is not consistent in favour of a gender bias, however. Davis (1981) found no support for the hypothesis that women were more likely than men to report chronic joint symptoms and Blair, Dowda, Pate, Kronenfeld, Howe, Parker, Blair and Fridinger (1991) concluded participation in physical activity as recalled10 years after a baseline measure was comparable in men and women.

Hence gender is a factor worth exploring whilst investigating factors affecting self-reported physical activity assessments.

iii. Proxy Measures

Proxy measures of activity require a person acquainted with the index participant (usually a spouse, other relative or non-related caregiver or friend) to rate the index participant's activity. Using proxies enables researchers to include a wider sample of participants (Neumann, Araki, & Gutterman, 2000), for example where potential participants have communication difficulties. Proxy measures are similar to self-report measures in that they may be prone to biases such as of recall, but also have similarities to observed performance measures as the activity rating is made externally to the participant; measures may be affected by subjective processes but those processes are likely to be relatively independent of the participant's subjective processes. Proxy raters are usually people who spend a lot of time with the index participant but do not witness all activity. It is, therefore, important to determine the extent to which proxy ratings are accurate as compared with index participant ratings before assuming that studies can freely substitute proxy-report measures for self-reported index participant assessments.

Fourteen studies comparing self and proxy-ratings on physical activity-related measures are reported in Table 1.1. It can be seen that, whilst agreement on some measures was found to be excellent, others showed only poor or moderate agreement. Of these, 11 papers reported bias acting in the same direction: proxies appear to consistently give more negative reports of index participants' abilities than do the patients themselves even when agreement coefficients are high. It is not possible to determine from a simple patient-proxy comparison who is in error, but it is clear that freely substituting proxy ratings for patient ratings could bias data sets.

Paper	Index population	Agreement	Bias
Andresen, Vahle, & Lollar (2001)	Adults with disability	Kappa from .33 to .82	Proxies reported greater impairment.
Bond & Clark (2000)	Community dwelling adults aged 65+ years	ICC from .57 to .82	No bias.
Epstein, Hael, Tognetti et al. (1989)	Adults aged 65+ eligible for Harvard Uni Health Services.	r = .73	No bias.
Magaziner, Zimmerman, Gruber-Baldini et al. (1997)	Post hip-fracture patients	ICC from .56 to .65	Most items: proxy reported greater assistance required
Moinpour, Lyons, Schmid et al. (2000)	Patients with brain metastases	LCC from .36 to .70	Proxies reported lower quality of life.
Peck, Smith, Ward, & Milano (1989)	Rheumatoid arthritis patients	r = .89	Not reported
Rothman, Hedrick, Bulcroft et al. (1991)	Chronically ill veterans	r = .72	Proxies rated patients as more impaired.
Rubenstein, Schairer, Wieland, & Kane (1984)	Hospitalised patients > 65 years old.	Not given	Proxies rated lower functioning level
Sneeuw, Aaronson, Sprangers et al. (1997)	Mixed cancer patients	ICC = .63 baseline, .61 at follow-up.	Proxies rated patients as more impaired at baseline.
Sneeuw, Aaronson, deHaan, & Limburg (1997)	6 months post-stroke	ICC = .85	Proxies rated patients as more limited
Sneeuw, Aaronson, Osoba et al. (1997)	Brain cancer patients	ICC = .67	Proxies rated patients' physical functioning lower.
Sneeuw, Aaronson, Sprangers et al. (1998)	Mixed cancer patients	R = .74, ICC = .73	Proxies rated patients' physical functioning lower.
Sneeuw, Aaronson, Sprangers et al. (1999)	Mixed cancer patients	ICC = .66	Significant others gave worse ratings, no bias by nurses.
Sneeuw, Albertsen, & Aaronson (2001)	Prostate cancer patients	ICC = .71	Proxies gave worse reports

Table 1.1: Agreement and bias between self and proxy-ratings on physical activity-related measures.
ICC = Intraclass correlation coefficient. LCC = Lin's concordance coefficient.
Interpreting coefficients: <.41 = poor, .41 - .60 = moderate, .61-.80 = good, >.80 = excellent.

It would appear that the patient's level of functioning may influence index-proxy agreement. Sneeuw, Aaronson, deHaan and Limburg (1997) found the magnitude of patient-proxy agreement to be clearly associated with the patient's functioning level. The tendency of proxies to rate patients as being more limited than patients' ratings was more pronounced where patients had a more impaired functioning level. The authors report that their sample was less impaired than some other studies using the same measure (Sickness Impact Profile); on a quality of life measure, Sneeuw, Aaronson, Sprangers, Detmar, Wever Lidwina and Schornagel (1998) found poorer agreement at intermediate levels of functioning than where quality of life was rated good or poor. It seems logical that minimal or maximal disability would be more obvious to people around the patient than at these intermediate levels of disability.

Most of these studies record the numbers of each type of proxy (spouse, parent, offspring etc) but do not assess measure comparability for each group. Bond and Clark (2000), however, found wives to be more accurate than husbands whilst both were more accurate than offspring in giving proxy ratings (index participant ratings were assumed to be the 'gold standard'). Rubenstein, Schairer, Wieland and Kane (1984) found that proxies who were spouses tended to rate ability as lower than more distant contacts such as other relatives or friends. It could be that it is harder to hide disabilities from closer individuals, or this could reflect a greater sense of burden in spouses (Rubenstein et al., 1984). In support of this, although finding no significant differences between patient and proxy scores overall, Epstein, Hael, Tognetti, Son, and Conant (1989) found that, the more hours proxies spent assisting patients, the more functionally impaired proxies rated patients, compared with the patients' self-reports. Hence different compositions of populations across studies could influence results. It was also rarely indicated whether index and proxy participants were required to complete the questionnaires or interviews separately although in a number of papers participants were asked to complete them independently (e.g. Sneeuw, Aaronson, Sprangers, Detmar, Wever & Schornagel, 1999) and Epstein et al. (1989) interviewed participants separately. It is therefore unclear to what extent participants may have influenced each other.

Self-report and Proxy Measures: Conclusion

Validity is the ability of an assessment method to measure what it is designed to measure. If an individual's every-day activity level is to be determined, self-report measures have the potential to have a high validity compared with laboratory performance tasks or automated monitors because items can assess behaviour both beyond the limits of the laboratory and of the relatively small sample of time for which an automated activity monitor can be worn. However, as has been seen above, a range of factors from self-report bias to ambiguously phrased questions can confound self-report scores. Proxy measures may also be vulnerable to biasing factors. The ability of performance measures to overcome such difficulties will be considered below.

iv. Observed Performance Measures.

Performance measures of activity require participants to demonstrate tasks to an observer who rates ability or records dependent variables such as time taken to perform a task. As such measures do not demand that respondents rate their own activity, they have certain advantages over self-report measures but are not without their own problems.

Ability to Overcome Problems of Self-Report Measures

Performance measures are likely to be less influenced by culture, language, education level and cognitive functioning than are self-report measures (Guralnik et al 1989). These factors could have some impact on performance measures though as they could affect how comfortable people feel in the laboratory setting and participants need to understand exactly what it is they are required to do.

A major problem with the self-report paradigm is the vulnerability to various types of bias. This is less of an issue for performance measures as behaviour is directly assessed; it is not first interpreted by the respondent. Recall bias clearly can have no influence in a performance test. However, it could be hypothesised that negative affectivity and social desirability may bias performance scores. Someone for whom it is important to make a good impression may put a lot of effort into a task compared with someone who is not bothered about the image portrayed. Similarly, an individual who is high in negative affectivity may not be motivated to perform well. Whilst performance tests are designed to measure what individuals *can* do and avoid the problem self-report questionnaires face of choosing to test either ability or what participants report *actually doing*, they can not differentiate between motivation and ability (Myers, Holliday, Harvey, & Hutchinson, 1993). Myers et al (1993) administered 14 IADL performance tasks to 99 community-dwelling older adults. Which performance tests participants attempted was predicted by their perceived capability, as assessed by self-report of whether they performed the various IADLs, rather than the frailty of the individuals (indicated, for example, by advanced age and health problems). It is possible, though, that both the self-report and performance tests were detecting lack of ability.

Observed Performance Measures: Problems

Whilst self-report measures can assess events in the home and outside, without time constraints, performance tests are restricted to the test locations, such as the laboratory, clinic or the participant's home. Adaptations to activity problems are likely to be ignored in performance tests; individuals may have overcome an activity problem at home that is not replicable in the laboratory and may prefer to perform activities that are not included in the test. Everyday ability can be influenced by different factors. For example, Harding et al. (1994) assert that a participant who does not leave the house without assistance may walk without aids in a performance test. Simple tests may not reflect performance on more complex tasks (Guralnik et al., 1989); many daily activities require a combination of movements (Myers et al., 1993). Myers et al. (1993) therefore used tests simulating complex activities such as sweeping. Harding et al. (1994) suggest those activities that are the most relevant to everyday activity performance should be selected for performance measures to maximise ecological validity. Assessments are task-specific and performance on one test will not necessarily be possible to extrapolate from performance on another (Harding et al., 1994).

Walker et al. (1997) assert that activity is determined by 3 factors: what an individual is able, needs and wishes to do. Whilst performance measures may estimate how well an individual can perform the selected tasks, they cannot assess the influence of the remaining factors. Ferrer et al. (1999) observed that self-report scales might be closer than performance measures to the disability concept as defined earlier. Questionnaires allow activity to be assessed in its sociocultural context whereas performance measures reflect something more like 'objective functional limitation'.

Compared with self-report measures, performance tests are generally more time consuming, require more space and equipment and may have the danger of injury (Guralnik et al., 1989). Additionally, bias may arise through the observer's behaviour. It has been shown, for example, that encouragement from the researcher influences performance on a walking task (Guyatt et al., 1984). The test situation itself could have a reactive effect on the participant's performance (see Harding et al., 1994).

The observer may also record data in a biased manner. Whilst many performance tests simply demand that the researcher notes, for example, a time or distance, others require participants to be rated according to difficulty in performing tasks or pain displayed. Observers rate behaviours in the context of their own experiences and expectations. Prkachin, Berzins, and Mercer (1994) took videotapes of shoulder pain patients making active and passive movements. It was found that specific facial actions as assessed using FACS were related to self-report measures for more tests than were observer ratings. The observed and FACS scores correlated only for passive movements, which are associated with higher pain levels than are active tests. Observers significantly underestimated pain. Hence observer reports can not be assumed to be more accurate than patient self-reports.

Validation Problems.

The construct validity of some measures of activity performance has been assessed using correlations with scores on ADL or IADL self-report measures (see Myers et al., 1993 for review). As Myers et al. (1993) point out, this is a questionable process when measures vary in content, scoring methods and rating format, and results can be biased when one researcher presents both measures. Also, given the various advantages and disadvantages of the two measure types, performance measures seem likely to access different aspects of activity from self-report measures and so it is doubtful whether 2 such measures should be expected to correlate very highly. Indeed, if a performance measure does correlate highly with a self-report measure, perhaps it is redundant given the extra costs it terms of time, space and equipment.

v. Self-report and Observed Performance Measures: Comparisons

Some researchers have reported good agreement between self-report and performance measures. For example, Hakala et al. (1994) reported rheumatoid arthritis patients' physical function score using the Arthritis Impact Measurement Scales (self-reported functional ability) to be correlated well with score on the Keitel Function Test, an observed performance measure of impairment ($r = 0.74$, $p < 0.0001$). Jette and Branch (1985) found impairment score and self-reported physical disability to be strongly associated. Peck et al. (1989) found patient-reported difficulty with ADLs to be highly correlated with an observer's assessment of the same items: $r = .89$, $p <.001$.

In contrast, Fisher and Johnston (1996b) found that the self-rated disability of chronic pain patients (the lifting section of the Oswestry Low Back Pain Disability Questionnaire (ODQ)) did not correlate with a lifting task either before or after a cognitive manipulation when time for which bags of rice were held was used. However, when the weight participants chose to hold was compared with the ODQ, a significant correlation was seen ($r = -.54$, $p < .01$). The ODQ refers to weight rather than the time of holding; participants could have been choosing weights they were confident of coping with or to be consistent with the self-report. Correlations of a similar order were seen by Spiegel, Leake, Spiegel, Paulus, Kane, Ward and Ware (1988). Time spent to walk 50m correlated significantly with self-reported scales assessing lower extremity function ($r = -.53$ to $-.68$).

Despite using identical tasks both for performance and self-report ratings, Rejeski, Craven, Ettinger, McFarlane and Shumaker (1996) found some variance to be shared between measure types but much remained unaccounted for. Observed performance scores were compared with participants' perceived difficulty and ability immediately after the tasks had been completed. For example, for the stair time task, observed time gave a correlation of $r = -.63$ with perceived difficulty and -.48 with perceived physical ability. Also comparing activity measures on identical tasks, Klesges et al. (1990) observed participants' non-structured activity in a public track and fitness area. Whilst ignorant of their having been observed, participants were asked to report activity engaged in on the same scale as the observers. Inter-measure correlations ranged from $r = .47$ ($p < .01$) for 'stationary with limb movement' time to $r = .76$ ($p < .001$) for 'stationary' time. Moreover, participants showed consistent biases in underestimating time spent in sedentary activity and overestimating time spent performing aerobic activities. In an earlier study by Klesges, Klesges, Swenson and Pheley (1985), student participants were observed for one hour, having been instructed to engage in whatever activity they desired. Participants were aware that they were being observed. Pearson correlations between observed and self-reported time spent in activities ranged from .21 for extreme activity and .36 for lying down to .81 for running and .88 for sitting, the mean correlation between observed and self-reported activity being .60. Again, participants underestimated time spent being relatively inactive (underestimated time sitting ($t = 2.59$, $p = .01$) and moderate activity ($t = 3.67$, $p < .001$)) and overestimated time in high activity categories (overestimated time spent running ($t = 3.93$, $p < .001$) and in extreme levels of activity ($t = 4.74$, $p < .001$)).

Merrill et al. (1997) compared older adults' self-report disability and performance scores for three tasks chosen for their comparability across measures (for example, reported ability to walk across a room was compared with time to walk 8 feet and back). It was found that, for the majority of participants, self-reported and performance measures matched accurately. Ferrer et al (1999) also found agreement between self-reported and performance difficulties, using similar performance measures to Merrill et al (1997) (chair raises and 4m walk), with an elderly Spanish sample. In both studies, specificity values (ranging from .71 to .94, Merrill et al. 1997 and .83 to .98, Ferrer et al. 1999) were higher than sensitivity values (.15 to .82, Merrill et al. 1997 and .58 to .63, Ferrer et al. 1999). (ie false positives: reporting themselves to have a disability where none was seen in the performance task occurred less than 'false negatives' – where individuals reporting no disability did have difficulties when observed). Both studies took the performance measure as the 'true' score, which is arguable, as seen earlier. As Ferrer et al. (1999) suggest, it is possible that individuals can find adaptations to tasks in their own environments, and limitations may not be perceived as disabilities if the task observed is not important for the individual in daily life. The agreement between performance and self-report measures was found to be moderate by Ferrer et al. (1999) (kappa values: .41, .55, .55). It is conceivable that someone who has difficulty rising from a chair 5 times may report no difficulty when asked about difficulty 'standing up and sitting down from a chair' (Ferrer et al., 1999) or self-reported 'difficulty stooping, crouching, or kneeling' (Merrill et al., 1997) as the task is different to the self-report question. It is likely that, the closer a self-report item matches an observed task, the better the agreement. This is supported by Myers et al. (1993), who found better agreement between performed tests and self-reports to be best when the wording of IADL items closely matched the test. This is a difficulty because, ideally, disability researchers generally wish to extrapolate from the performance or self-report measure to the participant's experiences in the wider world but self-report measures going beyond the limited range of performance tests can not easily be validated using such tests.

Myers et al. (1993) compared self-reports and performance assessments directly in 14 tasks of functional abilities with community-dwelling older adults. Good correspondence (>80% agreement) was found between observed and perceived difficulty for only a third of the tasks. In general, the observer underestimated difficulty compared with the self-assessment. In contrast, Ferrer et al. (1999) and Merrill et al. (1997) found participants reported not finding problems in tasks where observers reported difficulties. Perhaps the nature of the tasks, being instrumental activities of daily living such as sweeping the floor and writing is a factor in this: Myers et al. (1993) used complex activities requiring combined actions whereas Ferrer et al. and Merrill et al. used simpler, more isolated actions. Also, Ferrer et al. and Merrill et al. devised observed difficulty scales from measures such as time to perform a task whereas Myers et al. (1993) had observers estimate difficulty as a separate assessment from time to perform tasks.

Hence the evidence as to the comparability of self-report and performance measures is contradictory. Five of the studies looked at the direction of differences between self-report and observed measures; in three of these (Klesges et al. 1990, Klesges et al. 1985, Merrill et al. 1997 and Ferrer et al. 1999) self-report measures 'overestimated' activity compared with observed scores, only Myers et al. (1993) found participants to 'underestimate' activity compared with the observed measure. This finding is similar to comparisons of self- and proxy-reported measures in that self-reports yielded more optimistic results than proxy-reported measure.

With many studies reporting correlations that are significant but leaving much variance unaccounted for, it does seem likely that performance and self-report measures are often asking different questions. However, it is difficult to ascertain this without an objective means of assessing activity outside of the laboratory. A range of automated activity monitors have been devised with the potential to perform this function.

vi. Automated Measures

A range of automated activity monitors have been developed. Five groups of monitors are outlined below: electromyographic (EMG) measurement, devices using accelerometers, tilt-switch mechanisms, pedometers and combination devices.

1. EMG

Anastasiades and Johnston (1990) used thigh EMG. The records were found to covary closely with heart rate and to be a better predictor of heart rate than knowledge of the physical activity performed (running, walking and non-movement). However, it did not allow posture (sitting or standing) to be discriminated. It seems likely that such a measure will be of greater benefit with active, able populations than with more limited individuals: a floor effect could be found in a disabled population as amount of time spent standing as opposed to sitting or lying could be more relevant than differentiation between walking and running.

2. Accelerometers

The second group of monitoring devices is the accelerometers, containing an electrical transducer that detects changes of movement (Patterson, Krantz, Montgomery, Deuster, Hedges, & Nebel,1993). The Caltrac accelerometer is one such device, worn on the hip and recording vertical accelerations (see Levin et al., 1999). Accelerations and decelerations are summed over the relevant time periods and records can be converted to estimate total energy expenditure (in kilocalories), using metabolic rate estimated from the participant's height, weight, sex and age (Rennie & Wareham, 1998). However, whilst such a score is useful if individual fitness or energy expenditure is being assessed, in studies interested in activity or motion, dividing the kilocalorie score by the resting metabolic weight estimate may give a more useful score, in METs, especially as, with total energy expenditure, much depends on body size (Jacobs et al., 1993). A difficulty with such MET scores is that individual variation in resting metabolic rate is not taken into account. The 'Actigraph' also uses an accelerometer but differs from Caltrac in that reported scores reflect changes in frequency and intensity of activity throughout the recording period, rather than giving a summed MET score (Patterson et al., 1993). However, the monitor was attached to the wrist and so activities requiring high wrist movement, such as video games, would bias the activity score.

Unfortunately, as accelerometers require movement, different postures will not be detected. Tuomisto, Johnston and Schmidt (1996) compared accelerometric, electromyographic and hydrostatic posture measurements. Neither the accelerometer nor EMG reflected different postures well, whereas the hydrostatic measure differentiated postures effectively. The hydrostatic measure involved a plastic tube being placed along the body, from the ankle to the chest.

Tuomisto et al. (1996) found EMG to be more sensitive than accelerometry in differentiating between the tasks standing still, sitting, straight leg raising, lying, cycling and walking. However, they were similarly able to predict heart rate. Accelerometers can be fragile, but are easier to use than is EMG (Tuomisto et al., 1996).

It is difficult to recognise physical activities from Caltrac output and it does not relate to the demands of activities such as hill or stair climbing or cycling and arm movements are not represented (Levin et al., 1999). Tritrac, an accelerometer device using 3D movement may overcome the problems of stair-climbing and cycling as it does not rely only on vertical acceleration (see Rennie & Wareham, 1998). Both Caltrac and Tritrac are influenced by passive activities such as riding in a vehicle (Walker et al., 1997). Levin, Jacobs, Ainsworth, Richardson, and Leon (1999) found that 6 repeated measures with Caltrac were necessary to obtain reliable measurement. Swimming and showering can not be recorded as Caltrac is not waterproof (Rennie & Wareham, 1998).

3. Mercury Tilt Switches

The third group of monitors are those using mercury tilt switches to detect body posture, such as the leg mounted 'Up-timer' used by Sanders (1983) which measures the time spent standing or walking. Tran, Schwarz, Gorman, and Helme (1997) also used an up-timer, attaching the device with a waterproof dressing so that it did not interfere with showering, thus overcoming one problem found with other monitors. Tran et al. (1997) found the test-retest reliability (the monitor was worn on the same day 1 week later) to be 0.84. The large-scale integrated motor activity monitor (LSI) also uses a mercury switch (see Patterson et al., 1993). It was found that correlations between self-reported physical activity and LSI were only modest. The mercury switch was found to be insensitive to very low and very high activity levels (Klesges et al., 1985).

4. Pedometers

Pedometers, such as the electronic Yamax DW-500 used by Bassett, Cureton and Ainsworth (2000) perform a step count and estimates of distance walked can be acquired by multiplying step count by step length. Devices can also be small and lightweight, easily attached to a belt or waistband and so cause little inconvenience to participants.

5. Combination Devices

A couple of devices have been designed which combine features of other monitor types. Vitalog (Taylor, Coffey, Berra, Iaffaldano, Casey & Haskell, 1984, see also Haskell, Yee, Evans, & Irby, 1993) has both a mercury tilt switch attached to the thigh detecting leg/body movement and 3 chest electrodes assessing heart rate. It was found to correlate well with a 7-day activity self-report.

The Numact activity monitor combines tilt switches attached to the thigh and chest with an accelerometer also placed in the chest sensor. Output shows posture (lying, sitting, standing) detected by the tilt switches and the number and vigour of steps taken over the recording period (maximum 72 hours) recorded via the accelerometer. This device will be discussed in more detail in the Pilot Study (Chapter 2), as it is the monitor used here.

Conclusion

Hence automated activity monitors are somewhat imperfect: they do not record activity type (Patterson et al., 1993) and effectiveness of recording activity level varies between monitors and depends on the activity being performed. They can also restrict activity as participants must usually be asked to not swim or shower.

However, they do provide an objective means of assessing activity outside of the laboratory. Whilst questionnaires may lead to differentially representing the activity of men and women as women more frequently perform childcare and household activities which are difficult to report (Kriska & Caspersen, 1997), monitors will detect such activity, and other non-discrete, unplanned activities which may be difficult to recall.

Results may be affected by social desirability bias similarly to performance measures, but it is likely that extra effort is harder to sustain over the longer time periods used. Activity monitoring technology is likely to be important both for the validation of existing performance and self-report measures and in assessing factors with possibly differential influences on different measurement methods. For example, negative affectivity, social desirability, perceived control and distress could influence activity assessments differently, depending on the methodology type chosen.

vii. Self-report and Automated Measures: Comparisons

Williams, Klesges, Hanson and Eck (1989) compared Caltrac score with two self-report activity scales. Whilst the self-report scales correlated and were highly reliable, they were minimally related to the Caltrac score and Caltrac had poor test-retest reliability.

Jacobs et al. (1993) found that, out of 10 physical activity questionnaires used, none related to energy expenditure as assessed using the Caltrac monitor. However, several measures of heavy intensity physical activity were associated with the Caltrac motion score calculated by dividing the caloric expenditure by resting metabolic rate. That is, the MET score corresponded more closely with self-reported activity than the energy score, which largely depends on body weight. Nevertheless, a significant, if fairly low, correlation ($r = .35$) was seen between Caltrac records and 7-day diary records by Dishman, Darracott and Lambert (1992).

The correlation between monitor and self-report may depend on the activity being assessed. Bassett et al. (2000) found participants underestimated daily walking distance. Pedometer readings for 7 consecutive days were compared with self-reported walking distance from the Paffenbarger Physical Activity questionnaire (the College Alumnus questionnaire), which asks participants to report how far they walk each day. In contrast, Matthews and Freedson (1995) found Tritrac gave a lower estimate of daily energy expenditure than self-report measures but a higher estimate of time spent in sedentary activities in comparing 7 days of monitored activity with 7-day recall of that activity.

Automated up-time was compared with hourly self-monitoring over 4 consecutive days by Sanders (1983). Three participant groups were compared: 6 low back pain inpatients, 6 psychiatric inpatients with anxiety and depression and 6 hospital staff. It was found that all groups reported less up-time than the monitors suggested. The discrepancy was greatest with low-back pain patients. This is interesting with respect to the proposal that individuals with negative affectivity would be expected to view their performance as less than others and so under-report because it would be expected that anxiety and depression patients would show the greatest negative affectivity.

In addition to the relationship between monitor records and questionnaire scores depending on how the monitor activity is assessed (ie METs or energy output), the type of questionnaire used may also affect the association. As mentioned earlier, Levin et al. (1999) used 2 self-report questionnaires and Caltrac accelerometers to assess seasonal variation in activity. Whilst accounting for the month of the year reduced intra-individual variance by 7% using a 4-week activity survey, variance reduced by only 4.2% in a self-report measure taken over the 48 hours for which the monitor was worn (generally every 4 hours) and 5.2% for Caltrac records. As well as requiring participants to recall activity over a longer time period, the 4-week history was the only measure not to include activities of daily living. Thus, in order to have self-report and automated measures match closely, it is necessary to ask questions that allow people to give the appropriate activity information and using the same time period. This is supported by Patterson et al. (1993) who compared Actigraph output with self-report diary records. The average correlation between Actigraph readings and average activity was .57, ranging from .29 to .86 for individual participants. Lower correlations were found among participants who made fewer and less complete diary entries. Miller, Freedson and Kline (1994) compared Caltrac readings with five questionnaires (7-day recall, Baecke, NASA, Godin and 3-day record). Caltrac readings from a 7-day period were found to significantly correlate with two of these: 7-day recall ($r = .79$) and the Godin ($r = .45$). However, as the sample was quite small ($n = 33$), it is possible that other questionnaires could give significant, if smaller, correlations. Similarly, Pols, Peeters, Kemper and Collette (1996) found only a low correlation between a one-day Caltrac reading and a modified Baecke questionnaire (r = .22) but when compared with diary records taken over the same 24-hour period, the correlation improved, if only to r = .45. Conn, Minor, Mehr and Burks (2000) found the Physical Activity Index, a diary-based measure to correlate with TriTrac readings (r = .49) whilst the Baecke physical activity scale and the Physical Activity Scale for the Elderly correlated at lower levels: r = .37 and .24 respectively. Unfortunately significance levels do not appear to be indicated.

Hence it would appear that correlations between monitors and self-report measures tend to be variable and no better than moderate; in general, higher correlations are seen where activity is reported in detail (ie using activity diaries) over the same period of time for which the monitor is worn. One impressive exception to this is the finding by Voorrips, Ravelli, Dongelmans, Deurenberg and Van Staveren (1991) who were seeking to validate a physical activity questionnaire for the elderly with a reference time of 1 year, arguing that, as this population has weaker short-term memory, the reference frame should not be too short. The correlation between this measure and a pedometer worn for 3 consecutive days was r = .72.

viii. Self-report, Observed Performance and Automated Measures: Comparisons.

Up-timer records were compared with both self-report and performance measures by Tran et al. (1997). Elderly nursing home residents were given a range of self-report measures, as well as the Timed Up and Go test. This performance test times participants to rise from an armchair, walk 3 metres, walk back and sit down (Podsiadlo & Richardson, 1991). Moderate levels of correlation (-.47 to .55) were found between the up-timer and all the other measures, whereas high levels of association were found among the self-report measures (.78 to -.92) and the associations between the performance and self-report measures were intermediate (.59 to -.70). It is interesting that the correlation between up-timer and performance scores was -.51, similar to that with self-report measures, given that both up-time and performance measures are considered to be objective. Tran et al. (1997) suggested that this could be a result of people feeling a need to perform both in the self-report and performance tests.

Walker, Heslop, Kay, and Chandler (1998) used the Numact activity monitor with patients suffering osteoarthritis of the knee. One patient sample, suitable for an anti-inflammatory medication trial, responded to the Osteoarthritis Severity Index (OSI), a structured disability questionnaire asking about symptoms, activities and rated pain. Both patients and physicians completed 5-point scales on the patients' conditions. A second sample, awaiting knee replacement, completed the Nottingham Health Profile (NHP). Both samples wore the monitor for 24 hours and X-ray and joint space data were collected. Whilst correlations between Numact output, patient opinion of condition, pain scale and OSI were non-significant, a correlation of 0.4 was found with the physician's opinion of patient condition for the first population. In the second population, Numact did not correlate with the mobility, pain, energy, emotion or social scales of the NHP; the best correlation was found with the scale assessing poor sleep ($r = .33$). Correlations of Numact with x-ray grade and medial joint space were significant ($r = .45$, $r = .3$ respectively). Numact was significantly associated with objective impairment measures rather than assessments of activity limitations. Walker et al. (1998) suggested that patients could assess themselves with reference to how they usually are, and have adapted to their condition; changes in ability could be more likely to be detected with questionnaires. In contrast, the physician would have a wider experience of osteoarthritis, and so be able to make comparative judgements between patients. However, whilst this would explain patient pain ratings not corresponding to Numact output, it is not clear how ratings of activity and mobility would be affected. Also, as discussed earlier, proxy reporters can not necessarily be taken as the objective standard; a more objective performance comparison measure could have been used.

Steele, Holt, Belza, Ferris, Lakshminaryan and Buchner (Steele et al., 2000) compared the Tritrac monitor with repeated 6-minute walks and self-reported activity over a 3-day period. For the 3 walking tests, accelerometer records during the test time and distance walked correlated highly: $r = .84$, .85 and .95 for the 1st, 2nd and 3rd walks respectively. Accelerometer records over a continuous free-living 3-day period correlated with 'exercise capacity' (the furthest of the 3 distances walked in 6 minutes) at a very reasonable level: $r = .74$ ($p <. 001$). In contrast, the correlation between this accelerometer score and self-reported activity over the 3 days was non-significant ($r = .14$). These results could be due to both accelerometer and walking test detecting activity in a limited domain: walking, whereas the self-report measure requested the recall of all physical activity undertaken during this time period.

Thus existing evidence suggests that self-report, performance and automated measures of performance are not necessarily highly related. The dilemma faced by questionnaire designers, whether to ask participants about activity they are *able* to perform or activity that *actually is* performed is resolved differently by automated and performance measures. Performance measures assess what participants are observed to do in a laboratory situation and automated activity monitors record activity performed in the 'real world'. However, there is a shortage of evidence examining all three measurement methods in comparable tests. Tran et al. (1997) compared up-time with general disability measures and Walker et al. (1998) used non-self-report measures other than an objective performance test with which to compare Numact activity results.

Measurement Summary

It has been seen that self-report, proxy-report, performance and monitor measures of activity each have benefits and disadvantages to their use. Comparisons of self-report and performance measures have yielded contradictory results. This could depend, to some extent, on the particular measures used in studies and how comparable the activities assessed are. However, this is difficult to ascertain without an objective measure of use beyond the laboratory's confines. Studies comparing self-report and performance data with monitor results have, so far, provided limited evidence as to the relationships between the three measurement types. It would appear that the more similar the activity assessed across measures, the closer the correlations between the measures, but limited work has been done using such comparable measures.

Part 2: Prediction of Activity Limitations

ix. Emotion

The role of emotional state in the prediction of activity limitations, amongst other health outcomes, has been investigated. The emotions commonly researched are anxiety, depression and more general distress measures.

Self-Report Measures: Predictive Studies

Mood has predicted self-reported disability with ALS/MND patients (Johnston, Earll, Giles, McClenahan, Stevens, & Morrison, 1999) and Johnston, Morrison, MacWalter and Partridge (1999) found anxiety 1 month after hospital discharge to be correlated with self-reported recovery in stroke patients, although when entered into a multiple regression analysis with perceived control and gender as independent variables, anxiety did not maintain significance.

Bruce, Seeman, Merrill and Blazer (1994) found depressive symptoms to be associated with an increased risk of onset ADL disability two and a half years later in a large cohort of 1058 high-functioning men and women aged 70-79 years. Baseline measures including physical health (self-report and physical examination) and cognitive functioning were taken at time 1 and, where approaching significance ($p < .10$), these variables were included in the model to control for their effects. It was found that the effects were enhanced when these baseline factors were controlled, giving significant odds ratios for both men (3.55) and women (5.47). The outcome measure of activity was a self-report 7-item ADL scale (as selection criteria required all participants to be free of ADL disability, any ADL disability reported at time 2 could be considered onset disability). As mood at time 1 was associated with the onset of disability rather than with disability at time 1 it would seem that more is happening here than a simple bias of negative mood influencing reporting. However, alternative measures of activity would throw light on whether participants were doing less, reporting less, or were prevented by physical impairment of performing the same ADLs at time 2 as at time 1.

Observed Performance Measures: Predictive Studies

Kaivanto, Estlander, Moneta and Vanharanta (1995) found depression to be non-significant in the prediction of isokinetic performance of patients with chronic back pain.

Self-Report and Observed Performance Measures: Predictive Studies

Härkäpää, Järvikoski, Mellin, Hurri and Luomo (1991) found increased distress to be associated with the poorer performance of low-back pain exercises. Distress was not found to predict self-reported disability or frequency of exercise.

Johnston, Morrison et al. (1999) found neither stroke patients' anxiety nor depression at 1 month post discharge to predict observed disability recovery at 6 months, but anxiety did predict self-report (Barthel Index) recovery ($r = -.29$, $p < .05$). However, when entered into a multiple regression equation with perceived control (Recovery Locus of Control measure) and gender as independent variables, only perceived control entered the equation. Also with stroke patients, Bonetti, Johnston, Morrison, MacWalter, and Pollard (2001) found anxiety and depression at both within 2 weeks of hospital discharge and 6 weeks later to predict 6-month self-reported (Modified FLP) recovery (time 1: anxiety: $r = -.15$, $p < .05$; depression: $r = -.23$, $p<.001$; time 2: anxiety: $r = -.16$, $p < .05$; depression: $r = -.27$, $p = .001$). Correlations with self-reported Barthel Index or observed performance measures were not significant.

Interventions

Manipulations of mood were found to result in short-term effects on physical activity performance of patients with chronic low back pain (Fisher & Johnston, 1996a). During a normal clinical interview, participants were asked to talk about either upsetting events (increase anxiety condition) or good events (decrease anxiety condition) before performing a lifting task. The manipulation was found to successfully alter anxiety in the predicted directions.

Borson, McDonald, Gayle, Deffebach, Lakshminarayan, and VanTuinen, (1992) performed a 12-week randomised control trial of nortryptiline with 30 chronic obstructive pulmonary disease (COPD) patients with comorbid depression. Not only was the antidepressant superior to the placebo in treating depression, but the nortryptiline condition also showed improvements in self-reported physical symptoms and functional health status (illness-related activity and behaviour). However, no improvement was seen using the performance measure of exercise tolerance (a 12-minute walk). It should also be noted that many of the items on the self-report physical symptoms checklist, such as shortness of breath and hyperventilation, are found in patients with depressive or anxiety disorders as well as in patients with COPD. The authors suggest that the severity of the patients' physiological impairment was such that mood was unable to have a significant independent contribution. It would appear that change in health outcome was only reliably found with the self-reported illness-related activity measures. The lack of change in performance on the 12-minute walk suggests that physical ability may not have improved, although changes in daily activities performed, as assessed by the ADL-type items of the self-report scales were seen. It could be that patients are more active when their mood is more positive. On the other hand, it could also be that, with improved mood, people perceive and report themselves to have been more active even if the actual activity levels have not changed. The 'negative-affectivity bias' hypothesis would also predict changes in breathing difficulty to be reported, however, and such a change was not seen.

A second suggestion as to how mood might affect physical activity is found in a paper by Zelman, Howland, Nichols and Cleeland (1991). Students and hospital workers performed cold pressor trials before and after undergoing a randomly assigned Velten mood induction procedure to induce depressive, neutral or elative mood. Participants in the depressive condition decreased their tolerance, the time for which they kept their arm immersed in the cold water, by an average of 10.6s whilst neutral and elative participants increased tolerance time by 2.1

and 10.4s respectively. The difference between elative and depressive groups was significant. It would not appear that participants were merely responding to demand effects as self-reported pain ratings were not significantly different across groups. It could be, therefore, that participants with better moods are more likely to persevere in physical activities.

Hence emotional states have been found to predict physical activity in some studies. However, results are not consistent and it is not clear to what extent associations are dependent on measure type. Emotion is also of interest as depression and anxiety are components of the more general negative affectivity construct which has been proposed as a potential source of bias in self-report measures (see Watson & Pennebaker, 1989).

x. Control Cognitions

According to the Theory of Planned Behaviour (Ajzen, 1985), perceived behavioural control, an individual's perception of the ease or difficulty of performing a behaviour (Ajzen, 1991), is an important determinant in predicting behaviour (see Figure1.1).

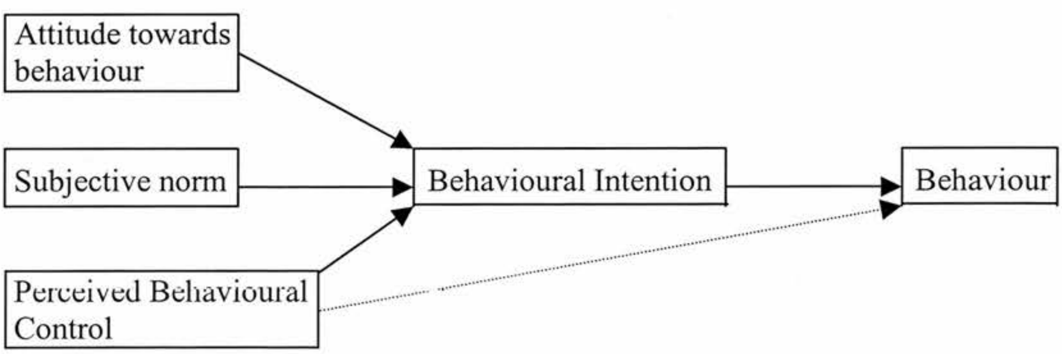


Figure 1.1: Theory of Planned Behaviour (adapted from Conner and Norman, 1996).

A similar construct is Bandura's (1977) 'self-efficacy', the 'conviction that one can successfully execute the behaviour required to produce the outcomes'. Bandura proposes that individuals will choose to do tasks that they believe they are able to perform and, the stronger their perceived self-efficacy, the more persistent their efforts will be (Bandura, 1986).

There is some debate as to just how similar these constructs are. Ajzen (1991) suggested that the 'perceived behavioural control' construct is compatible with Bandura's 'self-efficacy', with the Theory of Planned Behaviour placing the self-efficacy construct within a wider framework. Some research has been done incorporating the two models, for example, Courneya and McAuley (1994) used the Theory of Planned Behaviour but substituted perceived behavioural control with self-efficacy. However, Conner and Norman (1996) argue that studies provide evidence for distinguishing between the two over a wide range of behaviours. Nevertheless, as some behaviours have been predicted by self-efficacy and others by perceived behavioural control, both of these control constructs will be reviewed here in their ability to predict physical activity. These are also not the only control constructs operationalised in the literature in the prediction of activity; mastery, helplessness and locus of control are also seen. The main issue of interest here is whether predictive findings have been found across measurement types.

Self-Report Measures: Predictive Studies.

The relationships of perceived control and self-efficacy with self-reported disability in patients with rheumatoid arthritis (RA) were examined by Schiaffino and Revenson (1992). In an apparently cross-sectional analysis (few procedural details are given), self-efficacy was found to correlate with disability ($r = -.37$, $p < .05$); perceived control (the perceived controllability of RA) correlated with disability at a trend level of significance ($r = -.31$, $p < .1$). Self-efficacy was found to relate to disability when perceived control was partialled out, but, when self-efficacy was controlled for, the association between perceived control and disability became insignificant. Longitudinally, using disability assessed 4 months later as the outcome variable, a similar pattern was observed, but the trend-level relationship between perceived control and disability became non-significant. It does not appear that earlier disability was controlled for in this second analysis. It is not clear from this study whether self-efficacy is a stronger predictor than perceived control because, whilst perceived control was assessed with 2 general-level items ('RA is controllable' and 'RA is controllable by one's self', self-efficacy was assessed with 3 more specific items measuring perceived ability to manage pain, to deal with physical limitations and to perform daily activities. Bandura (1986) suggests that discrepant results may arise where different levels of generality are used between independent and dependent variables; an inappropriate level of assessment for the outcome measure could, therefore, cause the lack of effect with the perceived control measure.

Schiaffino, Revenson and Gibofsky (1991) found self-efficacy over ability to manage pain, deal with physical limitations and to continue daily activities despite rheumatoid arthritis to predict lower functional disability 14 months later in patients who had recently been diagnosed with rheumatoid arthritis (within 2 years of the study's start). The outcome measure was the self-reported Arthritis Impact Measurement Scale – a 21-item functional disability scale including dexterity, mobility, physical activity and household activity components. The level of self-efficacy questions seems to be reasonably specific to that of outcome measurement.

A relationship between self-efficacy and activity was found even with a general-level measure of self-efficacy by Schwarzer and Schröder (1997). It was found that, among patients going for heart surgery, pre-surgery general self-efficacy levels predicted self-reported physical activity: ambulatory activity in the hospital ward ($r = .16$, $p < .05$) and early attempts to stay physically fit ($r = .22$, $p < .01$).

Lorish, Abraham, Austin, Bradley, and Alarcón (1991) followed up 155 patients with rheumatoid arthritis over a 12-month period. Helplessness, patients' beliefs in their abilities to exert some control over their symptoms, was found to be related to self-reported physical functioning both cross-sectionally at baseline and longitudinally, at 12 months. In a stepwise multiple regression equation, helplessness was found to be a significant predictor of physical functioning accounting for .11 R^2 change after disease severity had been entered.

A predictive effect of self-efficacy on activity in community-residing elderly people was found by Mendes de Leon, Seeman, Baker, Richardson and Tinetti (1996). Self-efficacy regarding confidence to perform activities of daily living without falling predicted self-reported help required with activities of daily living 18 months later at trend level after controlling for baseline variables including ADL score ($b = .072$, $p < .10$). Orbell, Johnston, Rowley, Davey and Espley (2001) found self-efficacy at 3 months post hip or knee joint replacement surgery to predict disability at 9 months after controlling disability pre-surgery and at 3 months. Orbell et al.'s (2001) self-efficacy measure was designed to match activities reported on in the FLP, the self-report activity limitations measure used.

Kempen, van Sonderen and Ormel (1999) found both 'mastery' (the extent to which life changes are regarded as being under one's control) and general self efficacy expectations at baseline to be associated with change in self-reported disability over a 2 year period in low-functioning community-living older people ($r = -.091$; $r = -.107$; $p < .05$ for mastery and self-efficacy respectively). Participants who had reported high levels of mastery or general self-efficacy expectancies at baseline reported a significantly lower disability increase than participants who reported medium or low levels of mastery or self-efficacy. On repeating the analysis, controlling for each of mastery and self-efficacy expectations as covariates, it was found that mastery's effects were still significant ($p < .01$) whereas the effects of general self-efficacy expectancies were not, suggesting that mastery was the stronger predictor. It is unfortunate that little information is provided on the scales' content and so it is not clear whether the mastery and general self-efficacy measures differed in to what extent they matched the activity level of the disability measure.

In contrast to the studies above, Courneya and McAuley (1994) researched an undergraduate student population. Self-reports of physical activity were obtained 2 and 4 weeks after an initial questionnaire including self-efficacy measures had been completed. Self-efficacy was found to predict frequency and intensity but not duration of physical activity. Courneya and McAuley (1994) argue that physical activity frequency and intensity are less controllable than duration; as duration is more controllable, a predictive effect was not seen. However, it seems that it is exercise, not physical activity according to Caspersen et al's (1985) definition that is being assessed here.

Results are not completely consistent, however. Gill, Robison and Tinetti (1997) found recovery of ADLs in community-living elders to be predicted by self-efficacy in bivariate analysis, but in multivariate analysis, where age, cognitive status, mobility and nutritional status contributed significantly, self-efficacy was no longer significant. Frank, Johnston, Morrison, Pollard, and MacWalter (2000) found change (regression residuals) from baseline to 1 month in perceived control (as assessed with the Recovery Locus of Control (RLOC)) to be associated with change in functional limitations (Modified FLP). However, whilst the correlation of RLOC with both the total Modified FLP score and the psychosocial dimension score were significant, the correlation with the FLP physical dimension was not. It is possible that, within this population of patients who had had a stroke up to 2 years ago (average time since stroke: 40 weeks), insufficient change in physical functioning occurred over the one-month period to produce significant correlations.

However, the majority of studies reviewed suggest that control cognitions fairly consistently predict physical activity outcomes assessed by self-report.

Self-Report Measures: Intervention Studies

Langer (1983) randomised nursing home patients to experimental (responsibility-inducing) and non-experimental groups according the floor on which they resided. The experimental condition emphasized participants' responsibility and control over their living environment. They were given the opportunity to choose a plant, for which they would care themselves. In contrast, the non-experimental participants were given a plant for which the nurses would care. Questionnaires assessed self-reported activity both one week before and 3 weeks after the intervention. The experimental group reported themselves to be more active on the second interview than at the first ($t\ (43) = 2.67$, $p < .01$). However, the mechanism by which this change occurred is not clear: no accompanying significant change in perceived control was found but the experimental group did report greater increases in happiness after the intervention than did the comparison group. Nevertheless, as over 20% of participants indicated that they did not understand what was meant by 'control', the lack of a perceived-control effect could reflect comprehension of the questionnaire rather than lack of effect.

A self-management programme designed to enhance self-efficacy beliefs was offered to people with arthritis by Barlow, Williams, and Wright (1999). Eighty-nine participants completed questionnaires both before the programme and at four months after the baseline measure. It was found that self-efficacy beliefs increased but, although pain decreased, no change was seen in physical functioning. As most forms of arthritis are progressive, the lack of change in physical functioning was considered to be a positive result by the researchers. It is unfortunate that no control group was included to determine whether, without the intervention, a decrease in physical functioning would indeed have been seen over the four months.

O'Leary, Shoor, Lorig and Holman (1988) allocated rheumatoid arthritis patients to either a treatment including training in cognitive and behavioural pain management strategy or a control condition. A significant reduction in pain was seen in the treatment group and perceived self-efficacy to manage pain and other effects of rheumatoid arthritis were enhanced by the treatment. Whilst the treatment did not affect self-reported disability levels (the authors suggest this could be because participants were already quite active), changes in self-efficacy were related to changes in disability in the treatment group.

Observed Performance Measures: Predictive Studies

In two studies, the effects of self-efficacy on performance in back-pain patients were investigated. Lackner, Carosella and Feuerstein (1996) had patients complete the Functional Self-Efficacy Scale (FSE), reporting their confidence in performing 33 physical requirements of work. Several days later, a physical examination was conducted involving lifting, carrying, pushing and pulling. Partialling out anticipated re-injury and pain, participants with greater FSE showed better performance, with correlations ranging from .38 (lifting from waist to eye level) to .65 (static pulling, $p < .001$ for all measures). FSE also contributed significantly to performance on all tasks when entered into multiple regression equations alongside gender and pain. A later study incorporating a similar design replicated the finding that higher FSE scores were associated with greater lifting performance. Lackner and Carosella (1999) had 100 low back pain patients complete a questionnaire including items assessing self-efficacy expectancies of functional tasks (physical requirements of work) and perceived pain control several days before physical and functional capacity was evaluated in a physical exam including lifting tasks conducted by either an occupational health physician or a physical therapist. Functional self-efficacy was associated with lifting performance such that higher self-efficacy was associated with higher lifting performance.

With another patient group, wrist fracture patients, Partridge and Johnston (1989) found perceived control, assessed with the Recovery Locus of Control (RLOC) to predict the performance of wrist movements. The RLOC also predicted stroke patient recovery, as assessed by the performance of body movements (Partridge and Johnston, 1989).

Observed Performance Measures: Intervention Studies

A perceived control manipulation was found to have short-term effects on a lifting task performance with back pain patients (Fisher & Johnston, 1996b). The manipulation, which affected perceived control in the predicted directions, consisted of the participants telling the researcher about times when they either felt in control of situations (increase perceived control) or when they felt out of control (decrease perceived control).

Self-Report and Observed Performance / Monitor Measures: Predictive Studies

Rejeski, Miller, Foy, Messier and Rapp (2001) examined the role of self-efficacy in predicting activity as assessed by both self-report and observed performance measures 30 months later in over 300 older adults with knee pain. Self-reported difficulties with activities of daily living and performance of a stair-climbing task (time to ascend and descend a set of 5 stairs) were assessed both at baseline and at 30 months. The self-efficacy measure was task-specific: after a practice stair-climbing trial, participants were asked to report how certain they were they would be able to complete the task 2, 4, 6, 8 and 10 times without stopping. Measures of knee pain, knee strength, and x-rays were also taken. For both observed and self-reported measures of disability, knee strength, self-efficacy and an interaction of strength and self-efficacy were found to be significant predictors of disability. With both measures, self-efficacy was related to a greater decline in activity where knee strength was low.

Rejeski, Craven, Ettinger, McFarlane and Shumaker (1996) assessed the role of task-specific self-efficacy in participants with osteoarthritis before they completed a timed set of performance tasks (stair-climbing, lifting and carrying) and rated their perceptions of their performance. After controlling for maximal oxygen consumption, knee strength and pain, self-efficacy predicted both performance and self-report ability scores. For the stair-climb task, R^2 change was 7% for timed performance and 12% for self-reported physical ability. On the lift/carry task, self-efficacy added R^2 change values of 3% for timed performance and 7% for self-reported ability. Hence self-efficacy predicted physical activity with both performance and self-report measures, although the variance explained using the performance measure was about half that explained using the self-report measure.

Johnston, Morrison et al. (1999) found stroke patients' ($n = 71$) perceived control at 1 month post hospital discharge to predict disability recovery (regression residuals when disability at recruitment was regressed on 6 month scores) at 6 months on both self-report and observed measures ($r = .29$, $p < .05$; $r = .32$, $p < .05$ respectively). In a larger study (203 stroke patients), Bonetti et al (2001) also found perceived control measures (taken within 2 weeks of discharge) to predict 6 month recovery on both observed and self-report (Barthel Index) measures ($r = .26$, $r = .17$ for Recovery Self-Efficacy (RSE) and Perceived Control Index (PCI) measures, Barthel recovery; $r = .27$ and $r = .17$ respectively, observed measures. All $p < .05$). Perceived control at 6 weeks after interview 1 also predicted recovery but only on self-report (Modified FLP) disability recovery ($r = .21$, $p < .01$ for Perceived Behavioural Control over Recovery, RSE and PCI).

In contrast, Seeman, Unger, McAvay and Mendes de Leon (1999) found self-efficacy to predict self-report but not performance measures of disability. Highly functioning older adults (aged 70 – 79 years at baseline) were interviewed at two time points, the second being between 24 and 36 months after the first. Baseline self-efficacy was found to predict self-reported physical limitations with respect to strength and range of motion for both men and women and self-reported activities of daily living for men (the effect for women being in the expected direction but non-significant). However, performance, assessed with a composite measure of balance, gait, foot taps, leg strength and manual dexterity, was not predicted by self-efficacy. It could be that, to predict task-specific activity, it is necessary to use a task-specific measure of self-efficacy such as the self-efficacy over stair-climbing measure used by Rejeski et al (2001). Seeman et al (1999) were testing whether more general self-efficacy would predict activity limitations and used two measures of self-efficacy. The first, 'Instrumental Self-Efficacy' related to performing instrumental activities such as arranging transport and living arrangements; this was the measure found to have predictive effects with respect to self-reported activity limitations. The second measure, 'Interpersonal Self-Efficacy', related to managing relationships and had no predictive effects. However, Rejeski et al (2001) found an effect on more general self-reported activities of daily life when task-specific self-efficacy was assessed. It would seem to be important to assess self-efficacy relating to tasks involved in both the observed performance and self-reported measure to resolve whether it is the outcome measure of activity or the self-efficacy measure which determines whether or not an effect is seen.

Dishman, Darracott and Lambert (1992) also found only a self-report measure of physical activity to be predicted whilst the more 'objective' Caltrac accelerometer was not. Participants wore the Caltrac devices for one week and a structured daily diary of all physical activity was completed during the same week. The correlation of self-efficacy as assessed 2 weeks prior to the activity testing period was significant with the 7-day diary ($r = .46$, $p <.01$) but not with Caltrac movement counts ($r = .14$).

It would, therefore, appear that, whilst control cognitions have repeatedly predicted activity in studies with both self-report and observed performance measures of activity, in those studies combining measure types, results have not been consistent across measures.

xi. Control Cognitions or Emotion?

A link between control cognitions and emotional state has both been proposed and supported in the literature. Härkäpää, Järvikoski and Vakkari (1996) found feelings of helplessness to be associated with psychological distress in chronic low back pain patients ($r = .39$, $p <.01$). O'Leary et al. (1988) found arthritis self-efficacy to be associated with less depression in rheumatoid arthritis patients who had undergone a cognitive-behavioural treatment ($r = -.67$, $p < .001$).

The Learned Helplessness hypothesis (Abramson, Seligman, & Teasdale, 1978) proposes that learning outcomes to be uncontrollable leads to the motivational, cognitive, self-esteem and affective aspects of depression. Being unable to influence events with significant impact on a person's life may lead to despondency (Bandura, 1986). Smarr, Parker, Wright, StuckyRopp, Buckelew, Hoffman, O'Sullivan and Hewett (1997) found that, as rheumatoid arthritis patients' self-efficacy increased, depression decreased. Frank et al (2000) found perceived control to predict depression one month later in patients post-stroke ($r = -.38$, $p < .05$).

Control cognitions may also influence anxiety: those with low self-efficacy are likely to approach threatening situations more anxiously. The level of fear with which phobics perform anxiety-provoking tasks depends on their self-efficacy (Bandura, 1986). However, Frank et al. (2000) did not find perceived control to predict anxiety.

Riskind and Rholes (1985) argue that mood manipulations are mediated by cognitions and Rholes, Riskind and Lane (1987) argue that an emotion-producing event affects both affective state and cognitions. They manipulated mood state using either self-evaluative statements or statements describing somatic states that accompany mood states. Whilst both methods equally affected mood state, it was the more cognitive, self-evaluative statements that had the stronger impact on the recall latencies for memories. Such an effect could clearly influence the recall of physical activity. That it might also influence observed performance is less obvious, but Bargh, Chen and Burrows (1996) found participants for whom elderly stereotypes were primed walked more slowly down the hallway when leaving an experiment than did control subjects, suggesting that, if mood state lead to participants thinking more negatively about their activity, they would be likely to perform less activity. There are, therefore, theoretical frameworks both for an effect of control cognitions being mediated by emotion and for an effect of emotion being mediated by control cognitions.

A number of studies have supported an effect of control cognitions on activity independent of emotion. Lackner and Carosella (1999) reported self-efficacy but not psychological distress to contribute to the performance of lifting tasks in low back pain patients. In a population of patients with rheumatoid arthritis, Holm, Rogers and Kwoh (1998) found self-efficacy but not depression to predict self-reported ADL function. Self-efficacy was found to predict self-reported physical functioning after depression and anxiety had been controlled in coronary heart disease patients by Sullivan, LaCroix, Russo and Katon, (1998). Similarly, Johnston, Morrison et al. (1999) found perceived control to predict recovery from stroke, controlling for mood; there was no evidence of this effect being mediated by mood.

Evidence of an independent effect of emotion on physical activity seems relatively sparse. Nevertheless, Langer (1983) reports a nursing home study where residents were allocated to either a neutral condition or to a conditioned designed to increase perceptions of control. On follow-up, the experimental participants reported themselves to be more active than did the control group but no significant change in perceived control was seen. Greater increases in happiness were, however, found in the experimental group than the control group. Härkäpää et al. (1991) found distress but not health locus of control to be associated with the poorer performed accomplishment of low-back pain exercises whilst a more internal locus of control was associated with (seemingly self-reported) exercising more frequently. Finally, Fisher and Johnston (1998) found distress but not control cognitions to mediate between pain and disability in chronic pain patients. The authors suggested that the absence of control effects could be due to using a locus of control measure rather than assessing self-efficacy type control cognitions.

Hence, whilst much evidence suggests that perceived control is not dependent on emotion for its effects on activity and has an independent effect, less finds an effect of emotion on activity when control cognitions are controlled. However, both possibilities are, to some extent, supported. An experiment manipulating both perceived control and emotion would permit it to be confirmed whether these variables do have independent effects or whether one is dependent on the other.

xii. Summary

Measurement of Activity Limitations

The different methods of assessing activity limitations (self-report, proxy-report, observed performance and continuous monitoring) differ in their advantages and disadvantages. It seems that the measures may often ask different questions and address different aspects of activity. Where measure types have been compared, findings are inconsistent and, where measures are significantly associated, much variance remains unaccounted for. There is need, therefore, for further research comparing measure types, in particular, comparing measures assessing activity limitations on the same specific activity dimension (for example, walking activity).

Measures are likely to differ in the extent to which they are affected by biases of negative affectivity, social desirability and gender. Self-report measures are likely to be particularly vulnerable. However, only limited evidence exists to support bias hypotheses. Hence the extent to which these variables have biasing effects, especially on self-report measures, needs to be investigated.

Prediction of Activity Limitations

Both emotion and control cognitions have been found to predict activity limitations. However, the literature is not clear as to the extent to which the effects of these variables are dependent on measure type. Therefore, the question of whether the predictive effects of emotion and control cognitions depend on activity measure type, when activity measures address the same activity domain (walking) is addressed in this thesis.

A link between control cognitions and emotion is supported in the literature. It is possible that only one of these variables actually reliably predicts activity and the association of the other with activity is an artifact, caused by the variable's relationship with the other. Alternatively, one of these variables may mediate the effects of the other. Hence there is need to investigate whether emotion and perceived control each independently influence activity.

Research Questions

Four main questions have been identified:

1. To what extent do the different measure types give corresponding results?
2. Are measures, especially self-report measures, affected by biases of negative affectivity, social desirability and gender?
3. Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?
4. Do emotion and control cognitions independently affect activity?

These questions were investigated in four studies. The Pilot Study (Chapter 2) was an exploratory, small-scale pilot study of a student population. Self-report, performance and monitor ratings of activity were compared and psychological measures were taken to address the problem of bias and to explore the role of psychological predictors. The Calibration Studies (Chapter 3) addressed instrument calibration issues.

The Walking Difficulties Study (Chapter 4) was based on the pilot study but carried out with a sample of people with walking difficulties.

The first three questions were also investigated in the Stroke Study (Chapter 5), a larger study with stroke patients. Self-report, proxy-report and observed performance measures were compared.

Finally, whilst data collected included material addressing questions 1-3, the primary purpose of the Experimental Study (Chapter 6) was as to investigate whether emotion and control cognitions independently affect activity. The effects of emotion and perceived control manipulations on student participants' walking speed were examined.

CHAPTER 2

A PILOT STUDY: COMPARING SELF-REPORT, OBSERVED PERFORMANCE AND CONTINUOUS ACTIVITY MONITORING IN A STUDENT SAMPLE.

Abstract

Background

The relationships between different types of measures assessing activity in a similar domain are unclear. Biases of negative affectivity, social desirability and gender may affect activity ratings, especially on self-report measures. It is also not clear whether the effects of emotion and perceived control on activity are dependent on the measurement method used to assess activity outcome. This pilot study aims to test methods for investigating these issues.

Research Questions

1. To what extent do self-report, observed performance and continuously monitored assessments of activity correspond?

2. Are measures affected by biases of negative affectivity, social desirability and gender?

3. Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Methods

14 healthy adults completed attended laboratory sessions on two consecutive days separated by 24-hours wearing a device continuously monitoring activity. In the laboratory sessions, participants completed self-report and observed performance (session 1) measures of activity as well as measures assessing negative affectivity, social desirability, emotion and perceived control.

Results and Conclusions

The methodology employed was found to be feasible and acceptable to participants. Only activity monitor and self-reported 24-hour activity were found to be related at trend level. It would seem that measures that assess activity over identical time periods are more likely to be associated. On standardised measures, participants under-reported distance and over-reported time. No support for biasing effects of negative affectivity or social desirability on self-report measures was found. No significant effects of gender on measures were found, but findings were in the hypothesised direction suggesting this question merits investigation in further studies. No significant effects of emotion and perceived control on activity were seen, but there were trend effects of Depression and Generalised Self-Efficacy on self-reported usual activity, suggesting the predictive effects of these psychological variables may depend on the type of measure used.

Introduction

As demonstrated in the introductory chapter, the extent to which self-report, performance and monitor measures of activity are related is unclear. A particular problem in the literature is the lack of research comparing measures assessing the same domain of activity; if a study does not find activity output on a monitor assessing lower body movement to correlate highly with a general activities of daily living measure it is not possible to determine whether unexplained variance is caused by differences of the measure type or differences in the domains of activity addressed by those particular measures selected within the measurement category.

Watson and Pennebaker (1989) suggested that self-report measures of health status are related to negative affectivity because individuals high in negative affectivity are hypervigilant in their attention to symptoms and ambiguous stimuli are interpreted negatively. This negative affectivity hypothesis was supported by Watson and Pennebaker (1989) and is consistent with findings of McCrae et al. (1976) and Tessler and Mechanic (1978). However, this work has focussed on symptom report and participants' reporting of symptoms has been compared with more 'objective' measures of health status. Whether negative affectivity may also bias self-report measures of activity limitations is not clear. A second variable proposed to bias self-report measures is social desirability: those participants with a greater desire to appear good to the researcher may report higher activity. Third, there is some evidence to support gender as a biasing variable: some research (e.g. Merrill et al. 1997) suggests that men may be more likely to over-report and women to under-report activity as compared with performance measures of activity.

The literature indicates that both emotion and control cognitions have been found to predict activity limitations. However, the extent to which such effects are dependent on the activity measure type is not clear. For example, Härkäpää et al. (1991) found increased distress to be associated with poorer exercise performance but not to predict self-reported disability. In contrast, Johnston, Earll et al. (1999) found mood to predict self-reported disability with ALS/MND patients. Little research has been conducted assessing the roles of emotion and control cognitions in predicting activity measured with continuous monitoring methods; such methods allow an assessment that is more objective than self-report measures but allows a more natural assessment of activity than an observed performance measure. Hence further research is required to determine to what extent emotion and control cognitions predict activity across measure types, when measures are designed to assess, as closely as possible, the same activity domain.

These issues are later examined with a sample of people with walking difficulties (Walking Difficulties Study, Chapter 4). To determine whether the methodology would be acceptable to participants and feasible for the researcher, the present pilot study was first performed with a healthy, mainly student sample. The measures selected for assessing activity in this population are discussed below.

Physical Activity Measures

The three main measures used were the Paffenbarger Physical Activity Questionnaire (PPAQ) (Paffenbarger, Wing, & Hyde, 1978), Harding et al.'s (1994) 10-minute walk (a performance measure) and the Numact activity monitor.

1. Self-report

The criteria for selecting a self-report measure of activity were that it be appropriate for the mainly student sample and that it ask about activity that could legitimately be compared with the walking-based observed performance and continuous monitoring measures described in detail below. A decision that had to be made with the healthy adult pilot sample was whether to ask only about walking activity performed or a wider range of physical activity. Asking about walking activity only would mean that self-reported walking could be compared with an observed walking test and walking as assessed by the activity monitor. However, whilst this might be valid in a population of people with walking difficulties, it is unlikely to be valid in the present population. For someone who rarely walks but runs and swims on a regular basis, how far they *can* walk in a walking test seems more likely to be associated with total usual activity than with usual walking activity. Also, the activity monitor does detect non-walking ambulation such as running and jumping. Hence it was decided to select a self-report questionnaire that would include walking but that would also ask about other types of activity performed.

As studying is a largely sedentary occupation, the questionnaire needed to include leisure-time activity. Most students spend part of the year in a different, non-university environment, so a questionnaire with a relatively short time frame was preferred to one asking about 'usual' activity or activity over the past year. A measure assessing activity over the 24-hours for which the monitor was worn as well as a measure of more usual activity (within a fairly short time frame) was required. Whilst a diary measure completed at regular intervals throughout the day could have minimised recall bias and given a more accurate breakdown of activity, it was decided not to include one here as it could lead to improved recall of events when participants returned to the laboratory and so obscure one of the problems with self-report measures that is being investigated.

The Paffenbarger Physical Activity Questionnaire (or the 'College Alumni Physical Activity Questionnaire') (PPAQ) was selected from Kriska and Caspersen's (1997) 'A collection of Physical Activity Questionnaires for Health-Related Research' according to the criteria described above. The time frame of recall can be selected as either the past week or past year; the past week was used here. A score in METs can be devised from three questions asking about distance walked each day, number of stair flights climbed each day and sports or recreational activity performed during the past week. The original questionnaire asked how many city blocks are walked each day. As the participants in this study were European, not American, the question was rephrased to ask how far participants normally walked, with the answer requested in miles or kilometres. Item 4, requesting information about sport, recreational or other physical activity was also modified for this study. 'Number of times per year' of performance was replaced with 'number of times per week, and 'years participation' was excluded. After 5 participants had completed the experiment, an extra component was added, asking for activity intensity to be rated as light, moderate, vigorous or very vigorous (see Appendix A.i for the questionnaire in full). The PPAQ includes an item asking for a breakdown of activity over an average day into time spent in vigorous, moderate, light, sitting activity and sleeping or reclining. This item was also included in analyses here as it allows self-reported time spent neither sitting nor lying to be compared with monitor posture reports. Whilst the PPAQ asks for usual weekday and usual weekend hours in these activities, the items have been combined here into a single item of activity in a usual day as, with this population, the difference between weekday and weekend activity is less clearly marked than in the general population. For example, Wednesday is a common day for sports. The PPAQ is easily adapted to request information from the previous day only (Appendix A.ii).

A number of validity and reliability studies have been performed on the PPAQ (see Kriska & Caspersen, 1997). Ainsworth, Leon, Richardson, Jacobs and Paffenbarger (1993) found significant correlations of 0.29 with Caltrac, 0.60 with maximal oxygen capacity (VO_2 max) and –0.44 with percentage body fat. Similarly, Jacobs et al. (1993) found the PPAQ to correlate with Caltrac, VO_2 max and percentage body fat by 0.30, 0.52 and –0.30 respectively. These items have been found to predict heart attack risk (Paffenbarger et al., 1978). Whilst these correlations are not consistently high, the measure has mostly been validated against variables assessing fitness and risk of heart disease rather than activity level.

A range of test-retest reliability coefficients is reported by Kriska and Caspersen (1997), from 0.34 to 0.73 for the total index. Time of retest seems to be an important factor: Jacobs et al. (1993), for example, reported test-retest correlations of 0.5 for the total index, 0.3 for the stairs items, 0.39 for blocks walked and 0.63 for sports when months 1 and 13 were compared. For months 13 and 14, correlations were 0.72 (total), 0.78 (stairs), 0.63 (blocks) and 0.75 (sports). As activity patterns can conceivably change over time, high test-retest correlations over long time periods would not be expected.

Self-reports of both usual activity (activity over the past week) and activity during the time wearing the monitor were taken with the PPAQ. Additionally, participants were asked to give self-reports of their activity in the controlled laboratory environment. They were asked to estimate distance walked and the duration of walking in the walking task, and the amount of time for which they completed questionnaires. They were asked to lie, sit and stand for time periods set by the researcher; estimations of these time periods were recorded.

2. Observed Performance

A performance measure was required that would be comparable with both Numact readings and the self-report measure and be suitable for a young adult population. Tests based on walking were investigated.

A wide variety of walking tests have been devised, ranging from participants taking 2 steps, walking 30m to distances walked in 2, 6 or 12 minutes (Myers et al., 1993). McGavin, Gupta, and McHardy (1976) propose that using a standard time and assessing distance could allow a more uniform test of endurance than setting a standard distance and measuring time taken. However, the reverse of the argument would then be that setting a standard distance would give higher variability between participants, therefore being a more discriminating measure. It seems unlikely to matter whether a distance or time is set in a student population if individuals are asked to walk as briskly as possible. Both types of measure allow participants to be placed on a continuous scale of ability.

Cooper (1968) devised a 12-minute performance test, where 115 male US air force officers were instructed to cover as much distance as possible in the time, preferably running but walking when necessary. On repeats of the test, little training effect was found, and performance significantly correlated with maximal oxygen consumption (0.897).

A 12-minute walking test, based on Cooper's task, was used by McGavin et al. (1976) to assess disability in patients with chronic bronchitis. In contrast with Cooper (1968), a significant increase in the distance walked was found at the second trial (on a different day); no significant change was found on the third walk. It was therefore suggested that the test must be performed twice for test-retest reliability. McGavin et al. (1976) suggested that this kind of test is affected by similar factors to 'everyday disability', such as respiratory, cardiovascular and neuromuscular function, motivation and endurance. Walking tests are closer to ordinary function demands than tasks such as cycle ergometer testing in patient populations and are also simple, safe and inexpensive (Guyatt et al., 1985).

Guyatt et al. (1985) had 8 patients with chronic heart failure and 25 chronic lung disease patients perform a 6-minute walk 6 times over 12 weeks. Participants were randomly allocated to encouragement or no encouragement conditions. Encouragement was found to improve performance (this was also found by Guyatt et al., 1984). Stable results were found after the first 2 tests: improvement in scores occurred up to the third walk. This seems not to be unusual with this type of test: Mungall and Hainsworth (1979) found the distance covered in a 12-minute walking test to increase over the first 3 trials in a population of 13 patients with chronic obstructive pulmonary disease (6 trials at 2-3 weekly intervals). Guyatt et al. (1984) used 2- and 6-minute walks for six fortnightly tests with chronic airflow limitation and chronic heart failure patients. The first two days' results were compared with the last four visits; it was found that performance had improved with repetition.

The Incremental Shuttle Walking Test was developed by Singh, Morgan, Scott, Walters and Hardman (1992) in response to the difficulty of standardising performance and the potential for influence from motivation or encouragement with the 6 or 12 minute walking tests. The shuttle walking test standardises walking speed and gradually increases the pace required through the test, thus reducing biases caused by the manner in which tests are conducted and enabling valid comparisons between studies to be made. Participants are instructed to walk up and down a 10 metre length. A tone sounds from an audiocassette each time the participant is required to be at the starting point. After each minute, the speed level at which the participant walks increases such that, at level 1, 3 shuttles are walked. Thereafter, the number of shuttles to be walked increases by one each minute, so, for level 2, 4 shuttles are walked and 5 for level 3, with a total of 12 speed levels. When participants can no longer keep up with the audiocassette, the total distance walked is recorded. Unfortunately for the purposes of this pilot study, the test was devised for use with chronic obstructive pulmonary disease patients and ceiling effects were likely to occur with the healthy student population assessed here. On piloting with a postgraduate student, the highest speed level was reached. Léger and Lambert (1982) developed a more challenging 20m shuttle-run test which would better have suited healthy student participants. Unfortunately facilities were not available to set up this test.

Harding et al. (1994) tested walks of 10 and 5 minutes. The 5 and 10-minute performances were highly correlated. However, as the 5-minute score was taken as the distance participants had walked half way through the 10-minute test, it seems likely that scores would differ if individuals knew they would be walking for a shorter time. Whilst 6 and 12 minute walks have been found to show training effects between first and second performances, Harding et al. (1994) found excellent test-retest reliability: both 10 and 5-minute walks were found to have good test-retest reliability ($r = 0.944$ for the 10 minute scores). This could be influenced by the time interval between performances - Harding et al used a 12 week interval whereas Guyatt et al. (1984), Guyatt et al. (1985) and Mungall and Hainsworth (1979) had fortnightly inter-test intervals. Harding et al. (1994) also showed inter-rater reliability to be high ($r = 0.994$). The risk of ceiling effects is greater in shorter tests (Harding et al., 1994). Hence the 10-minute walk was selected over the 5-minute walk for use in this chapter's pilot study.

Harding et al. (1994) chose the time of 10 minutes to sufficiently test limitations in fitter participants, whilst less able patients were able to take time to rest. No walking aids were allowed. Whilst this is unlikely to be an issue in the pilot study conducted here, if this test were to be used in a similar paradigm with a patient population, allowing people to use aids may give a result more comparable with scores taken from the participants' usual environment (self-report and activity monitor scores).

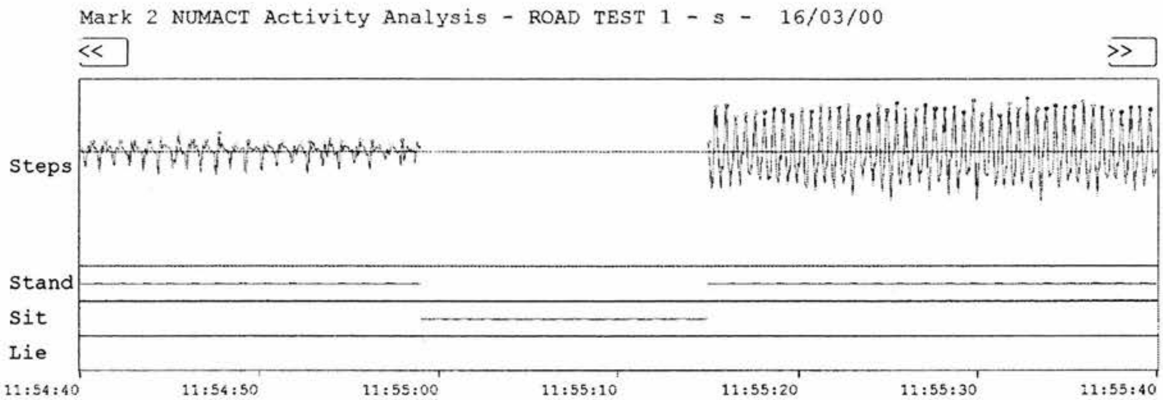
3. Automated monitor

The Numact monitor consists of 2 sensors, one attached to the chest and one to the outer thigh, wired to a microcomputer carried in a belt pouch (see Walker, Heslop, Plummer, Essex and Chandler, 1997). Both leg and chest sensors contain tilt switches which, combined, detect whether the participant is lying, sitting or standing. The chest sensor also contains an accelerometer, allowing amplitude, or force of steps, to be detected. An example of output obtained is shown in Figure 2.1. A period of slow walking, followed by sitting and then brisk walking is displayed. Steps are only recorded when the participant is judged to be standing.

The vertical impulse at heel strike recorded by the accelerometer is used to detect steps (Walker et al., 1997). The peak value has to exceed a set threshold value (Numact analysis software default value: 136); values below the threshold are assumed to be caused by background noise. Impulses also have to occur at intervals greater than a set value (default: 0.5s); ie, for a 'step' to be recorded, it must not occur at more than 0.5s after the previous step.

As with comparing any two systems of measurement, it is necessary to ensure that inter-monitor bias (errors systematically occurring in the same direction, such that one monitor gives scores higher than the other) is low and agreement (the extent to which monitor scores concur) is high. Walker et al. (1997) videotaped an individual's activity. Two observers watched the tape independently and steps were counted. No systematic error was detected and the variation between observers was similar to that between observers and monitor. Over 110s of recording, the agreement between observation and monitor had a kappa value of 0.933. Between 2 different monitors, kappa = 0.857.

As seen in Figure 2.1, data output includes the product of step number x amplitude, which should be a reflection of energy expenditure. Using a treadmill test, a good linearly correlated relationship was seen between step number x amplitude and METs (Walker et al., 1997).



Time	Duration secs	Lie secs	Sit secs	Stand secs	Steps	Amp Mean	Steps x Amp	St Int Mean	C Walk Mean	C Walk SI Mean	C Walk n
TOTAL	17110	1522	7201	8387	11577	53.08	614512	0.619	81.30	0.512	4857

Five minute energy - 0:14:06:57 Energy 58347 Mean amplitude 100.08 Number 583
Longest uninterrupted walk - 0:11:55:15 Duration 541 Energy 71148 Mean amplitude 69.28 Number 1027
Data period - Start Thursday, March 16, 2000 10:00:00 Stop Thursday, March 16, 2000 16:00:00

Figure 2.1: Numact output.

Research Questions:

1. To what extent do self-report, observed performance and continuously monitored assessments of activity correspond?

2. Are measures affected by biases of negative affectivity, social desirability and gender?

3. Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Methods

The study was approved by the University of St Andrews School of Psychology Ethics Committee (see Appendix G).

Design

Participants, mainly students, attended laboratory sessions on 2 occasions 24 hours apart. Questionnaire measures were completed, including a self-reported activity measure, and a walking task was performed at the first time point; further questionnaire measures were taken at the second time point. An activity monitor was worn throughout the intervening 24 hours. The three measure types were compared cross-sectionally. The effects of bias on measure types were also explored cross-sectionally. Where possible, the associations of control cognitions and emotion with different measurement methods were assessed predictively.

Participants

14 participants were recruited via a sign-up sheet in the psychology department foyer requesting healthy volunteers (Appendix F.i), or by word-of-mouth. £25 was paid on completion of all tasks. 5 of the participants were undergraduate students, 7 were postgraduate students and 2 reported ‘other’ occupations. Participants’ ages ranged from 18 to 32; 11 were female, 3 were male.

Measures

Activity Measures: Self-Report Measures

Self-Reported Activity: PPAQ (Appendix A.i, A.ii)

The PPAQ was presented to participants at both first and second laboratory visits. At time 1 (PPAQ1), scores for activity in the past week were obtained; the time 2 questionnaire (PPAQ2) asked about activity in the past day. PPAQ1 kilocalorie scores were obtained using number of miles walked and number of stairs climbed per day and types and duration of sports, recreation or other physical activities performed in the past week (see Appendix Ai/ii.1,3,4). The formula used to calculate PPAQ1 is shown below (from Kriska and Caspersen’s (1997) collection of physical activity questionnaires). The kilocalorie scores for distance walked and stairs climbed were defined by Paffenbarger et al. (1978); MET activity intensities as defined by Ainsworth, Haskell et al. (1993) were used. PPAQ2 scores were calculated using the same method, except that miles walked and stair flights climbed per day were not multiplied by 7, and activity undertaken during only the past day was recorded, giving a score in kcal/day.

(miles per day x 7days per week x 96kcal) + (stair flights per day x 7days per week x 4kcal/flight) + (activity MET intensity x duration(minutes) per occasion) = PPAQ1 kcal per week

A second score derived from the Paffenbarger Physical Activity Questionnaire was taken from the final item of the scale (see Appendix A.i.8, A.ii.6): time spent in each of 5 activity levels. The times of the 'standing' categories (light, moderate and vigorous activity) were summed, giving a score to compare with monitor records of time spent standing.

Self-reported time and distance walked in walking task. (Appendix A.iii) Participants were asked to estimate the distance and time for which they walked in the 10-minute walk.

Self-reported time spent sitting, standing and completing questionnaires. At the start of the first laboratory session, participants were asked to sit and to stand for a short time, until the researcher asked them to stop. Participants were later asked to estimate for how long they were sitting and standing for this task (Appendix A.iii). The time period selected for both sitting and standing was1 minute 45 seconds because it was less likely to be chosen as an anchor point than 1 minute. Participants were told they were being asked to do this in order to calibrate the monitor (the task's secondary purpose). Participants were also asked to estimate the time spent completing questionnaires over the whole of the first study session. The researcher recorded time completing questionnaires with a stopwatch.

Activity Checklist

At the second laboratory session, participants were asked to complete a table (Appendix A.iv), ticking boxes to indicate whether they had been lying, sitting, standing, running or cycling during each hour wearing the monitor. This was as a check for monitor anomalies when unexpected output appeared (e.g. 'lying' readings in the middle of the day) and was not included in analyses.

Activity Measures: Observed Performance

10-minute Walk (Harding et al. 1994)

Harding et al. (1994) used an empty corridor with marks 20m apart as a course. Patients were told the time of each lap, or at each minute if laps were very slow. In the present study, as an empty corridor was not available, participants were asked to walk between 2 walls 10m apart in the laboratory. The floor was marked at 1m intervals to facilitate the recording of distance. Participants were told they would be asked to walk for somewhere between 10 and 15 minutes, and that their aim should be to cover as much distance as possible in the time allowed. They were informed that, to keep the task constant, the experimenter would not speak other than to let them know when they were ¼, ½ and ¾ of the way through the test. Care was taken to conceal the time taken to perform the test because participants were asked to report their perception of the time later in the laboratory session. The distance participants walked in 10 minutes was recorded.

Activity Measures: Automated Activity Monitor Assessment

Numact Activity Monitor

The Numact Activity Monitor, as described earlier, was used to assess activity over a 24-hour period. Activity between the time at which participants left the laboratory after the initial session and the time at which participants returned to the laboratory was analysed. Time of leaving was taken as the time at which Numact output indicated the participant standing on completing the questionnaire; the end time was the time at which participants sat down on their return to the laboratory. Difficulties were found in obtaining data read-outs from the Numact device for the time period required: the analysis programme would not always give output for the exact time period but presented activity data for a small extra time period (range: 1-55s). This is a small proportion of 23-hour data and was always time sitting. As the time between laboratory sessions varied depending on arrival times and the speed of filling out questionnaires, activity was averaged, giving a 23-hour reading for all participants.

Two readings were used from the monitor: steps x amplitude, a measure related to METs, and the total time for which participants were recorded to be on their feet: 'standing' time.

The study's original Numact monitor failed after 6 participants and was replaced with a second Numact monitor.

Measures Assessing Bias Variables

Negative Affectivity: Positive and Negative Affect Schedule (PANAS) (Watson, Clark, & Tellegen, 1988)(Appendix B.i)

PANAS contains 20 adjectives describing 10 positive and 10 negative moods. Participants rated the extent to which each word fits the way they feel 'at the present moment' on a 5 point scale. Scores for both positive and negative affectivity (NA) can be derived; the NA scores will be used here. Internal consistency of the NA subscale has been demonstrated (.85); test-retest reliability over an 8 week period was found to be moderate ($r = .45$) (Watson et al., 1988); a high test-retest correlation of state measure over a long time period would not be expected. PANAS external validity was addressed using longer PANAS time frames and is discussed in the Walking Difficulties Study. Whilst both positive and negative affectivity scales were administered here, only the negative affectivity scores were analysed.

Social Desirability: Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960) (Appendix B.ii)

This measure was developed to improve upon that devised by Edwards (1953). Crowne and Marlowe (1960) suggest that, since the Edwards scale drew items from the Minnesota Multiphasic Personality Inventory it was not possible to determine whether responses reflected social desirability or a genuine absence of pathologic symptoms. Crowne and Marlowe (1960) aimed to find items reflecting culturally sanctioned but improbable behaviours. Items had to fit the criteria of cultural approval with minimal pathological or abnormal implications in whichever direction participants responded. Fifty items were scored by 10 judges with true/false categories in the socially desirable direction. 90%

agreement was reached by 47 items. These and Edwards' items were rated by 10 judges for maladjustment implied by undesirable responses on a scale from 1 (extremely well adjusted) to 5 (extremely maladjusted). Whilst the Marlowe-Crowne items were given a mean rating of 2.8, suggesting neither good nor bad adjustment, Edwards' items received a higher rating of 3.9. It would, therefore, appear that the Marlowe-Crowne items were less related to maladjustment. The Marlowe-Crowne items were administered to 76 students and item analysis was performed. The 33 items discriminating at .05 level between high and low total scores became the Marlowe-Crowne Social Desirability Scale.

Administration of the scale to 39 students revealed an internal consistency score of .88 (Kuder-Richardson formula 20); completion of the scale 1 month later by 31 of these participants gave a test-retest correlation of .89. The correlation between the Marlowe-Crowne Scale and Edwards' Scale was .35 ($n = 120$, $p < .01$) (Crowne and Marlowe, 1960). It would, therefore, appear that the scale is reliable and, whilst related to another measure of social desirability, less associated with symptoms of psychological pathology. Hence the scale consists of these 33 items which were marked as true or false by participants.

Measures of Psychological Predictors: Emotion and Perceived Control Variables

Anxiety and Depression: Hospital Anxiety and Depression Scale (HADS) (Zigmond & Snaith, 1983) (Appendix C.i)

The HADS was designed to assess anxiety and depression without being confounded by somatic symptoms indicating physical disorders. The scale contains 14 items, 7 to assess depression and seven for anxiety. For each item, participants are asked to choose one of four responses to best indicate how they have been feeling over the past week.

For example:

I feel tense or 'wound-up': 1. Most of the time 2. A lot of the time 3. Time to time, occasionally 4. Not at all.

Each item was scored from 0 – 3; higher scores indicated greater distress.

Moorey, Greer, Watson, Gorman, Rowden, Tunmore, Robertson, and Bliss (1991) found HADS to have good psychometric properties. The internal consistencies (Cronbach's alpha) were 0.90 for the depression and 0.93 for the anxiety subscales. Moorey et al. (1991) found anxiety and depression items to load onto separate factors except for item 7 from the anxiety subscale: 'I can sit at ease and feel relaxed'. Johnston, Pollard and Hennessey (2000) also found the HADS to perform satisfactorily. Using exploratory factor analysis, loadings on the psychological factor were higher than loadings on the somatic factor for all items for patients with breast disease. Confirmatory factor analysis indicated that, for MI and stroke patients, anxiety and depression items separated appropriately. Internal consistency scores were calculated for all 3 samples (breast disease, MI and stroke patients); for each population Cronbach's α exceeded .7; the majority of analyses reached $\alpha > .8$. Concurrent validity is suggested by the significant correlations found between HADS anxiety and depression scores and psychiatric ratings of 100 medical out-patients: anxiety, $r = .54$, depression, $r = .79$ (Zigmond & Snaith, 1983).

Generalised Self-Efficacy: Generalised Self-Efficacy Scale (GSE) (a short version of this scale is outlined in Jerusalem and Schwarzer (1992); for further details see Johnston, Wright & Weinman, 1995). (Appendix C.ii)

This is a 10-item scale assessing beliefs in managing to deal with general difficult situations. High internal consistency has been found. Test-retest reliability in a group of migrants from the then East Germany was found to be 0.47 for men and 0.63 for women. Concurrent validity has been shown through correlations with assessments of, for example, self-esteem (.52), internal control beliefs (.40), optimism (.49), general anxiety (-.54) and pessimism (-.28) (see Johnston et al 1995).

Perceived Control – over walking task and 24 hour activity. (Appendices C.iii, C.iv)

Five items assessing perceived control over the walking task and 5 for control over activity during the 24 hours wearing the monitor were developed from Conner and Norman (1996). All items were rated on 7-point scales.

Other Measures

Intention – to perform well in walking task and to be active over 24 hours (Appendices D.i, D.ii)

One item assessing intention to perform well in the walking task/to be active over the next 24 hours was developed for this study: 'How much effort do you intend to put into the walking task?' Six items were adapted from Conner and Norman (1996). 7-point rating scales were used for all intention items. This variable was not included in analyses.

Whether could have walked further, if so how far (P7 onwards) (Appendix D.iii)

It was noticed that some participants, despite indicating a strong intention to try hard in the walking test, did not appear to be putting in much effort. An item was therefore included (participant 7 onwards) asking whether participants thought they could have walked further, and if so how many more laps they could have walked. This item has not been included in analyses.

Distance of home from Union

St Andrews is a small town and it is common for students to make their way around by foot or by bike and, for some, this could be a major factor determining activity level. Therefore the distance lived from the town centre (the Union building) was assessed. This was assessed by asking for participants' address, and then measuring distance on a map.

Procedure

Participants were contacted before the experiment to arrange an appointment and to inform participants that they would not be able to swim, shower or play contact sports during the 24 hours wearing the monitor.

The procedure was outlined to participants in an information sheet and informed consent was obtained (see Appendix F.iv). Details of the walking task's requirements were given verbally to enable participants to respond to pre-walk perceived control and intention items referring to the task. Height and weight were recorded and the Numact activity monitor was attached. Participants completed the first questionnaire containing scales in the following order: age, gender, occupation, address, PANAS, HADS, PBC over walking task, intention to perform well on walking task, and PPAQ 1. The experimenter timed participants completing both the first and second questionnaires as a comparison for participants' estimates of the time taken.

Participants were asked to lie, sit and stand for 1minute 45 seconds each. They were told: 'In order to calibrate the machine, I'm going to ask you to lie down, sit and stand for a short time each'. This was to disguise the experimenter's intention to have participants estimate time in different postures under controlled conditions. The walking test was then performed as described earlier.

The second questionnaire was presented to participants, containing the following scales: whether further could have been walked and, if so, how many more laps; GSE; Marlowe-Crowne Social Desirability Scale; Perceived Control and Intention to be active in the next 24 hours; self-report distance and time for the walking task; self-report time spent lying, sitting and standing; and self-report time spent filling in questionnaires over the whole experiment.

At the same time the following day, participants returned to the laboratory where the third questionnaire was completed consisting of PANAS, PPAQ2, and the activity checklist.

Analyses

Unless otherwise stated, throughout the thesis, analyses were performed with SPSS version 9 or 10. Analyses were 2-tailed.

Non-parametric statistics were used because of the small sample size and, on many measures, scores did not fit a normal distribution. The sample size of this exploratory pilot study was such that statistically significant findings were unlikely; whilst the significance level of $p = .05$ is used, trend effects with $p < .1$ are discussed.

> *Research Question 1: To what extent do self-report, observed performance and continuously monitored assessments of activity correspond?*

Correlations (Spearman's rho) between walking test performance and self-reported distance covered in the test, walking test performance and self-reported usual activity (PPAQ1), 24-hour monitor records and self-reported 24-hour activity (PPAQ2) and self-reported usual (PPAQ1) and 24-hour (PPAQ2) activity were performed.

Sign tests were conducted to test whether there were differences in the numbers of participants over- and under-estimating time spent walking and distance walked in the walking task, time filling in questionnaires and time lying, sitting and standing when requested to do so by the researcher.

> *Research Question 2: Are measures affected by biases of negative affectivity, social desirability and gender?*

The small sample size of this pilot data allowed the following analyses: Correlations were performed assessing the relationships of negative affectivity and social desirability with the activity measures. The Negative Affectivity score of the same time point as each activity measure was taken as the effect of negative affectivity on responses rather than its predictive effects were of interest. Self-reported and actual mean distances walked in the walking task by female and male participants were compared, as were self-reported and monitor-rated 24-hour activity means; these data were used in an exploratory capacity to inform as to whether further investigation into gender bias would be merited.

Research Question 3: Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Correlations were performed between Perceived Control over Walking Task and self-report and actual walking task outcomes, between Perceived Control over 24-hour Activity and monitor ratings and between GSE, Anxiety, Depression and all activity measures. The psychological variables were assessed as predictors for all correlations except those of GSE with self-report Usual Activity and walking task measures.

Results

The monitor data of 4 participants have been excluded because of monitor faults. The social desirability scale, PPAQ2 and estimated time walking were incomplete in one case each. Due to experimenter error, the following data is missing: total questionnaire time (2 participants), estimated walking time and distance, estimated time lying and standing (1 participant each) and estimated time sitting (2 participants). Hence N values vary throughout the results section.

Data for all measures to be analysed were scanned for outliers ($z > 3.29$, Tabachnick and Fidell, 1996). None were found.

Skew and kurtosis scores are shown in Table 2.2. It can be seen that a number of scales have values of greater magnitude than +/- 1. A number of scales had extreme values, greater than +/- 2: self-reported time lying and sitting (kurtosis = 6.00 and 5.58 respectively), monitored time spent standing (skew = 2.37, kurtosis = 6.15, Negative Affectivity time 2 (skew = 3.02, kurtosis = 9.60) and Perceived Control over walking task (kurtosis = 2.09 (all items) and 2.69 (no item 3). As many of the scales did not appear to fit a normal distribution, non-parametric statistics were used.

Reliability

Internal consistency analyses (Cronbach's α) were conducted (see Table 2.1). Overall, scores were moderate to good. It can be seen that, whilst α for the PANAS negativity scale was only .24 at the first laboratory session, it reached .92 at the second. The scale measures a range of negative affect states, including 'nervous', 'hostile' and 'distressed', and it could be that these states are not always consistent. However, given that other researchers have found higher internal consistency levels (see Measures section), this result could be due to sampling error in this small population. The very high consistency at the second session is probably due to a floor effect: 10 out of the 14 participants gave the lowest possible score. It is likely that participants were more relaxed about the second session, which was both more familiar to them and less demanding.

One item was detrimental to consistency in the perceived control measure: item 3: 'I would like to perform well on the walking task/be very active in the next 24 hours but I don't really know if I can'. This item can be answered in one of two ways – either as how much participants would like to perform/be active or as how uncertain they were as to their ability to do so. One participant actually voiced uncertainty over this item. Item 3 was therefore excluded from further analyses.

Mean scores and standard deviations for all measures are shown in Table 2.2.

Scale	**Cronbach's α**
Perceived Control over walking task	.83 (.05 if item 3 included)
Perceived Control over 24-hour activity	.84 (.72 if item 3 included)
Negative Affectivity time 1	.24
Negative Affectivity time 2	.92
Anxiety	.86
Depression	.78
Social Desirability	.72
GSE	.88

Table 2.1: Reliability analyses

Variable	Measure	Mean	SD	Skew	Kurtosis	N
Activity	*Self-Report (SR):*					
	Usual activity (PPAQ1, kcal/wk)	2544.74	1195.95	.37	-.26	14
	Usual daily time standing (PPAQ1, s)	32528.57	12756.56	-.52	-.95	14
	24-hour activity (PPAQ2, kcal/day)	340.91	243.54	1.02	-.36	13
	24-hour time standing (PPAQ2, s)	26572.29	12619.07	.45	-.80	14
	SR distance -walking test (m)	620.00	365.88	.27	-.74	13
	SR time - walking test (s)	664.62	113.48	1.62	1.32	13
	SR time lying (s)	120.00	24.49	.00	6.00	13
	SR time sitting (s)	132.50	50.29	1.98	5.58	12
	SR time standing (s)	120.00	34.64	.00	1.04	13
	SR time completing questionnaires (s)	805.71	393.60	.02	-.92	14
	Observed Performance:					
	Walking test distance (m)	1007.73	124.76	-.70	-.25	14
	Time completing questionnaires (s)	1109.56	311.39	.31	1.00	12
	Automated Monitor:					
	Steps x Amplitude	713929.03	295408.49	.56	-.76	10
	Time spent standing (s)	16544.90	7017.73	2.37	6.15	10
Negative Affectivity	Negative Affectivity (time 1)	11.43	1.16	.03	-1.43	14
	Negative Affectivity (time 2)	11.93	4.67	3.02	9.60	14
Social Desirability	Social Desirability	14.23	4.85	-.75	.34	13
Emotion	Anxiety	5.64	3.89	.51	.32	14
	Depression	3.14	2.87	.94	-.14	14
Perceived Control (PC)	PC - walking test, all items	27.00	6.32	-.98	2.09	14
	PC - walking test, no item 3	22.64	4.81	-1.21	2.69	14
	PC - 24-hour activity, all items	27.79	4.93	-.45	-.77	14
	PC - 24-hour activity, no item 3	22.50	4.70	-.96	.13	14
	GSE	30.36	4.13	.55	.82	14

Table 2.2: Descriptive statistics. Mean score and standard deviation for each measure.

Question 1: To what extent do self-report, observed performance and continuously monitored assessments of activity correspond?

Correlations between the various measurement methods are shown in Table 2.3. None were significant, although the correlation between monitor results and self-reported 24-hour activity neared significance (rho = .60, p = .09). The relationships of three main comparisons (self-reported usual activity with observed walking performance, observed walking performance with monitor results and monitor results with self-reported activity over the same time period) are illustrated in Figures 2.2 – 2.4. In Figure 2.2, showing self-reported usual activity plotted against observed distance walked in the walking test, it would appear that there is a cluster of distances walked in the 1000-1100m range. This lack of walking task variation would make it difficult to detect a relationship between the variables. Figure 2.3 illustrates the trend-level correlation between 24-hour self-reported activity and monitor readings. Little indication of a relationship between walking task performance and monitor readings can be seen in Figure 2.4.

Measures	Spearman's rho	P	N
Self-Report – Observed Performance			
SR distance – Observed distance (walking task)	-.10	.75	13
SR usual activity – Observed distance (walking task)	-.10	.73	14
SR 24-hour activity – Observed distance (walking task)	-.06	.86	13
SR questionnaire time – Observed questionnaire time	.18	.57	12
Observed Performance – Monitor			
Observed distance (walking task) – Steps x Amplitude	-.41	.24	10
Self-Report – Monitor			
SR usual activity – Steps x Amplitude	.08	.83	10
SR 24-hour activity – Steps x Amplitude	.60	.09	9
SR usual time standing – Monitor (standing time)	.06	.88	10
SR 24-hour standing time – Monitor (standing time)	.21	.56	10

Table 2.3: Correlations between activity measures

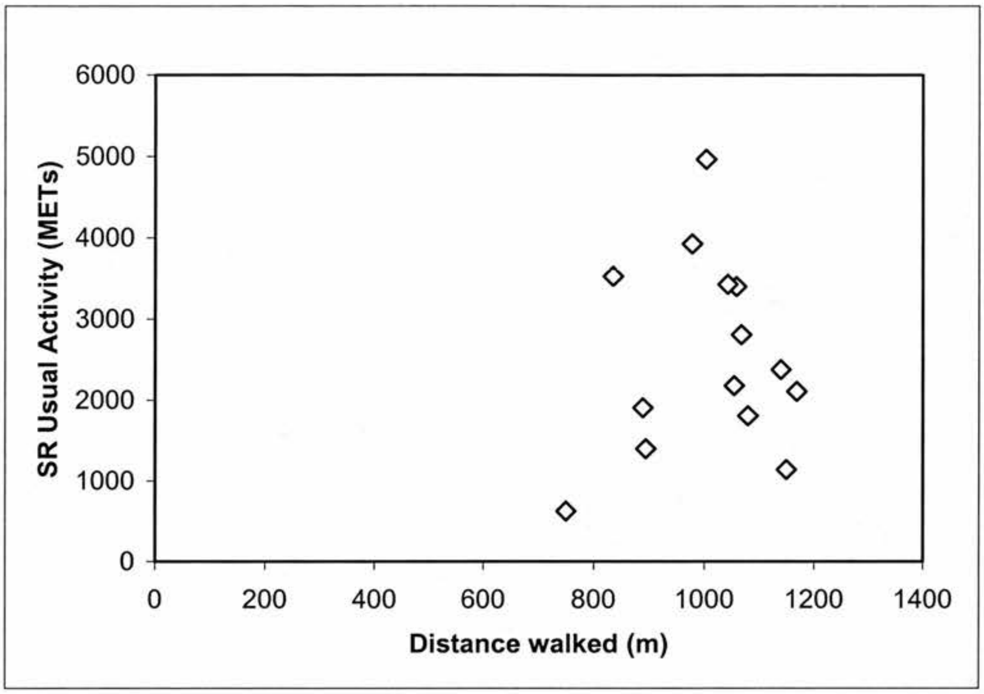


Figure 2.2: Scattergram showing distance covered in walking test and self-reported usual activity (PPAQ1). Rho = -.10, p = .73.

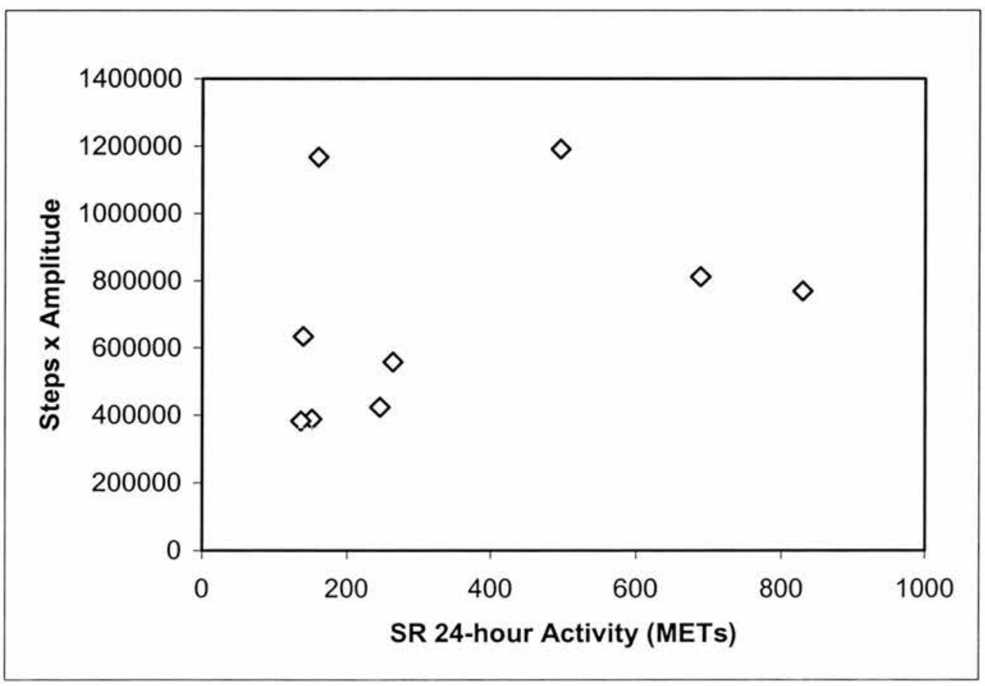


Figure 2.3: Scattergram showing monitor Steps x Amplitude and Self-Reported Activity over the same time period (PPAQ2). Rho = .60, p = .09.

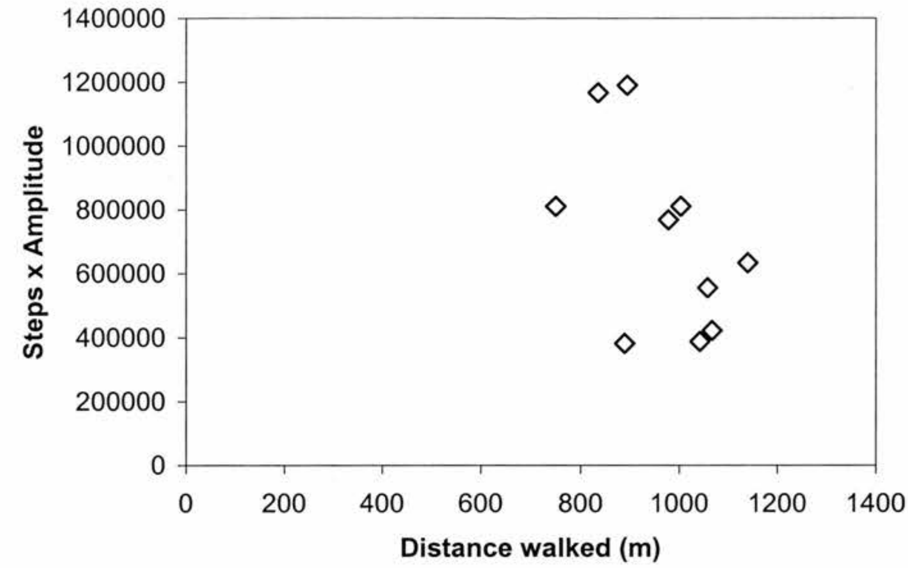


Figure 2.4: Scattergram showing monitor Steps x Amplitude and distance covered in walking test. Rho = -.41, p = .24.

The frequencies of participants over- and under-reporting distance on the walking task and time lying, sitting, standing and filling out questionnaires and time standing over 24 hours are shown in Table 2.4. The numbers of participants over-reporting and giving correct responses for walking test time are also shown. Walking test time was the only one of these measures for which participants had anchor points: the researcher had informed them that they would be walking for between 10 and 15 minutes. This is reflected in a number of participants correctly guessing 10 minutes. As the actual time was 10 minutes, the lower anchor point, no participants under-reported time spent walking. Appendix E.i contains frequency tables with the distributions of estimates. Sign tests showed that significantly more participants under-reported than over-reported walking task distance ($p = .003$). Time lying, sitting and standing was more often over- than under-reported ($p = .003$, $p = .003$ and $p = .006$ respectively). More people over than under-reported time standing (ie not sitting or lying) over 24-hours (compared with the monitor) but this was not significant ($p = .11$).

The frequencies of over- and under-report of distance on the walking task are illustrated in Figure 2.5; Figure 2.6 shows the means and standard deviations of reported and observed scores. It can be seen that there is much variability in the estimated distance scores. On average, participants under-reported distance walked by 387.7m.

Task	No. participants over-reporting	No. participants under-reporting	P (sign test)	N
Walking test (distance)	1	12	.003	13
105s lying (time)	12	1	.003	13
105s sitting (time)	11	1	.006	12
105s standing (time)	11	2	.022	13
Questionnaires (time)	4	8	.388	12
Time standing (24-hour)	8	2	.109	10
Walking test (time)	4	0[1]	See footnote	13

Table 2.4: Frequencies of participants over- and under- reporting self-reports of activities, with sign test results.

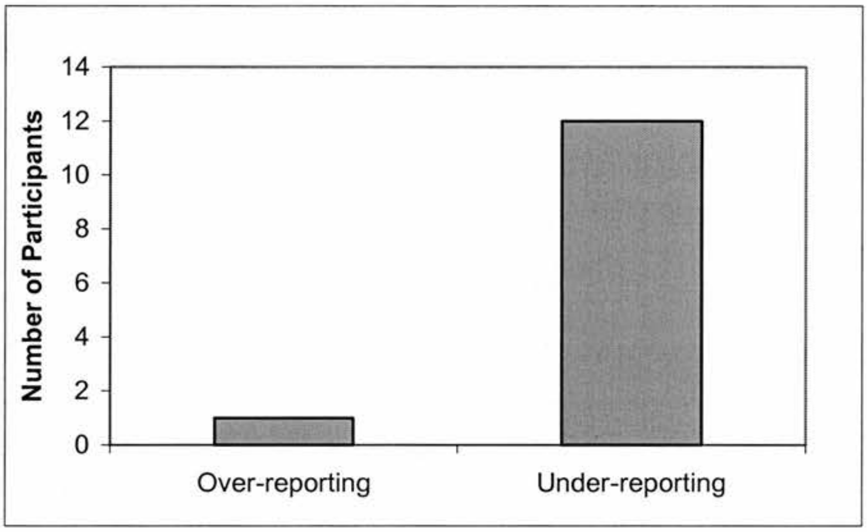


Figure 2.5: Frequencies of over- and under-reporting distance walked in the walking test.

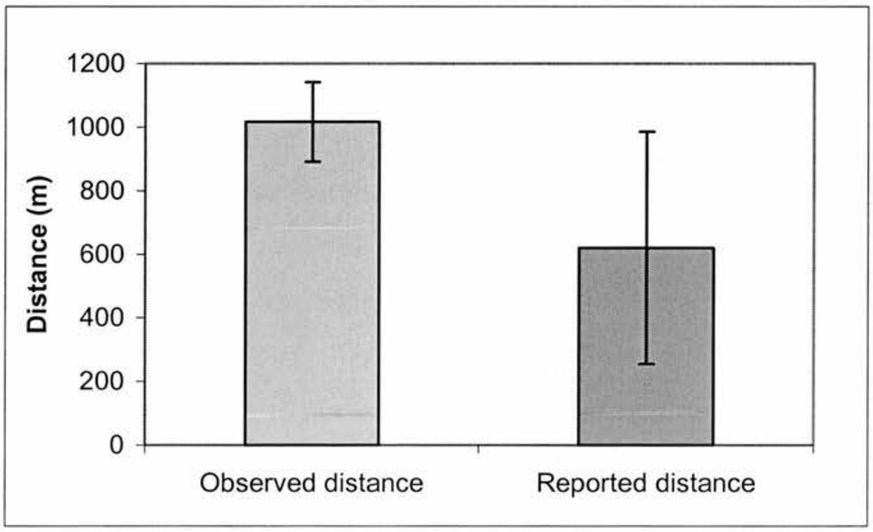


Figure 2.6: Mean observed and reported distances walked in the walking test (± SD)

[1] Participants knew that the walking task time was between 10 and 15 minutes. As the actual time spent walking was 10 minutes, under-reporting of time was very unlikely.

Question 2: Are measures affected by biases of negative affectivity, social desirability and gender?

Negative Affectivity and Social Desirability

Correlations of Negative Affectivity and Social Desirability with activity measures are shown in Table 2.5.

Bias Variable	Activity Measure	Spearman's rho	P (2-tailed)	N
NA (time 1)	SR usual activity	-.49	.07	14
	SR walking test distance	-.08	.79	13
	Observed walking test distance	-.49	.07	14
	Monitor (steps x amplitude)	.07	.85	10
NA (time 2)	SR 24-hours activity	.21	.49	13
	Monitor (steps x amplitude)	-.03	.94	10
Social Desirability	SR usual activity	.32	.29	13
	SR 24-hours activity	.19	.54	13
	SR walking test distance	.02	.96	12
	Walking test	-.05	.87	13
	Monitor (steps x amplitude)	-.46	.21	9

Table 2.5: Correlations of the bias variables Negative Affectivity (NA) and Social Desirability with activity measures. Correlations at trend level ($p < .1$) are highlighted.

The negative affectivity bias hypothesis would predict negative affectivity to correlate more highly with self-report measures of activity than with other measure types. As seen in Table 2.5, Negative Affectivity was found to correlate at trend level with Self-reported Usual Activity but the correlation with Walking Test score was at the same level (for both: rho = -.49, p = .07) whilst self-reported walking test distance was not associated with Negative Affectivity (rho = -.08, p = .79). Negative Affectivity did not near significance in correlations either with the other self-report activity measure (Self-report 24-hours activity) or with other activity measures.

Social Desirability did not near significance in any correlation with activity measures.

Gender

Tables 2.6 and 2.7 show the mean distances walked and self-reported and the mean monitor and self-reported 24-hour activity scores according to gender. In calculating these scores, any participant with a score for only one of the 2 comparison measures was discarded. Using Mann-Whitney U, no significant differences between male and female participants were found on any of these measures (observed distance walked: $U = 6.00$, $p = .16$; self-report distance walked: $U = 14.50$, $p = .94$; monitor steps x amplitude: $U = 6.00$, $p = .89$; self-report 24-hour activity: $U = 6.00$, $p = .89$) although the small size of the samples, with only 2 or 3 male participants in each calculation, does limit the likelihood of finding significant results.

	Mean distance walked	**SD**	**Mean distance estimated**	**SD**	**N**
Female	1041.90	110.63	656.00	403.44	10
Male	931.67	158.53	600.00	264.58	3
Total	1016.46	125.32	620.00	365.88	13

Table 2.6: Mean distance walked and estimated in the walking test by female and male participants.

	mean Steps x Amplitude	**SD**	**Mean SR 24-hour activity (PPAQ2, kcal/day)**	**SD**	**N**
Female	732375	331168	324.39	255.34	7
Male	600440	299407	419.52	379.69	2
Total	703056	311198	345.53	262.07	9

Table 2.7: Mean monitor (Steps x Amplitude) and self-reported 24-hour activity scores of female and male participants.

The gender bias hypothesis would predict that women would be more likely to under-report and men would be more likely to over-report walking test performance than the other gender. Here, both men and women under-estimated the distance covered in the walking test (see Figure 2.7). Figure 2.7 suggests an interaction between gender and measure could be present however: the difference between actual and estimated distance is larger for women than for men. This is supported by Wilcoxon Signed Ranks test: the difference between self-reported and observed distance walked is significant for female ($Z = -2.40$, $p = .02$) but not male ($Z = -1.60$, $p = .11$) participants. However, this finding should be treated with caution as there were only 3 male participants.

Figure 2.8 shows mean monitor and self-reported activity over 24 hours. Whilst the monitor appears to have detected more activity by women than men, women reported being less active than did men. This effect is in the direction predicted by the gender bias hypothesis but neither gender difference is significant (Mann-Whitney U = 6.00, p = .89 for both comparisons).

With these measures, much variance was found in the scores and the sample was very small, with only 7 female and 2 male participants in the comparison between monitor scores and PPAQ2 and 10 female and 3 male for the comparisons between self-reported and actual distance walked. Nevertheless, these results do suggest that the gender bias hypothesis would be worth investigating in larger samples.

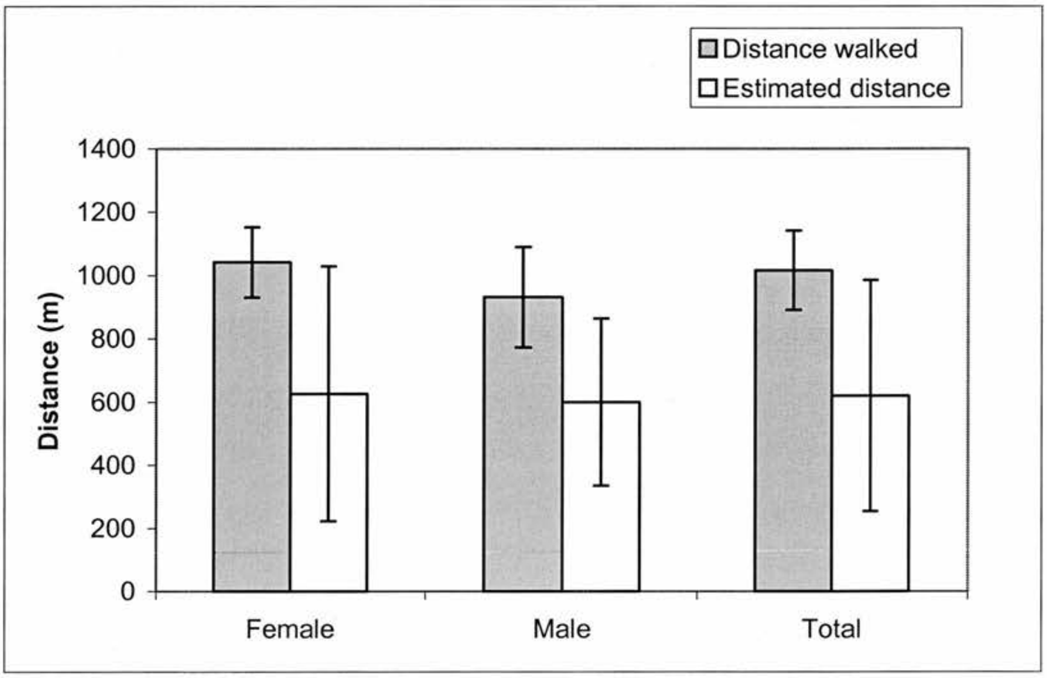


Figure 2.7: The walking test: Mean observed and self-reported distances walked (± SD) , according to gender.

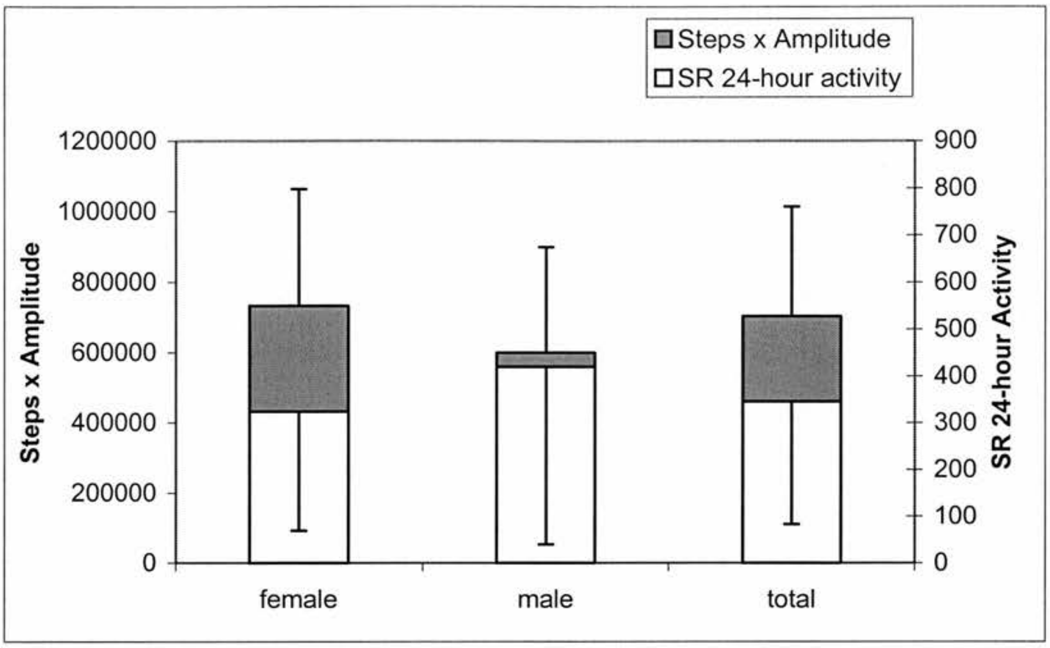


Figure 2.8: Mean monitor records (Steps x Amplitude) and Self-Reported 24-hour Activity (± SD) according to gender.

Question 3: Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Most correlations between potential predictor variables and activity measures did not near significance (see Table 2.8). The two exceptions to this were GSE and Depression, both of which almost significantly correlated with self-report usual activity (rho = .51, p = .06; rho = -.49, p = .07 respectively) such that increased GSE was associated with greater reported activity and increasing Depression was associated with less activity reported. However, as 1 in 10 correlations would be expected to have $p < .1$ by chance, 2 significant findings out of 19 correlations should be treated with caution. As the GSE measure was presented after the self-report usual activity measure, this correlation is associative but not predictive.

Depression and GSE were not associated: rho = -.10, p = .72, suggesting that the relationships of each of these with self-reported usual activity were unlikely to be caused by common variance.

PredictorVariable	Activity Measure	Spearman's rho	P	N
Perceived Control: walking test	Walking test	-.33	.25	14
	SR Walking test distance	-.20	.51	13
Perceived Control: 24-hour activity	SR 24-hours activity	-.07	.82	13
	Monitor (steps x amplitude)	.24	.51	10
GSE[2]	SR usual activity	.51	.06	14
	SR 24-hours activity	.29	.34	13
	Walking test	.32	.26	14
	SR Walking test distance	.41	.17	13
	Monitor (steps x amplitude)	-.11	.76	10
Anxiety	SR usual activity	-.24	.40	14
	SR 24-hours activity	.17	.58	13
	Walking test	.28	.33	14
	SR Walking test distance	.35	.24	13
	Monitor (steps x amplitude)	.34	.34	10
Depression	SR usual activity	-.49	.07	14
	SR 24-hours activity	-.03	.93	13
	Walking test	.18	.53	14
	SR Walking test distance	.19	.54	13
	Monitor (steps x amplitude)	.20	.58	10

Table 2.8: Correlations of the predictor variables Perceived Control, GSE, Anxiety and Depression with activity measures.

As seen in Table 2.9, no correlations between Distance from Union and relevant activity measures were significant. This suggests that this approximation of the distance they would need to travel for work purposes had little impact on records of activity.

Activity Measure	Spearman's rho	P	N
SR usual activity	-.16	.58	14
SR 24-hours activity	-.28	.36	13
Monitor (steps x amplitude)	.16	.65	10

Table 2.9: Correlations between Distance from Union and relevant activity measures.

[2] GSE was presented after SR usual activity and the walking task. The correlations for these measures are therefore associative, not predictive.

Discussion

This pilot study demonstrated that the study procedure was both feasible and acceptable to participants. The activity monitor was not found to be uncomfortable or to cause practical problems for participants. Data collection was highly time-consuming; with having only one activity monitor it was usually only possible to test one participant every 2 days. Time was also lost in waiting for a replacement monitor when the first one broke down. However, a second monitor was acquired for the second monitor study (the Walking Difficulties Study) to minimise these limitations.

Research Question 1: To what extent do self-report, observed performance and continuously monitored assessments of activity correspond?

Correlations between self-report, performance and monitor results were found to be poor, with only the correlation between monitor output and self-reported activity over the monitored period nearing significance (rho = .60, p = .09). Participants underestimated walking task distance, and overestimated time lying, sitting and standing.

LaPorte, Black-Sandler, Cauley, Link, Bayles and Marks (1983), in a sample of 76 women, found only a low correlation between the Paffenbarger Physical Activity Questionnaire (PPAQ) and monitor ratings (LSI monitor) at $r = .23$, $p < .05$ and suggested that the two methods were accessing different aspects of activity, with the LSI monitor assessing frequency of movement but not intensity, whereas intensity of movement is, to some extent, assessed by the PPAQ. However, Ainsworth, Leon et al. (1993) found that PPAQ scores did not significantly correlate with fourteen 48-hour records of activity as assessed by the Caltrac monitor in 78 participants ($r = .19$). In this study, it does not seem surprising that the correlation between activity recorded over a 24-hour period and a 'usual' activity measure should not be highly associated (rho = .08, p = .83), as they are assessing activity over different time spans. In contrast, the association between monitored and self-reported activity over the same time span did near significance. It also seems that the observed walking test could be assessing a different aspect of activity to self-reported and monitored activity performed outside of the laboratory, although participants had difficulty to accurately recall activity performed in the laboratory.

Myers et al. (1993) found that the more closely self-report and performance items were matched, the closer the relationship between scores. Of the main physical activity measures used here, monitored and self-reported 24-hour activity were the only ones measuring the same activity, the other measures being the walking test and self-reported activity over the past week. The different measures seemed to be assessing different aspects of activity, but these findings are tentative given the small sample size.

However, in the laboratory, participants seemed to have difficulty accurately reporting activity, with distance walked being under-reported and time lying, sitting and standing over-reported compared with the researcher's observed records of the same activities. It is unclear whether the difference in error direction is due to the nature of the tasks (active as opposed to passive) or the medium of response (time or distance). Whilst participants were also asked to report walking task time, responses were constrained by the information that they would be walking for between 10 and 15 minutes. Anchoring effects were seen, with most (8) participants estimating 10 minutes, 2 choosing 15 minutes and 2 guessing 12 minutes. Given that the difference in the number of participants over or under-reporting time standing in the field compared with monitor recordings was not significant, these findings could be artifacts of the artificial laboratory setting. In a slightly more natural setting, where participants were allowed to perform any activity they wished, Klesges et al. (1985, 1990) found participants to under-report time spent in sedentary activity and to over-report time spent being active, relative to an observer's ratings.

Overall, these results would suggest that, not only are different measure types accessing different aspects of activity, but that, in the laboratory setting, participants have clear difficulties in reporting their activity in the way it was assessed here and clear biases of over- and under-estimation occur.

Research Question 2: To what extent are measures affected by biases of negative affectivity, social desirability and gender?

No support was found for biasing effects of negative affectivity or social desirability on self-report activity measures: Negative Affectivity neared significance in correlation with only 2 measures, one of which was performance and one self-report (rho = -.49, p = .07 for both). If the negative affectivity bias hypothesis were true, associations between Negative Affectivity and self-report measures would be expected to be stronger than between Negative Affectivity and other measures. Social Desirability did not near significance in association with any activity measure.

Mean monitor activity scores were higher for females than for males whereas mean self-reported activity was higher for males than females. This finding is in the predicted direction (Merrill et al., 1997) but not significant. Given the large variation in scores and extremely small sample size for this comparison (7 women, 2 men), further work must be conducted before any weight can be given to this finding.

Research Question 3: Do the abilities of emotion and control cognitions to predict activity vary according to the measurement method used?

Correlations of control cognitions and emotions with activity measures were, in general, low. However, both GSE and Depression were almost significantly associated with self-reported usual activity (rho = .51, p = .06; rho = -.49, p = .07), with higher Self-Efficacy and lower Depression predicting higher activity, suggesting that further investigation of these variables is merited. It seems unlikely that these results reflect common variance between GSE and Depression as the correlation between these measures was low (rho = -.10, p = .72). Nevertheless, with 2 of 19 comparisons nearing p = .1, these findings could be due to chance. If not chance findings, these results suggest that the psychological variables may be more strongly associated with self-report measures rather than with other assessment types when walking behaviour is the outcome variable.

Limitations

The activity scores given by the Numact monitor could be an underestimation of total activity as activity of the upper body would have gone undetected. This would not be expected to have an impact on associations with the observed walking test scores as both measures assess walking activity. However, relationships with self-reported activity could have been weakened as the PPAQ assessed a wider range of activity. The next studies (the Walking Difficulties, Stroke and Experimental Studies) use measures more specific to walking activity.

Participants' activity was limited as they were asked not to swim or play contact sports, and some participants reported being unable to shower as a deterrent for performing their usual level of activity. Matthews and Freedson (1995), using the Tritrac accelerometer, overcame these problems to some extent by allowing participants to remove the monitor for water activities and taking self-reported scores for those periods. Whilst errors associated with self-report measures could have been introduced, possibly more accurate estimates of usual activity performance could be gained with such a method. However, Conn, Minor, Mehr, and Burks (2000) found that, with 32 participants reporting wearing TriTrac devices daily, only 10 of the total 36 participants had TriTrac data confirming their reports, so having an easily detachable monitor can introduce extra sources of variance. Whilst TriTrac was carried on a belt pouch, the Numact monitor's sensors needed to be taped to the participant and would be more difficult to correctly reattach. As it is, it is possible that Numact recordings were affected by sensors being knocked out of place or inaccurately replaced when tape came unstuck.

Numact was prone to recording sitting as lying if participants were leaning backwards; this will not have affected results here as these scores were not used in analyses. However, Numact sometimes recorded short sitting periods (e.g. 1-2 seconds) whilst participants were standing or walking. The effect of these errors was likely to be relatively small over the whole data collection period. It was also observed (noted also by Bussmann, Tulen, van Heral and Stam (1998)) that the angle of participants' sternum varied to a fair extent, which could affect how likely the monitor was to record 'sitting' as 'lying'. The possibility of differently positioning the chest sensor to avoid this problem is discussed in the Calibration Studies.

It is possible that the PPAQ encouraged participants to underestimate walking. Whilst participants were asked how far they walk each day, they were not instructed to take steps taken in light activity such as movement in the home or workplace into account. As the Numact monitor counts all steps, the mismatch of activity could have increased as a result.

Participants seemed to vary on the amount of effort they put into the walking test. Such variability could have been reduced with an incremental task similar to the incremental shuttle-walking task discussed earlier (Singh, Morgan, Scott et al., 1992). A modified version of the shuttle-walking test has recently been developed for use with young cystic fibrosis patients; 3 extra speed levels were added to the test (Personal communication (2000) from Singh's team at The Glenfield Hospital, Leicester). This test, if used here, may have allowed the use of an incremental test with less of a ceiling effect than the Shuttle Walking Test would have produced.

It was unfortunate that only measures assessing negative affectivity in the 'at the present moment' time frame were included in this study as it did not allow how a participant's affectivity assessed over the same time period as the Self-reported Usual Activity and the 24-hour Activity measures might affect results. That is, it is not possible to suggest from such data whether it is how someone thinks and feels at the moment of recall that determines any bias or whether how someone thinks and feels over the time period for which that information is encoded is important. Watson and Pennebaker (1989) would suggest the latter: in terms of symptom perception, they suggest that people with high negative affectivity attend more to symptoms than those with low negative affectivity. However, whilst Watson and Pennebaker (1989) found state negative affectivity to have the strongest associations with symptom perception, state and trait negative affectivity were found to be strongly related, so it seems unlikely that a clear association between negative affectivity and activity measures would have been missed as a result of this less than ideal methodology.

An error in study design was that the GSE measure was placed after self-reported usual activity and walking test measures were taken as it cannot therefore be discussed as a predictor of activity. However, as no association was seen between GSE and walking test measures, it is unlikely that performing the walking test altered perceptions of self-efficacy. Nevertheless, although the GSE assesses self-efficacy in general and not over activity specifically, it is possible that completing the self-reported 'usual activity' measure could have influenced control cognitions.

It would appear that, overall, the methodology used in this exploratory study was feasible and effective. These methods were therefore taken forward, and improved where appropriate, to assess a larger sample of people with walking difficulties (the Walking Difficulties Study). In the Pilot Study, only one activity monitor was used, leading to a slow rate of data collection and, on technological failure, data collection was delayed whilst awaiting a replacement. A second monitor was therefore acquired for the next study. However, with the arrival of the second monitor, issues of calibration between the two monitors gained importance. These issues are discussed and investigated in Chapter 3: Calibration Studies.

Conclusion

In conclusion, activity measures of different types were not generally found to be closely related. Support was not found for biases of negative affectivity or social desirability, although potential gender effects were in the expected directions. Results did tentatively suggest that the variables depression and self-efficacy may prove to differentially predict scores in self-report, performance and automated monitor results in a larger study: here, trend effects were found with self-report but not other measure types.

The methodology employed was found to be feasible and acceptable to participants.

Acknowledgements

I am grateful to Colin Chandler (University of Northumbria at Newcastle) and Peta Heslop and Jackie Vasseghi (Musculoskeletal Unit, The Freeman Hospital, Newcastle Upon Tyne) for technical support.

CHAPTER 3:
NUMACT ACTIVITY MONITOR CALIBRATION STUDIES

Abstract

Background

The Pilot Study found the positioning of the Numact 'chest' sensor to be less than ideal. The first aim of Chapter 3 was, therefore, to determine whether the sensor could be attached to participants' sides and still give data comparable to that obtained when the sensor was placed on the sternum. A second monitor was acquired to perform the Walking Difficulties Study. Hence the second aim of these calibration studies was to test how reliable the monitors were. Monitor breakdown lead to further monitors being used, increasing the need for such inter-monitor reliability testing.

Research Questions

Study 1: Are readings from monitors comparable when 'chest' sensors are attached either to the chest or to the side?

Study 2: Five different monitors were used over the course of the main study. To what extent do their readings correspond?

Methods

Study 1: Two participants performed a series of standardised activities wearing two monitors simultaneously. Monitor scores (posture, step number, step amplitude) were compared when both 'chest' sensors were attached to the sternum, when one was attached to the side and when both were attached to the side.

Study 2: Standardised activities were performed whilst wearing 2 or 3 monitors concurrently. With the exception of one set, all tests were carried out by the same participant. Monitor scores (posture, step number, step amplitude) of monitors worn simultaneously were compared.

Results and Conclusions

Study 1: Attaching the 'chest' sensor to the participants' sides did not appear to reduce the accuracy of either postural information or step number/amplitude data collected.

Study 2: Inter-monitor reliability was not found to be high. Combined with the tendency of monitors to break down, it would be necessary to reconsider how data from the monitors would be analysed in the Walking Difficulties Study.

Introduction

This chapter addresses two issues concerning the reliability of the Numact monitor: the positioning of the 'chest' sensor and the use of multiple monitors.

As noted in the Pilot Study, the positioning of the Numact 'chest' sensor was less than ideal. Hence, the first study in this chapter aimed to determine whether attaching the sensor to the side instead of the sternum would give results comparable to those acquired using the more conventional placement.

A second issue to be considered was the use of more than one Numact monitor. Data collection in the first study was dependent on a single monitor, restricting testing to one participant in any 24-hour period and causing testing to stop altogether on monitor failure. A second monitor was therefore obtained to conduct the study described in the Walking Difficulties Study and, because of technical failure; these were substituted with new monitors over the duration of the study. It was therefore necessary to calibrate the monitors, to test whether they were recording the same activity. The calibration work reported in the present chapter was performed in parallel with running the Walking Difficulties Study.

Study 1: Sensor positioning.

The main advantage of attaching the 'chest' sensor to the centre of the chest is the closeness of the sternum to the skin, allowing maximum conductance of movement to the accelerometer and lessening differences in recordings between participants with different amounts of body fat. Taping the sensor to the side could add variance due to body fat.

However, it was observed in the pilot study that the angle of the sternum to the horizontal differed greatly between participants; whilst some people have an almost vertical sternum, in others it slopes quite considerably. The risk of 'sitting' behaviour being recorded as 'lying' therefore increases in such individuals. Attaching the 'chest' sensor to the side would be expected to reduce this effect.

Male participants with a large amount of chest hair would have to choose between shaving the area to which the sensor would be taped, putting up with pain caused by the tape pulling the hair, and not participating. As this could cause a sampling bias, the possibility of attaching the 'chest' sensor to participants' sides was worth exploring to lessen this problem.

Trials were therefore performed comparing readings from monitors worn for identical activities but taped to different body locations with two participants differing in build.

Study 2: Inter-monitor reliability

A second issue to be considered in this chapter is inter-monitor reliability in detecting step amplitude, step number and posture.

Walker et al. (1997) videotaped an individual's activity. Two observers watched the tape independently and steps were counted. No systematic error was detected and the variation between observers was similar to that between observers and monitor. Over 110s of recording, the agreement between observation and monitor had a kappa value of 0.933.

A series of nine one-hour trials during which time two monitors were worn simultaneously, with sensors adjacent to each other, are reported (Walker et al., 1997). Five different monitors appear to have been involved in these trials; systems 1 and 2: 6 trials, systems 5 and 6, 6 and 7, 7 and 5: one trial each. A maximum of 10% error at typical activity level is reported. Using the average amplitude of recordings provided, the percentage of agreement between monitors can be calculated (smaller average amplitude / larger average amplitude x 100). Apart from a system1-2 percentage agreement of 88.03%, the minimum remaining percentage agreement was 93.52%; and 3 trials gave agreement of over 98%. The activities undertaken during these trials are not reported, but it would appear that inter-monitor reliability should be high.

A kappa value of .857 is reported for 78 s of concurrent recording using 2 monitors; it would seem this statistic is for agreement in step-counting between the two monitors. The errors are reported to have occurred during small-amplitude shuffle-like steps.

As seen in Figure 2.1, data readouts from the Numact monitor include the product of step number x amplitude, which should be a reflection of energy expenditure. Using a treadmill test, a good linearly correlated relationship was seen between steps x amplitude and METs (Walker et al. 1997).

In the Walking Difficulties Study, it was usual to have the use of 2 monitors. Due to technical problems, it was necessary, at times, to acquire replacement monitors. Summaries of monitors used and participants are found in tables 3.1 and 3.2. It was, therefore, necessary to compare the output of the different monitors where possible to ensure that the high levels of inter-monitor reliability reported by Walker et al. (1997) generalised to the monitors used here.

Participant	Monitor No.	Comments
S2P1	22	
S2P2	22	
S2P3	20	
S2P4	22	Odd readings → M 22 not used further.
S2P5	20	
S2P6	20	
S2P7	20	Chest sensor faulty → M20 not used further
S2P8	14	
S2P9	9	
S2P10	14	File was blank other than the following data: duration (70556), lie s (39004), sit s (17090), stand s (14462). M 14 not used further.
S2P11	11	
S2P12	9	
S2P13	11	Monitor repeatedly detached
S2P14	9	
S2P15	11	
S2P16	9	
S2P17	11	
S2P18	9	Faults → M9 not used further.
S2P19	11	
S2P20	11	

Table 3.1: Summary of monitors used

Monitor	Successful recordings	Unsuccessful recordings
22	2	1
20	3	1
14	1	1
9	4	1
11	6	0

Table 3.2: Summary of monitor success ratings.
NB: These results exclude data lost as a result of sensors becoming detached during the testing period.

Research Questions:

Study 1: Are readings from monitors comparable when 'chest' sensors are attached either to the chest or to the side?

Study 2: Five different monitors were used over the course of the main study. To what extent do their readings correspond?

Study 1

Methods

Design

A series of trials were performed where 2 monitors (M20 and M22) were worn simultaneously, with varying sensor placements. Trials were performed using 2 participants.

Participants

The 2 participants (Participant A and Participant B) were both postgraduate psychology students. Participant B's body mass index was higher than that of Participant A.

Procedure

Set 1: Participant A performed a series of set tasks wearing both monitors.

Tasks: Lie:1 minute*

Sit:1 minute*

Stand:1 minute*

Jumps: 10 high, 10 low, on the spot.

Shuttles (walking laps) in laboratory between obstacles c. 9.5m apart. One shuttle = c. 19 metres. Shuttles performed: 3 brisk walk, 3 slow walk, 3 running.

Brisk and slow walking shuttles repeated without shoes.

* Times aimed for – actual time in each posture varied slightly.

Trial 1: Both 'chest' sensors were attached side-by-side to the sternum, M20 being to the left. Leg sensors were both attached to left leg, M20 below M22.

Trial 2: Monitors' positions were swapped so that M20 'chest' sensor was to the right of M22; M20 leg sensor was above M22.

Trial 3: 'Chest' sensor was attached to the side (left) for M20, to the sternum (centre) for M22. M20 leg sensor was placed below M22.

Trial 4: 'Chest' sensor was attached to the side (left) for M22 and to the sternum (centre) for M20. M22 leg sensor was placed below M20.

Set 2: With the exception of Trial 6 (see below), participant A performed set tasks wearing both monitors. The range of tasks is simpler and closer to what actual participants would be performing whilst wearing the monitor.

Tasks: Lie: 1 minute

Sit: 1 minute

Stand: 1 minute

Walk: 5 minutes (normal walking pace, wearing shoes).

Trial 1: Both 'chest' sensors were attached side-by-side to the sternum, with M20 to the left of M22. On the left leg, M20 was placed below M22.

Trial 2: M20's 'chest' sensor was attached to the side (left); M22's to the sternum. M20's leg sensor was placed below M22 on the left leg.

Trial 3: The 'chest' sensor of M22 was attached to the side (left); M20's to the chest. M20's leg sensor was placed below M22's on the left leg.

Trial 4: Both monitors' 'chest' sensors were attached to the side: M20's to the right side, M22's to the left. M20's leg sensor was placed below that of M22 on the left leg.

Trial 5: 'Chest' sensors were placed as for Trial 4; the position of the leg sensors was swapped such that M20's sensor was above M22's on the left leg.

Trial 6: M20's 'chest' sensor was placed on the side (left); M22's on the sternum. M20's leg sensor was placed above M22's on the left leg. For this trial, instead of the participant performing the set tasks above, the recording was made as Participant A went about her normal daytime activities.

Set 3: The tasks and trials were performed as in Set 2 by Participant B.

Analysis

A score percentage (M20/M22 x 100) was calculated for each of step no., amplitude mean and steps x amplitude. Time spent lying, sitting and standing was not similarly compared as it was not always possible to start monitors running at exactly the same time so such calculations regarding time values could potentially be misleading. However, the actual times were compared between sensor positions. The activity (steps) data was also the most relevant to the Walking Difficulties study.

Results

Set: Trial	Monitor	'chest' sensor position	Duration (s)	Lie (s)	Sit (s)	Stand (s)	Step no.	Amp mean	Steps x Amp
1:1	M20	Sternum	605	90	191	324	153	35.84	5483
	M22	Sternum	601	91	187	323	183	40.25	7365
1:2	M20	Sternum	1753	64	347	1342	553	45.82	25341
	M22	Sternum	1754	71	301	1382	564	52.69	29716
1:3	M20	Side (left)	525	112	80	333	387	56.41	21831
	M22	Sternum	524	77	125	322	356	56.51	20116
1:4	M20	Sternum	486	62	163	261	315	52.05	16396
	M22	Side (left)	586	130	195	261	325	60.57	19684
2:1	M20	Sternum	440	65	72	303	525	41.84	21965
	M22	Sternum	441	64	69	308	529	61.04	32288
2:2	M20	Side (left)	497	67	64	366	434	39.89	17311
	M22	Sternum	495	63	66	366	512	59.02	26636
2:3	M20	Sternum	503	65	71	367	548	44.74	24517
	M22	Side (left)	506	66	68	372	556	48.08	26733
2:4	M20	Side (right)	516	64	85	367	496	34.20	16962
	M22	Side (left)	517	67	82	368	523	40.04	20943
2:5	M20	Side (right)	504	66	67	371	479	34.43	16493
	M22	Side (left)	505	68	66	371	519	40.49	21013
2:6	M20	Side (left)	19227	699	14159	4369	2765	44.75	123730
	M22	Sternum	19207	260	14705	4242	2863	62.08	177731
3:1	M20	Sternum	530	78	65	387	597	31.89	19037
	M22	Sternum	558	105	67	386	579	46.86	27133
3:2	M20	Side (left)	509	70	64	375	551	37.94	20904
	M22	Sternum	534	90	63	372	582	59.19	34447
3:3	M20	Sternum	532	142	9	381	587	36.78	21590
	M22	Side (left)	531	87	62	382	572	45.02	25752
3:4	M20	Side (right)	503	66	65	372	495	35.77	17704
	M22	Side (left)	503	67	63	373	564	35.02	19751
3:5	M20	Side (right)	510	56	65	389	445	42.97	19122
	M22	Side (left)	511	54	66	391	584	42.25	24676
3:6	M20	Side (left)	6017	1438	4048	531	297	23.22	6896
	M22	Sternum	6015	1718	3759	538	301	29.51	8883

Table 3.3: Total duration, lying, sitting and standing time, number of steps, mean step amplitude and step number x amplitude for both monitors for trial sets 1-3.

Notes: 1:1: For both, many steps went undetected, especially with shoes. (when threshold lowered to 4/20, became much closer to 1:2 levels)

1:4: Unclear why duration time differed by 100s; experimenter error whilst starting the monitors suspected.

3:1: M22 – extra 28s at start (E error). Most: lying, and 1 s sit.

3:2: M22 – extra 12s at start. Mixture of lying and sitting.

Posture

From Table 3.3 it can be seen that, in general, posture readings were similar as recorded by the two activity monitors. The chest sensor's role was to distinguish between lying and sitting. Ignoring those trials where the recording start-time differed (trials 1:4, 3:1 and 3:2), the largest differences in posture recordings were seen at trials 1:3, 2:6, 3:3 and 3:6. For trials 2:6 and 3:6 it is not possible to say with certainty which recording was the more accurate as ordinary activity rather than lab tasks were being performed. However, as the activity was recorded during daytime, the lesser lying times are likely to be the more accurate. In trial 2:6, the sensor placed on the side (M20) recorded over 2 ½ times more lying time than did the sternum-placed sensor. However, in trial 3:6, it was the sternum-sensor (M22) that recorded more lying time (1718s compared with 1438). For the lab-task trials, the more accurate lying time recording should be that closest to 1 minute. In trial 1:3 it would appear that M20, the side-placed sensor has over-recorded lying time (112s compared with 77s). However, in trial 3:3, M20 with a sternum-placed sensor has recorded more lying time (142s as opposed to 87s). It should be noted that it was with Participant A that the sternum-sensor appears to be the most accurate and that, with Participant B, the side-sensor appears to be more accurate. The participants' sternums did differ in angle to the vertical (Participant B's was more sloped).

As, for most of the time, the monitors recorded lying times of similar duration, and as, when the times differed, there does not seem to be a systematic bias, it would not appear that attaching the 'chest' sensor to participants' sides will lessen posture recording accuracy overall.

'chest' sensor position	Set: Trial	Step no. Score %	Amp mean Score %	Steps x Amp Score %
Both: Sternum	1:1	83.61	89.04	74.45
	1:2	98.04	86.96	85.28
	2:1	99.24	68.55	68.03
	3:1	103.11	68.05	70.16
M20: Side M22: Sternum	1:3	108.71	99.82	85.28
	2:2	84.77	67.59	64.99
	2:6	96.58	72.09	69.62
	3:2	94.67	64.10	60.68
	3:6	98.67	78.69	77.63
M20: Sternum M22: Side	1:4	96.92	85.93	83.30
	2:3	98.56	93.05	91.71
	3:3	102.62	81.70	83.84
Both: Side	2:4	94.84	85.41	80.99
	2:5	92.29	85.03	78.49
	3:4	87.77	102.14	89.64
	3:5	76.20	101.70	77.49

Table 3.4: Score percentages (M20/M22 x 100) for step number, amplitude mean and step number x amplitude.

If attaching the 'chest' sensor to the side instead of sternum reduces sensitivity to steps, it would be expected that the percentage agreement between sternum- and side-placed monitors would be lower than agreement when both monitors are attached to either the sternum or to the side, and would systematically lead to a lower number of steps.

Step Number

When both sensors were attached to the sternum, score percentages for step number fell between 16.39 and .76 percentage points of 100% (ie the furthest score from 100% is 83.61% and the closest is 99.24% (Table 3.4). When both monitors' sensors were attached to the side, score percentages were within 23.8 and 5.16 percentage points and, when one monitor's sensor was attached to the side and the other to the sternum, score percentages fell within 15.23 and 1.33 percentage points of 100%. Error would not, therefore, appear to be highest when sensors were attached to differing locations.

Mean Step Amplitude

Score percentages for mean step amplitude fell between 31.95 and 10.96 percentage points of 100% when both 'chest' sensors were attached to the sternum and between 14.97 and 1.70 percentage points when both were attached to sides. When one sensor was attached at each position, agreement ranged from 35.9 and .18 percentage points from 100%. Again, agreement does not appear to be very much different from when both sensors are attached to the same position.

Steps x Amplitude

Although this measure of a composite of the step number and amplitude mean discussed above, it is worth also examining the agreement between monitors with this measure as, depending on how step number and mean amplitude measures combine, agreement could potentially be either much higher or much lower than those already discussed.

When both 'chest' sensors were attached to the sternum, score percentages ranged from 31.97 to 14.72 percentage points from 100% and from 22.51 to 10.36 percentage points when both were attached to the sides. Score percentages with measures taken with one sensor attached to the side and one to the sternum ranged between 39.32 and 8.29 percentage points from 100%. Thus, once more, agreement with sensors placed on different places is not dissimilar to agreement where sensors are placed in corresponding positions.

Between-Participant Comparisons

Placing the 'chest' monitor on the side could mean that, with increasing levels of body fat, conductance of vertical acceleration on making a step would lower. In this case, Participant B would have a larger discrepancy between measurements yielded when one sensor is placed on the sternum and the other on the side than would Participant A. However, this is not the case. Taking measurements on the same lab tasks (2/3:2 and 2/3: 3), percentage agreement is 15.23 percentage points from 100% for Participant A and 5.33 percentage points for Participant B when M20 is the side-placed sensor (set 2); in this case, it is Participant B that shows the highest agreement. When M22 is the side-placed sensor (set 3), agreement is 1.44 percentage points for Participant A and 2.62 percentage points from 100% for Participant B. This difference between agreements is in the hypothesised direction but is very small. Thus it would not appear that levels of body fat would have a negative impact on recording accuracy when sensors are attached to sides.

The evidence presented here would suggest that attaching the 'chest' sensor to the participant's side does not, in general, reduce the accuracy of either postural or step number/amplitude data collected. Participant build would also not appear to affect records made when sensors are attached to the side rather than to the sternum where the sensor is closer to the bone.

However, in general, whether monitors were positioned directly comparably (both sternum or both side) or whether they were differently placed, agreement on all measures varied from trial to trial and was consistently less than perfect. Study 2 aims to investigate the extent to which inter-monitor measurements are reliable.

Study 2.

Methods

Design

Two or three monitors were worn simultaneously, as available, whilst a series of set tasks and ordinary activities were performed.

Participants

All tests were performed with Participant A with the exception of Set 3, performed by Participant B.

Procedure

Participants performed tasks as described below, wearing 2 or 3 monitors (whichever monitors were in use at the time) simultaneously. Data from Study 1 is included where participants wore both 'chest' sensors in equivalent positions (ie both 'sternum' or both 'side').

Set 1: M20 and M22.

The tasks performed for Set 1's trials are as for Study 1 (apart from 1:5):

Lie: c. 1 minute

Sit: c. 1 minute

Stand: c. 1 minute

Jumps: 10 high, 10 low, on the spot.

Shuttles in laboratory between obstacles c. 9.5m apart. One shuttle = c. 19 metres. Shuttles performed: 3 brisk walk, 3 slow walk, 3 running.

Brisk and slow walking shuttles repeated without shoes.

The data from trials 1:1, 1:2, are as for Study 1. The details of trial 1:5 are newly described here. For 1:1, 1:2 and 1:5, both monitors' 'chest' sensors were taped to the chest.

Trial 1: M20's 'chest' sensor was placed to the left of M22's. Leg sensors were both attached to left leg, M20 below M22.

Trial 2: Monitors' positions were swapped so that M20 'chest' sensor was to the right of M22; M20's leg sensor was above M22.

Trial 5: Activity was recorded as the participant went about ordinary activities. Monitors were attached as for 1:1.

Set 2: M20 and M22. The data from these trials are as for Study 1. The tasks performed were as follows:

Lie: 1 minute

Sit: 1 minute

Stand: 1 minute

Walk: 5 minutes (normal walking pace, wearing shoes).

Trial 1: Both 'chest' sensors were attached side-by-side to the sternum, with M20 to the left of M22. On the left leg, M20 was placed below M22.

Trial 4: Both monitors' 'chest' sensors were attached to the side: M20's to the right side, M22's to the left. M20's leg sensor was placed below that of M22 on the left leg.

Trial 5: 'Chest' *sensors* were placed as for Trial 4; the position of the leg sensors was swapped such that M20's sensor was above M22's on the left leg.

Set 3: M20 and M22. The equivalent tasks and trials were performed as in Set 2 by Participant B. Trials selected: 1, 4 and 5.

Set 4: M20 and M9

Trial 1: Both 'chest' sensors were attached to the sides (M20: left side, M9: right side). The leg sensor of M20 was placed below that of M9, on the left leg. Laboratory tasks were performed as for Set 2.

Trial 2: Sensors were attached as for trial 4:1. Activity was recorded whilst the participant went about ordinary activity.

Set 5: M9, M11 and M14

Trial 1: 'Chest' sensors were attached to the sides (M9: left side, M11: left side, adjacent (to left of) M9, M14: right side). The leg sensor of M9 was placed below that of M11, in turn below M14, on the left leg. Laboratory tasks were performed as for Set 2.

Trial 2: Sensors were attached as for trial 4:1. Activity was recorded whilst the participant went about ordinary activity.

Analysis

Score percentages (Monitor A / Monitor B x 100) were calculated for each of step number, amplitude mean and steps x amplitude. As for Study 1, time spent lying, sitting and standing was not similarly compared as it was not always possible to start monitors running at exactly the same time so such calculations regarding time values could potentially be misleading. This is not the case for step data as both monitors were always running when the participant started moving. The activity (steps) data was also the most relevant to the Walking Difficulties Study.

Results

Set: Trial	Monitor	'chest' sensor position	Duration (s)	Lie (s)	Sit (s)	Stand (s)	Step no.	Amp mean	Steps x Amp
1:1	M20	Sternum	605	90	191	324	153	35.84	5483
	M22	Sternum	601	91	187	323	183	40.25	7365
1:2	M20	Sternum	1753	64	347	1342	553	45.82	25341
	M22	Sternum	1754	71	301	1382	564	52.69	29716
1:5	M20	Sternum	9954	84	5788	4082	2847	45.38	129185
	M22	Sternum	9954	5	5445	4504	2740	56.46	154698
2:1	M20	Sternum	440	65	72	303	525	41.84	21965
	M22	Sternum	441	64	69	308	529	61.04	32288
2:4	M20	Side (right)	516	64	85	367	496	34.20	16962
	M22	Side (left)	517	67	82	368	523	40.04	20943
2:5	M20	Side (right)	504	66	67	371	479	34.43	16493
	M22	Side (left)	505	68	66	371	519	40.49	21013
3:1	M20	Sternum	530	78	65	387	597	31.89	19037
	M22	Sternum	558	105	67	386	579	46.86	27133
3:4	M20	Side (right)	503	66	65	372	495	35.77	17704
	M22	Side (left)	503	67	63	373	564	35.02	19751
3:5	M20	Side (right)	510	56	65	389	445	42.97	19122
	M22	Side (left)	511	54	66	391	584	42.25	24676
4:1	M9	Side (right)	516	66	72	378	507	41.74	21162
	M20	Side (left)	531	75	86	370	420	37.71	15840
4:2	M9	Side (right)	7631	463	4491	2677	2535	62.03	157238
	M20	Side (left)	7758	613	4585	2560	2055	40.65	83535
5:1	M9	Side (left)	589	74	149	366	520	54.77	28480
	M11	Side (left)	591	71	152	368	539	47.26	25472
	M14	Side (right)	590	113	129	348	487	71.77	34952
5:2	M9	Side (left)	2351	10	1758	583	764	60.25	46028
	M11	Side (left)	2342	7	1710	625	786	53.03	41684
	M14	Side (right)	2351	16	1709	626	829	69.91	57953

Table 3.5: Total duration, lying, sitting and standing time, number of steps, mean step amplitude and step number x amplitude.

Notes: 1:1: For both, many steps went undetected, especially with shoes. (When threshold lowered to 4/20, became much closer to 1:2 levels).

3:1: M22 – extra 28s at start (Experimenter error). Most: lying, and 1 s sitting.

3:2: M22 – extra 12s at start. Mixture of lying and sitting.

Posture.

As it was not always possible to start the monitors at exactly the same time point, posture duration absolute scores are discussed instead of using percentage agreement scores.

Table 3.5 shows that, whilst duration spent lying, sitting or standing was extremely close between monitors in some trials, for others this was not the case. For example, in trial 3:4, monitors M20 and M22 showed almost exact agreement across lying, sitting and standing. However, in trial 1:5,where both monitors recorded identical total duration, M22 detected only 5 seconds of lying compared with M20's 84 seconds. This could reflect the tasks performed: trial 3:4 consisted of laboratory tasks as opposed to the more natural 'ordinary activities' measure taken in trial 1:5. In trial 5:1, M9 and M11 detected very similar levels of lying, sitting and standing whereas M14 appears, in comparison, to have over-estimated lying time (113s as opposed to 74 and 71 s) and underestimated sitting and standing time (129s compared with 152 and 149 s and 348 s compared with 366 and 368s). In contrast, on trial 5:2, M11 and M14 appear to give the closest results for sitting and standing, as compared with monitor M9 (1710s and 1709s compared with 1758s; 625s and 626s compared with 583s).

For the present project, differences between monitors in time spent sitting and lying are not the primary concern as the preferred outcome measure reflected walking activity. However, results could be confounded where standing time differs as Numact only records 'steps' when the monitors are sensing 'standing'.

Monitors compared	Set: Trial	Step no. Score %	Amp mean Score %	Steps x Amp Score %
M20 / M22	1:1	83.61	89.04	74.45
	1:2	98.04	86.96	85.28
	1:5	103.91	80.38	83.51
	2:1	99.24	68.55	68.03
	2:4	94.84	85.41	80.99
	2:5	92.29	85.03	78.49
	3:1	103.11	68.05	70.16
	3:4	87.77	102.14	89.64
	3:5	76.20	101.70	77.49
M20 / M9	4:1	82.84	90.34	74.85
	4:2	81.07	65.53	53.13
M11 / M9	5:1	103.65	86.28	89.44
	5:2	102.88	88.02	90.56
M11 / M14	5:1	110.68	65.85	72.88
	5:2	94.81	75.85	71.93
M 9 / M14	5:1	106.78	76.31	81.48
	5:2	92.16	86.18	79.42

Table 3.6: Score percentages for step number, amplitude mean and step number x amplitude.

Step Number

As can be seen in Table 3.6, agreement between monitors in step number in many trials would appear to be high. Out of the 17 trials, in 6 trials score percentages were within 5 percentage points of 100%. Percentages were out with 10% for 6 trials and by 20 percentage points for 1 trial. However, when comparing 2 monitors over a series of trials, the differences between their recordings were not consistent (e.g. M20 and M22: percentages ranged from within 1 percent to 23.80 percentage points of 100%).

Amplitude Mean

Score percentages for amplitude mean were generally lower than that of step number (see Table 3.6). Percentages were within 5 percentage points of 100% for only 2 trials. Percentages fell out by 10 percentage points for 3 trials and by 20 percentage points of 100% for 6 trials (Table 3.6).

Steps x amplitude

Score percentages were worst by this measure. Percentages were within 10 percentage points of 100% for only 1 trial and out with 20 percentage points for 9 out of the 17 trials; on one trial (4:2), score percentage dropped to only 53.13% (Table 3.6).

Threshold Adjustments

Numact software allows adjustments to be made of both the minimum time which must pass before a second impulse is counted as a step, and of the threshold step amplitude required before a step is detected, as opposed to being ignored as background noise.

Table 3.7 shows the step number, amplitude mean and steps x amplitude scores when the default minimum time between adjacent steps was lowered from 5/20s to 4/20s for trials 1:1, 1:2 and 1:5. As can be seen in Table 3.8, although the percentage steps x amplitude score percentage improved on each trial, this improvement ranged from only .13 to 14.92%. It is of note that agreement did not consistently improve on either step number (improvement on 1 of 3 trials) or amplitude mean measures (agreement improves on no trial), so it seems not unlikely that the agreement improvement seen using the steps x amplitude measure was due to chance. As improvement using the minimum time between steps of 4/20s was neither large nor consistent, and as the study population of people with walking difficulties would be likely to have a slower walking speed than the healthy participants A and B, it was decided to keep the default value of 5/20s.

Set: Trial	Monitor	'chest' sensor position	Step no.	Amp mean	Steps x Amp
1:1	M20	Sternum	464	48.74	22614
	M22	Sternum	434	58.31	25305
1:2	M20	Sternum	596	43.46	25902
	M22	Sternum	567	52.48	29755
1:5	M20	Sternum	2914	44.46	129555
	M22	Sternum	2768	55.96	154898

Table 3.7: Step number, amplitude mean and steps x amplitude with minimum time between adjacent steps set at 4/20s.

Monitors compared	Set: Trial	Step no. Score %	Amp mean Score %	Steps x Amp Score %
M20 / M22	1:1	106.91	83.59	89.37
		(83.61)	*(89.04)*	*(74.45)*
	1:2	105.11	82.81	87.05
		(98.04)	*(86.96)*	*(85.28)*
	1:5	105.27	79.45	83.64
		(103.91)	*(80.38)*	*(83.51)*

Table 3.8 Score percentages for step number, amplitude mean and step number x amplitude with minimum time between adjacent steps set at 4/20s. Figures italicised in brackets are the equivalent scores with default threshold (5/20s)

The step amplitude level could also be adjusted using Numact software. In the Pilot Study, amplitude level was adjusted such that the threshold appeared to be visually identical on the output from two Numact monitors where it had been impossible to calibrate the two monitors against each other. Thus, the same methodology was employed here, for trial 1:1, to determine whether adjusting amplitude threshold to match visually would improve agreement between two monitors.

The results, shown in Table 3.9 and Table 3.10, appear to be conclusive. Although step number score percentage increased from 83.61% to 92.80%, steps x amplitude score percentage decreased from 74.45% to 3.32% and amplitude mean agreement decreased from 89.04% to 3.22%. The default amplitude threshold levels were therefore selected.

Set: Trial	Monitor	'chest' sensor position	Step no.	Amp mean	Steps x Amp
1:1	M20	Sternum	464	48.74	22614
	M22	Sternum	450	1512.55	680647

Table 3.9: Step number, amplitude mean and steps x amplitude with amplitude threshold of M22 adjusted to visually match that of M20 on graphic output.

Monitors compared	Set: Trial	Step no. Score %	Amp mean Score %	Steps x Amp Score %
M20 / M22	1:1	92.80	3.22	3.32

Table 3.10: Score percentages for step number, amplitude mean and step number x amplitude with amplitude threshold of M22 adjusted to visually match that of M20 on graphic output.

Discussion

Although the Numact activity monitor had appeared to be performing satisfactorily in the studies of Walker et al (1997), the data reported in this chapter did not find the device to be reliable. The score percentage results between the monitors give cause for concern, especially where comparability of amplitude mean and step x amplitude are considered. A difference of 20 percentage points between 2 monitors could be manageable if sufficient data were collected with the same 2 monitors to allow inter-monitor effects to be controlled for in analyses. However, this was not the case with these monitors: monitor failure forced the use of 5 monitors to acquire only 16 accurate data sets; of these one data set is unusable because of monitor management problems (sensors coming detached). It would not, therefore, be possible to meaningfully compare monitor records of step number and amplitude across participants. Also, whilst monitor agreement on posture was at times very high, this was not consistent, even when the same 2 monitors were compared.

As a result it was necessary to rethink how the data of the Walking Difficulties Study would be analysed. It had been envisaged that steps x amplitude would be the main monitor measure of interest. This would have been correlated with self-reported and observed measures of activity. Negative affectivity would have been correlated with each of the three measure types and the predictive effects of control cognitions and emotion on these measures would have been investigated. Instead, as can be seen in the Walking Difficulties Study, limited between-participant comparisons were made using the more reliable 'step number' measure. Further analyses were conducted comparing self-reported and monitor data on a within-participant basis (see Walking Difficulties Study for details).

It would have been helpful to have performed more trials in calibrating the monitors. However, because of time constraints on data collection and the rapidity with which monitors broke down, requiring replacement, this was not feasible.

Conclusion

In conclusion, placing 'chest' sensors on participants' sides would appear to have the potential to improve the acceptability of the study procedure to participants without jeopardising existing monitor accuracy or inter-monitor reliability. This methodology was therefore used in the Walking Difficulties Study.

However, inter-monitor reliability, even with both 'chest' sensors attached at equivalent positions is inconsistent and cannot be depended on for subsequent work with the Numact activity monitor. Thus, ways of analysing study data alternative to the original plans needed to be considered.

CHAPTER 4:
COMPARING SELF-REPORT, OBSERVED PERFORMANCE AND CONTINUOUS ACTIVITY MONITORING IN PEOPLE WITH WALKING DIFFICULTIES.

Abstract

Background

The extent to which different types of activity measures assess the same construct in a similar activity domain is unclear. Biases of negative affectivity, social desirability and gender may affect activity ratings, especially on self-report measures. It is also not clear whether the effects of emotion and control cognitions on activity are dependent on the measurement method used to assess activity outcome. This study takes the methods tested in the Pilot Study and investigates these issues in a sample of people with walking difficulties.

Research Questions

1. To what extent do self-report, observed performance and continuously monitored assessments of activity correspond?
2. Are measures affected by biases of negative affectivity, social desirability and gender?
3. Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Methods

20 adults participated after responding to adverts requesting volunteers who had difficulty walking but who were able to walk 100 yards. Laboratory sessions on two consecutive days were completed, separated by 24-hours wearing a continuous activity monitoring device. In the laboratory sessions, self-report and observed performance (session 1) measures of activity as well as measures assessing negative affectivity, social desirability, emotion and perceived control were completed.

Results and Conclusions

Few significant associations were found between activity measure types. It appeared that correlations between measures were more likely to be seen when both measures were assessing either what participants could do or what participants actually did. On standardised measures, participants under-reported distance and over-reported time. No support was seen for biases of gender, negative affectivity or social desirability. Emotion and control cognitions most consistently predicted activity as assessed by observed performance. An association between control cognitions and reporting accuracy seems to have been confounded by associations between self-reported activity and control cognitions.

Introduction

The Pilot Study demonstrated that the methodology of having participants take part in 2 laboratory sessions, including a walking task and separated by 24-hours wearing the Numact activity monitor was both acceptable to the participant and feasible for the researcher. The present study was therefore able to build on these findings and to test the same hypotheses with a sample of people with health conditions: people with walking difficulties. Participants were people who responded to adverts requesting volunteers with difficulty walking but with the ability to walk 100 yards.

To recapitulate, the first area under investigation is the extent to which self-report, performance and monitor measures of activity are related. In particular, work is needed comparing measures assessing the same activity domain. Measures used in this study focus on the walking domain. One problem in the Pilot Study was that of using a self-report measure assessing walking (the activity assessed by the observed performance task and the main activity detected by the monitor) but also assessing overall activity as the activity monitor also detects some non-walking activity. Hence jogging, for example, could be detected by the monitor but would not have been reported in a questionnaire asking only about walking activity. This difficulty was lessened with a sample of people with walking difficulties as walking tended to be the main activity engaged in.

Second, the biasing effects of negative affectivity, social desirability and gender were investigated. From the work of Watson and Pennebaker (1989), it would seem that people high in negative affectivity are likely to report activity less positively than would be seen on other measures of activity. Social desirability is also likely to bias self-report measures: people concerned to impress the researcher may be more likely to over-report activity. Finally, there is some evidence (e.g. Merrill et al. 1997) to suggest that men are more likely to over-report and women are more likely to under-report activity compared with a performance measure. Little support was found for any of these bias hypotheses in the Pilot Study but the small sample size may have limited the findings.

Finally, emotion and control cognitions have been found to predict activity limitations, but whether effects depend on the measure type used is not clear. The Pilot Study found Generalised Self-Efficacy and Depression to predict only self-report activity at trend level. The present study was designed to investigate whether this finding would be replicated in a larger sample of adults with walking difficulties.

In response both to the experience of the Pilot Study and to make the measures more appropriate to people with walking difficulties, different self-report and observed performance measures of activity were selected. Also, more measures of negative affectivity were included. Finally, this study was intended to be a larger and more powerful study than the Pilot Study. This is discussed below.

Self-Report

The measure of self-report usual activity used was the Yale Physical Activity Survey (DiPietro, Caspersen, Ostfeld, & Nadel, 1993). Being designed for an elderly population, it was more suited to the activity levels of this study's population than the PPAQ and walking was a main feature of its questions. With walking being the activity assessed by the observed performance and continuous monitor measures it was important to ensure that it also featured highly on the self-report measure.

Observed Performance

A different walking test was selected to the Pilot Study as the 10-minute walk was likely to be too strenuous for this study's population. As it seemed that participants differed in the amount of effort they put into the 10-minute walk, the Shuttle Walking Test (Singh, Morgan, & Hardman, 1992) was selected as participants constantly have targets with which to keep up. The test is also designed so that participants can stop walking as soon as they start to feel uncomfortable, so participants would not feel under pressure to walk further than they felt able to. As described below, a different lay-out of walking track was also used. Whilst healthy students were able to make the abrupt turns that walking between 2 walls required, participants with walking difficulties were likely to be less agile so a curved walking course was introduced so that the walking test would, indeed, assess walking rather than turning agility.

Negative Affectivity

In addition to asking participants to give negative affectivity ratings for 'the present moment' at each laboratory session, measures of activity for over the past month (the same time frame of the Yale Physical Activity Survey) and over the past 24-hours (the time frame of monitor and further self-report data) were included (see discussion, Pilot Study (Chapter 2)).

Participant Number

This study was intended to have the power to clarify the findings of the Pilot Study. Data collection was planned to involve 40 participants. Unfortunately, as outlined in the Calibration Studies, technical problems were met with. Activity monitors repeatedly failed so that those participants who did take part did not always have complete data sets. Additionally, as the study progressed, it became clear that inter-monitor reliability was not high (the calibration studies were performed in parallel with the present study). Difficulties with recruitment were met: despite repeating the study's advert in local papers and advertising the study in a number of institutions in St Andrews. Because of the difficulties in acquiring and interpreting monitor data, alternative means of recruiting participants were not pursued. As a result, the study was stopped after 20 participants had been tested. The present study is therefore more exploratory than definitive.

Research Questions:

1. To what extent do self-report, observed performance and continuously monitored assessments of activity correspond?

2. Are measures affected by biases of negative affectivity, social desirability and gender?

3. Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Methods

The study was approved by the University of St Andrews School of Psychology Ethics Committee (see Appendix G).

Design

Participants with walking difficulties were assessed on self-report and observed walking activities before a 24-hour period of continuous activity monitoring. Self-reported activity over the 24-hour period was recorded after this period.

The extent to which different measurement methods correspond was examined cross-sectionally: observed walking task performance was compared with self-reported walking task performance, 24-hour monitor records were compared with self-reported activity over the same time period and observed walking task performance was compared with self-reported 'usual' activity and 24-hour monitor records.

Relationships between the potential sources of bias (negative affectivity, social desirability and gender) and activity measure were also examined cross-sectionally.

The effects of perceived control and emotion on activity were assessed predictively. The predictive relationships of these psychological variables with the different activity measures were compared.

Participants

20 participants, 11 female and 9 male, were recruited using advertisements in the local newspaper and supermarket inviting people who had difficulty walking but who were able to walk 100 yards to contact the researcher (see Appendix F.ii). Participants were paid £15 on completion of the tasks. It was not possible to pay £25 as for the pilot study because paying a sum greater than £15 would have caused complications for participants receiving state benefits.

Measures

Activity: Self-Report.

Self-reported usual activity: Modified Functional Limitations Profile (FLP) (Pollard & Johnston, 2001): Ambulation subscale (Appendix A.v)

The Functional Limitations Profile (Patrick & Peach, 1989) is a British version of the American Sickness Illness Profile (SIP). The scale consists of 136 items grouped into 12 categories, one of which is 'Ambulation'. The items take the forms of statements describing activities such as 'I do not use stairs at all'. Participants are asked if they agree or disagree with each item. Where participants agree, they are asked if it is due to their health. Each item has a weighting reflecting the severity of the item's limitation: the higher the score the greater the limitation. The sum of checked items' weightings is used to calculate percentage limitation scores for each category (sum of checked items weightings in category divided by maximum possible category score, multiplied by 100) and for the overall disability score (sum of all checked items weightings divided by sum of all weightings, multiplied by 100). Repeatability over 2 days was tested with a sample of 30 disabled patients. The magnitude of changes of category scores ranged from –2.68 to 6.5 (Charlton, 1989). Unfortunately the change scores for each category are not reported. The Functional Limitations Profile was criticised by Pollard and Johnston (2001) for problems such as its illogical scoring (for example, someone unable to walk at all could achieve a lower score than someone with limited walking ability as the latter would be able to agree with more items) and its length. The method of scoring was argued to be incompatible with the underlying theoretical framework (Pollard & Johnston, 2001). A new method of scoring (the Modified Functional Limitations Profile) was therefore proposed whereby the category's score was the weighting of the item with the highest weighting checked within a category. Rearranging the items from the most severe limitation to the least severe allowed the questionnaire to be shortened: when participants agreed with an item they could move on to the next category rather than having to respond to all 136 items. In a sample of 101 disabled patients, Johnston and Pollard (1996) found the scale to have internal reliability (Cronbach's α) of .71 (.81 excluding work category). Frank et al. (2000) found Cronbach's αs of .89 and .87 with stroke patients interviewed on two occasions one month apart. Hence the Modified Functional Limitations Profile would appear to be an internally consistent measure. The Ambulation Subscale was used by itself for this study as items relate to walking. The interviewer reads a list of statements from the most serious 'I do not walk at all' to the least serious 'I walk more slowly'. Participants are asked to decide whether or not each statement described them on that day. These statements have weightings reflecting level of disability; higher weightings indicate more serious disability. When participants agree with a statement they are asked if it is due to their health and, if so, no more statements are read. The weighting for that item is recorded as the subscale's score. Higher scores indicate greater limitation.

Self-reported usual activity (past month): Yale Physical Activity Survey, Section 2 (YPAS) *(DiPietro, Caspersen, Ostfeld and Nadel, 1993)* (Appendix A.vi)

The YPAS was developed to assess the current physical activity of an older population. It contains two sections: the first consists of a checklist measuring time spent in work, exercise and recreational activities. The second section assesses participation in different types of activity: vigorous activity and leisurely walking (walks lasting at least 10 minutes but not sufficiently strenuous to cause large increases in breathing, heart rate, leg fatigue or perspiration – such activity would be classified as 'vigorous activity') (over the past month) and moving about, standing and sitting (on a typical day during the past month). As the second section relates more closely to the activity domain of interest (walking, standing, sitting), only this section was included in the present study (see Appendix A.vi). For vigorous activity and leisurely walking, participants are asked to report the number of times and for how long the activities were performed over the past month. For moving, standing and sitting, participants are asked to report about how many hours per day, on average, they carried out that activity over the past month. The test's authors use a categorical response system whereby participants select a category reflecting the frequency and time of their activities, rather than making a free response. However, here, as 2 scoring methods were used (see below), participants were asked to make a free response, which was recorded both as itself and placed by the researcher in the authors' categories to enable the measure to be scored as the authors intended.

Following DiPietro et al. (1993), a frequency score was multiplied by a duration score for vigorous activity and leisurely walking; for moving about, standing and sitting, a score was given according to the amount of time spent doing that activity per day. For each exercise category, these scores were multiplied by a weighting factor based on the relative intensity of that activity: 5 (vigorous), 4 (leisurely walking), 3 (moving about), 2 (standing), 1 (sitting). The final index was the sum of these 5 individual scores.

A second scoring method ('Modified YPAS') was developed to create a score more consistent with the energy-expenditure based PPAQ used in the Pilot Study. Instead of using DiPietro et al.'s (1993) categorical responses of activity frequency and duration, actual recalled frequency and duration of activity were multiplied. These scores were multiplied by intensity codes (kcal/minute). Intensity codes were taken from Section 1 of the YPAS where scores for Section 2 activities were available; other codes were collected from Ainsworth, Haskell et al.'s (1993) compendium of energy cost codes: 6 (vigorous, DiPietro et al 1993), 3.5 (leisurely walking, DiPietro et al. 1993), 2.3 (moving about, Ainsworth, Haskell et al. 1993), 1.2 (standing, Ainsworth, Haskell et al. 1993) and 1 (sitting, Ainsworth, Haskell et al. 1993). The YPAS also contains an item asking how many flights of stairs were climbed up each day. Whilst the authors' scoring system ignores this item, it was included in this modified scoring method and given a value of 4kcal per flight of stairs climbed (as for the PPAQ, (Paffenbarger et al., 1978), used in the Pilot and Experimental Studies).

DiPietro et al. (1993) performed a test-retest reliability study with 76 older adults (56 female, 12 male), with the YPAS being administered on two occasions 14 days apart. Test-retest correlations for the individual activity dimensions were each highly significant ($p<.0002$): $r = .42$ (sitting), .48 (standing), .49 (moving about), .48 (leisurely walking) and .61 (vigorous activity). For the total summary index, the test-retest correlation was .65. Although significant, these correlations are moderate in size. It could be that a shorter time interval than 2 weeks is necessary for a questionnaire assessing activity over the past month to give high test-retest correlations as activity could have changed. The only activity for which the score significantly differed between time points was vigorous activity ($p = .02$). As some research has indicated that such activity may be better recalled than lower intensity activity (Wareham & Rennie, 1998), it seems unlikely that this finding was due to responding problems such as recall, but perhaps peoples' activity differed between the two time points.

DiPietro et al. (1993) also report a validation study. Twenty-five participants (11 female, 14 male) completed the YPAS immediately before skinfold measurements and treadmill testing with oxygen consumption were performed. Participants also wore Caltrac motion sensors in the home environment for 2.5 days. Total weekly activity time and weekly kilocalorie expenditure (from YPAS Section 1) were found to correlate with Section 2 scores for moving about (Spearman's rho = .44, $p = .03$; rho = .47, $p = .02$) and inversely marginally significantly with sitting scores (rho = -.38, $p = .05$; rho = -.35, $p = .08$). Section 2's total score was significantly associated with estimated maximal oxygen uptake (rho = .58, $p = -.004$) and percentage body fat (rho = -.43, $p = .03$). The Section 2 vigorous activity index also correlated with estimated maximal oxygen uptake during treadmill testing (rho = .6, $p = .003$). The Section 2 total score and the leisurely walking index approached significance in their correlations with Caltrac motion counts (rho = .37, $p = .09$; rho = .31, $p = .18$). The moderate correlations reported above suggest that, whilst YPAS Section 2 scores are associated with a range of activity-related assessments, the measure does show divergent validity. Of the Section 2 measures, the moving about and sitting index were found to be significantly inversely associated (rho = -.44, $p = .03$). No figures are reported for other inter-item correlations.

Hence, for the YPAS, adequate test-retest reliability and validity have been demonstrated. It would, however, be beneficial to see results from a larger sample as it is unclear here whether the marginal significance level shown in some correlations (such as Section 2 total score and Caltrac counts: rho = .37, p = .09) is due to low power or to a genuine lack of relationship.

Self-reported walking task performance

Participants were asked to estimate first, how far they had walked in doing the Shuttle Walking Test (see below) and second, for what length of time they were walking (Appendix A.iii)

Self-reported time spent sitting, standing and completing questionnaires.

At the start of the first laboratory session, participants were asked to sit and to stand for a short time, until the researcher asked them to stop. Participants were later asked to estimate for how long they were sitting and standing for this task (Appendix A.iii). The time period selected for both sitting and standing was 45 seconds because it was less likely to be chosen as an anchor point than 1minute; being asked to stand for longer than 1 minute could have caused discomfort for some participants. Participants were told they were being asked to do this in order to calibrate the monitor (the task's secondary purpose). Participants were also asked to estimate the time spent completing questionnaires over the whole of the first study session. The researcher recorded time completing questionnaires with a stopwatch.

Self-reported 24-hour activity

Two records of self-reported 24-hour activity were taken: a modified version of the YPAS and a 24-hour record.

The Modified YPAS took items from Section 2 of the YPAS (DiPietro et al., 1993) (see above) but altered the wording such that all items referred to the previous 24-hour period (see Appendix A.vii). It was scored according to the method developed for this study (see above) rather than that the authors recommended for the original YPAS as some of their scoring categories made little sense for reporting 24-hour activity compared with activity over the period of a month.

The 24-hour record was developed from the KIHD 24-hour Total Physical Activity Record (Salonen & Lakka, 1987). The interviewer worked through the past 24 hours in half-hour blocks with the participant, asking, for each half hour, into which category of activity their behaviour fitted. Instead of using Salonen and Lakka's (1987) activity categories (9 categories ranging from 1: sleeping, resting supine to 8: very strenuous activities and 9: other), the categories vigorous activity, walking, moving about, standing, sitting and lying were included so that the measure would be more comparable with the YPAS and with Numact monitor records. Whilst the record could be scored using the kcal weights discussed above, the measure was developed in order to compare time spent in different activity as reported by the Numact monitor with time reported spent in those activities across the day. Therefore, 'Minutes Active', combining 'Vigorous Activity', 'Walking' and 'Moving About' was the primary measure taken. Records were not taken for half-hour blocks for which participants were in the laboratory. As time spent in the laboratory varied between participants, to gain a score better reflecting the relative proportion of active time the final score was calculated as follows:

$$\boxed{\text{(total Minutes Active / number of half-hour blocks) x 48}}$$

with 48 being the total possible number of blocks in the 24-hour period. Minutes Active in each individual half-hour block was also recorded for each participant.

Activity: Observed Performance

The Shuttle Walking Test (Singh, Morgan & Hardman, 1992)

It was noted in the Pilot Study that participants seemed to vary in the amount of effort they put into the 10-minute walking task. A walking test was therefore chosen here that would set goals for participants in order to lessen this effect and to gain an indication of participants' maximal performance. It was decided to base the assessment on the progressive 10-metre Shuttle Walking Test (Singh, Morgan & Hardman, 1992). The test designers advise laying out a 10m course with marker cones set .5m in from the course ends to avoid the need for any abrupt change in direction. The test begins with a triple tone sounding from a cassette tape provided by the test designers (Singh, Morgan & Hardman, 1992). At this signal, the participant starts walking from one end of the course. Thereafter, tones emitted by the cassette tape indicate when the participant is to reach the opposite end of the course. The speed of walking increases after each minute. Thus, three ten-minute shuttles are performed in the first minute (Level 1), four in the second minute (Level 2), five in the third minute (Level 3), and so on up to Level 12. The end point of the test is reached when either the participant wishes to discontinue, for example if they feel they can no longer continue or feel in any discomfort, or when the participant can no longer maintain the required speed. Walking test score is the distance walked in completed shuttles.

Test-retest reliability was assessed with 10 chronic airways obstruction patients (Singh, Morgan, Scott et al. 1992). Participants performed the test on 3 occasions one week apart. Performances on each occasion were very highly correlated ($r = .98$, .99 and .98 for time 1and time 2, time 1and time 3, time 2 and time 3 respectively). An increase in distance walked from time 1 to time 2 was seen (mean distance walked in trials: 345, 376 and 378m for trials 1, 2 and 3 respectively); suggesting a small training effect; little improvement occurred in performance in comparing trials 2 and 3 (mean change: 2m). These results that the most accurate indication of participants walking ability could be achieved after a practise trial, on the second attempt. In the present study's protocol, it would be necessary for participants to perform the test either twice in one laboratory session or twice in two days, re-doing the test on their return to the laboratory on the second day. There was concern that in a population of people with walking difficulties, requesting a strenuous walking task performance twice in a short amount of time would increase the study's work load considerably and it was very possible that fatigue could counter any effects of training. In terms of analyses to be performed, the impact of training effects as reported by Singh, Morgan, Scott, et al. (1992) would be minimal. Walking test performance scores were to be compared with self-reported scores of the same activity; the accuracy of self-report can be considered whether or not training effects occur. Other analyses were correlations between different measure types. The very high correlations between shuttle walking scores at times 1, 2 and 3 (time 1-2: $r = .98$, time 1-3, r = .99) suggest that correlations should be minimally affected by any training effects.

For 15 participants, Shuttle Walking Test performance was compared with performance in the 6 minute walking test (Singh, Morgan, Scott et al., 1992). A moderate correlation was seen between the two test scores (rho = .68, no significance level given). Higher maximal heart rates were found with the Shuttle Walking Test than the 6 minute walking test, suggesting that the shuttle-walking test was more effective in obtaining an indication of participants' maximal ability.

For this Walking Difficulties Study, the track design was modified because of space constraints: a track was laid out in a figure-of-eight pattern, such that the distance from the outer edge of one 'circle' was 10m from the opposite point on the other circle (each peripheral point being marked with a cross, see Figure 4.1). The use of a 10-metre space between 2 walls (as for the pilot study) was considered, but participants would have had to make abrupt turns and there was a risk that the task would then assess turning agility rather than walking ability.

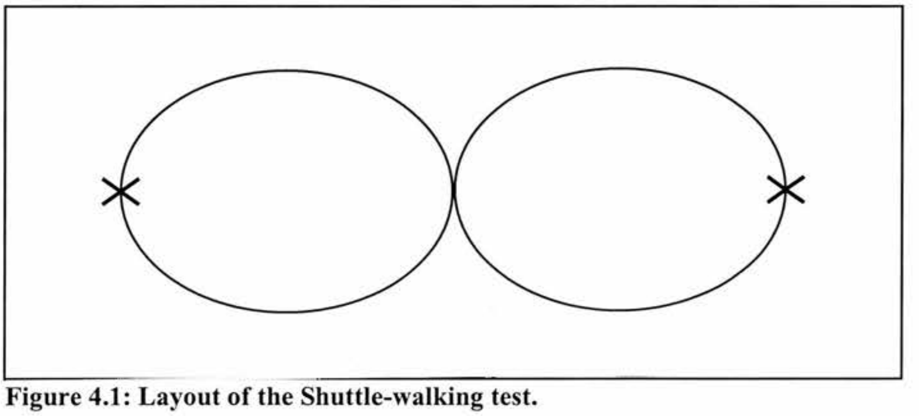


Figure 4.1: Layout of the Shuttle-walking test.

The test authors recommend that, if the participant is further than 0.5m from the end marker when a tone sounds then the test be discontinued. When the participant is within 0.5m of the marker it is recommended that the participant be allowed to walk one further shuttle to attempt to regain the distance. As this allows very little margin for misjudging required walking speed, a more generous error margin was allowed here. If participants were within 1.5m of the cross at the tone they were allowed to continue but asked to stop if they were unable to make up the distance to within 0.5 metres of the next cross; if some distance was gained but not within 0.5 metres of the cross, participants were allowed one more shuttle to catch up. This was found to be satisfactory: on a number of occasions, participants who would have been forced to discontinue using Singh et al's recommendation were able to make up the distance, giving a more accurate representation of what they were able to do.

Participant instructions are provided on the shuttle-walking test audiocassette. However, as this was written for chronic obstructive pulmonary disease patients, the text was adapted and read by the researcher instead. Participants were instructed that their objective was to walk for as long as possible, keeping to the speed indicated by the tones on a cassette. On hearing a tone, they were to aim to be at an extreme point on the figure-of-eight track, marked by a cross; on hearing the next tone they were to aim to be at the cross marking the opposite point on the track. They were informed that the required walking speed would initially be slow, allowing 20 seconds to reach the opposite end of the track, but that it would increase. The experimenter walked alongside participants for the first minute to aid them in judging the walking speed required.

As recommended by Singh, Morgan and Hardman (1992), the distance walked in the completed shuttles was recorded. In addition, actual total distance was recorded as many participants ended the task partway through a shuttle, and total time for which participants were walking was noted. The actual total distance is used in analyses for this study.

Unaware Observed Walking Speed

Where possible, participants' walking speed as they left the laboratory was measured using 2 photoelectric detector transmitter/receiver pairs 8m apart as described in detail in the Experimental Study (Chapter 6). This measure was used with less success than in the Experimental Study because the slower walking speed allowed more time for passers by to trigger the photoelectric detectors and some were accompanied by caregivers, on occasion being pushed in a wheelchair. Participants were not explicitly informed beforehand that this measure would be taken. However, where a measure was successfully taken, all participants were later fully debriefed and their consent for this data to be used was requested. However, as time to walk between the two photoelectric detectors was successfully recorded for only 12 participants it was not included in analyses.

Continuous Activity Monitoring

The Numact Activity Monitor

The Numact activity monitor (see Pilot Study, Chapter 2 for details) was worn by participants from the beginning of the first laboratory session until the end of the second laboratory session. Step Number and Steps x Amplitude were recorded for the entire time the monitor was worn.

To match the self-reported 'Minutes Active' score (see above), scores were taken for the half-hour blocks for which participants were out of the laboratory. Scores reflecting the proportion of active time over the recording period were calculated:

Score (Step Number or Steps x Amplitude = (raw step number or steps x amplitude / number of half-hour blocks) x 48.

Because of the inter-monitor reliability limitations of the Steps x Amplitude measure (see the Calibration Studies), only the Step Number score was used in between – participant analyses.

Activity scores for each individual half-hour block (both Step Number and Steps x Amplitude) were also recorded.

Potential Bias Sources

Negative Affectivity: The Positive and Negative Affect Schedule (Watson et al., 1988) (PANAS, Appendix B.i)

PANAS is described in the Pilot Study (Chapter 2). A limitation to the Pilot Study was that negative affectivity was only measured 'at the present moment' and so it was not possible to investigate the effects of negative affectivity during a time period measured by activity assessments. For example, would negative affectivity over the past month be associated with self-reported activity over the past month? Negative affectivity was therefore assessed using a range of time frames in the present study. PANAS can be worded to fit the time scale required and was used 4 times to refer to 4 different time periods. First, participants were asked to report 'to what extent you feel this way right now, that is, at the present moment' immediately before completing self-report usual activity measures. Second, participants were asked to report 'to what extent you have felt this way during the past month' to gain an estimate of negative affectivity corresponding to the interval that self-report usual activity measures referred to. Third, participants were asked about negative affectivity at the present moment immediately before completing self-reported 24-hour activity measures. Finally, participants were asked 'to what extent you have felt this way during the past 24 hours' to estimate negative affectivity during the period of wearing the activity monitor.

Watson et al. (1988) tested the scale with the time instructions: at the present moment, today, past few days, past few weeks, past year and in general on large samples mainly of undergraduate students (n ranging from 586 to 1002). Internal reliability (alpha) was acceptable in all cases, ranging from .84 to .87. No systematic differences were found between student data and data from non-student adult populations. Test-retest reliability was assessed over an 8-week period with 101 undergraduate students. Reliabilities tended to increase as the time frame lengthened from 'moment' (.45) to 'year' (.60) and 'in general' (.71), suggesting that test-retest reliability was appropriate to the time frame. Taking the time frames most relevant to this study, test-retest Negative Affectivity reliabilities were .45 (moment), .39 (today) and .48 (past few weeks). External validity of the measures was suggested by correlations with other commonly used measures of psychological distress: Hopkins Symptom Checklist ($r = .74$ (PANAS time frame: past few weeks); $r = .65$ (PANAS time frame: today), Beck Depression Inventory ($r = .58$, PANAS time frame: past few weeks) and STAI State Anxiety Scale ($r = .51$, PANAS time frame: past few weeks).

Social Desirability: Short form (Reynolds, 1982) of the Marlowe-Crowne Scale (Crowne & Marlowe, 1960) (Appendix B.iii)

The Marlowe-Crowne Scale is described in the Pilot Study. For the present study, a short-form of the Marlowe-Crowne Scale was selected to minimise the load for participants completing an already long interview. Following factor analysis, Reynolds (1992) selected items with factor loadings of at least .40. Further items were added according to item with total scale correlation. Reynolds' (1982) 13-item short-form has a correlation with the standard Marlowe-Crowne Scale of .93 ($p<.001$) (the scale was completed by 608 undergraduate students), and internal reliability of .76 (Kuder-Richardson Formula 20). The scale was administered by the interviewer, who read each statement, e.g. 'I am always courteous, even to people who are disagreeable'. True or false responses were recorded. Higher scores indicate higher social desirability.

Psychological Predictors: Emotion and Perceived Control

Emotion: Anxiety and Depression: Hospital Anxiety and Depression Scale (HADS) (Zigmond & Snaith, 1983) (Appendix C.i)

This scale is described in the Pilot Study (Chapter 2).

Generalised Self-Efficacy: Generalised Self-Efficacy Scale (GSE) (a short version of this scale is outlined in Jerusalem and Schwarzer (1992); for further details see Johnston, Wright and Weinman, 1995). (Appendix C.ii)

The scale is described in the Pilot Study (Chapter 2).

Perceived Control over Shuttle Walking Test

As for the Pilot Study, 5 items assessing perceived control over walking task performance were developed from Conner and Norman (Conner & Norman, 1996) (see Appendix C.iii). Participants gave responses on 7-point scales. For example, for the item 'Whether I do or do not perform well on the walking task is entirely up to me' was rated from 1 = strongly disagree to 7 = strongly agree. The Pilot Study found internal reliability for the scale was found to be .83 with one item excluded. The item in question, ' I would like to perform well on the walking test but don't really know if I can' was kept in the questionnaire initially, with the expectation that it might be necessary to exclude that item's responses.

Perceived Control over 24-hour Activity

Similar perceived control measures were developed to assess perceived control over the 24-hour period for which participants would be wearing the monitor. (Appendix C.iv) Questions were phrased to ask about control over being 'very active in the next 24 hours', as for the Pilot Study. The Pilot study found internal reliability to be .84 (one item excluded). Similar to the perceived control over walking task performance scale, the excluded item was 'I would like to be very active in the next 24 hours but I don't really know if I can'.

Intention to be Active (Appendices D.i, D.ii)

Items assessing intention to perform well in the walking task and intention to be active over the 24-hour period were included in the questionnaire but were not analysed.

Demographic Variables

Participants' age, gender, educational level and condition causing their walking difficulty was recorded. Information on the condition was gained from two sources: first, participants were asked what the condition causing the walking difficulty was. Second, after receiving consent from participants, GPs were contacted and asked to return a form indicating the cause of the participant's walking difficulties. Where a discrepancy arose, the GP's diagnosis was accepted.

Procedure

Adverts inviting people who had difficulty walking but who could walk 100 yards were placed in local newspapers, supermarket, public library and community centre (see Appendix F.ii). People expressing an interest were invited, by telephone, to attend a preliminary meeting with the researcher. At this meeting, the study's procedure was explained and the Numact activity monitor was demonstrated to participants to ensure that they were fully aware of what the study involved and what they would be required to do. Participants also received an information sheet (see Appendix F.v). Twenty-three participants attended such a meeting; three did not continue to the study sessions. One did not continue because she would not personally benefit from participating; two discontinued because of difficulties in finding a convenient time to participate. For other participants, dates were arranged to take part in the study itself. Two consecutive days where participants could attend at the same time on each day were selected.

At the first actual study session, the procedure was once more briefly run through and signed consent to take part in the study was received from the participant (consent forms in Appendix F.v). Participants were also asked if they would consent to the experimenter contacting their GPs and informed that not consenting to this would not affect their being able to take part in the study. The Numact activity monitor was attached to participants at the start of the session so that they could become accustomed to wearing it whilst in the laboratory. Participants were then asked to sit and to stand for a short time (45 seconds) each. They were told that this was in order to calibrate the monitor. A secondary purpose was to allow the researcher to compare actual time sitting and standing at this point with participants' recalled time when asked later in the study. The first part of the questionnaire (see Table 4.1) was given by interview, and participants were asked to perform the Shuttle Walking Task. Participants then completed the second part of the interview (Table 4.1), ending the first session. Participants were asked to go about their normal activities as much as possible over the next 24 hours, although participants were asked to keep the monitor dry, so bathing, showering and swimming were not permitted.

Participants returned to the laboratory at the same time the following day. They completed the third and final part of the questionnaire and were paid £15 for participating. Where an 'unaware observed' walking speed measure was successfully taken, participants were debriefed and asked for consent to use this measure.

The order in which questionnaire scales were given is shown in Table 4.1. After the first ten participants the position of the GSE and Social Desirability scales was changed so that they were placed before any measures of activity to better allow their predictive value to be assessed.

Participants 1-10	**Participants 11-20**
Part 1 A. HADS B. PANAS – now C. FLP } Self-report usual D. YPAS activity E. Perceived control – walking task F. Intention – walking task	Part 1 A. HADS B. GSE C. Social Desirability D. PANAS – now E. FLP } Self-report usual F. YPAS activity G. Perceived control – walking task H. Intention – walking task
SHUTTLE WALKING TASK	
Part 2 G. PANAS – month H. GSE I. Social Desirability J. Self-reported distance and time during walking task, time sitting and standing at start of study. K. Perceived control – 24 hours L. Intention – 24 hours M. Self-reported time spent completing questionnaires.	Part 2 I. PANAS – month J. Self-reported distance and time during walking task, time sitting and standing at start of study. K. Perceived control – 24 hours L. Intention – 24 hours M. Self-reported time spent completing questionnaires.
24-HOURS WEARING MONITOR	
Part 3 N. PANAS – now O. Modified YPAS } Self-report P. 24-hour record 24-hour activity Q. PANAS – 24 hours R. Education level S. Gender, date of birth.	*Part 3* N. PANAS – now O. Modified YPAS } Self-report P. 24-hour record 24-hour activity Q. PANAS – 24 hours R. Education level S. Gender, date of birth.

Table 4.1: Presentation order of questionnaire.

Analysis

Monitor data:

Monitor data was graphed for each participant over the 24-hour period along with comparable activity self- reports. An example of this work is given in Figure 4.2; the charts for all participants are found in Appendix E.iii.

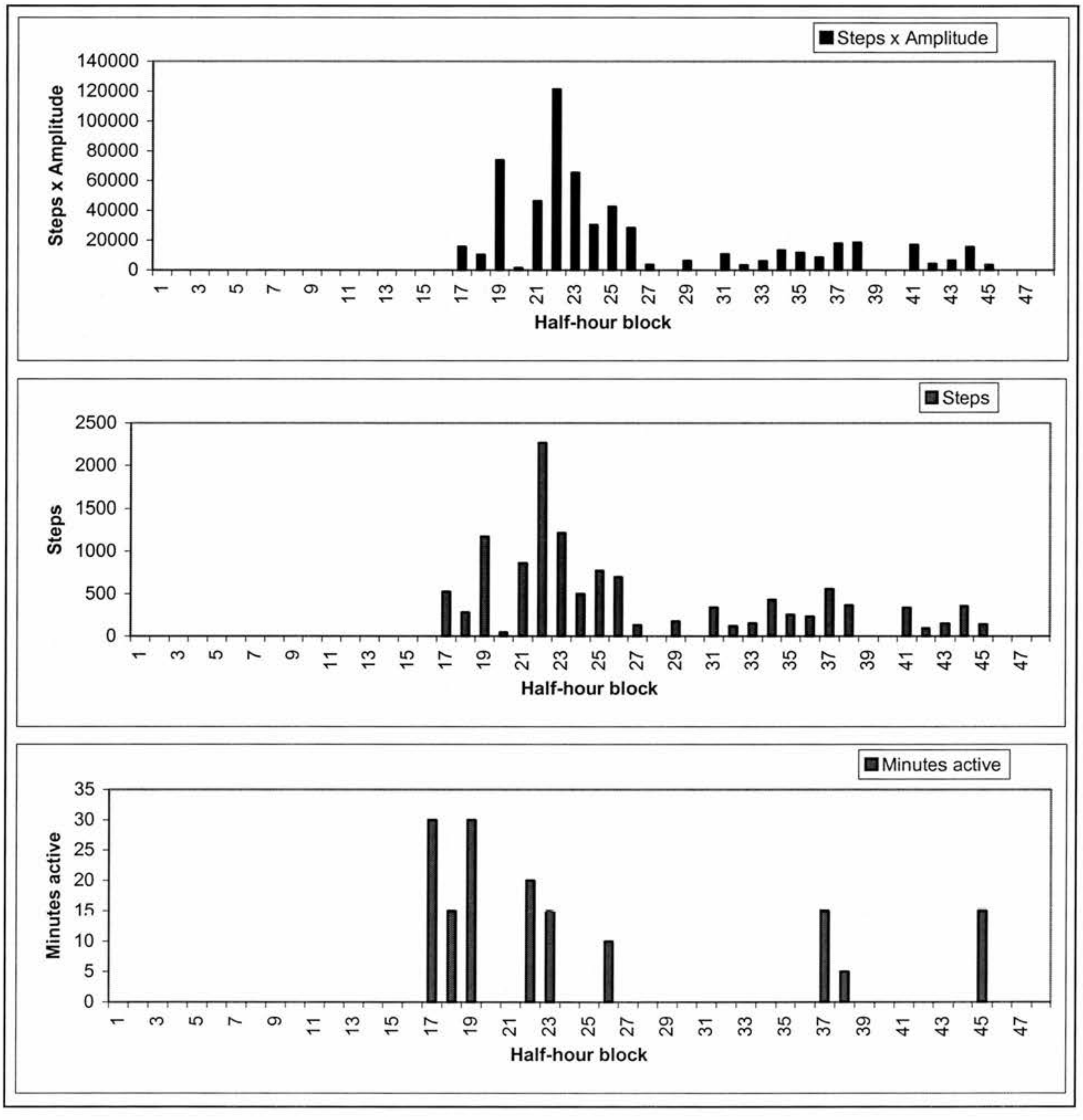


Figure 4.2: Continuously monitored (Steps x Amplitude and Steps) and self-reported (Minutes active) measures of activity in ½ hour time blocks over a 24-hour period.

Two measures reflecting the relationship between self-report and monitor records were devised. For each participant, each half-hour block was reduced to a binary report of whether or not activity was reported by a) the participant and b) the monitor (frequency tables can be found in Appendix E.ii). From this data were calculated:

i. *Kappa*: a measure of the agreement between monitor and self-report. Kappa values can range from 0 to 1; the higher the score, the closer the agreement.

ii. *Bias*: This measure was designed to assess the extent to which participants over- or under-report activity compared with the monitor. A measure of bias would be useful in partial correlations: would associations of potential bias sources (Negative Affectivity and Social Desirability) and predictors (Emotion and Control Cognitions) with Kappa (Agreement) change on partialling out bias?

$$\text{Bias} = \frac{\Sigma\ (\text{blocks self-reported activity})}{\Sigma\ (\text{blocks monitor-reported activity})}$$

Scores ranged between 0 and 1 (no participants over-reported activity. Lower scores reflect greater 'under-reporting' of the participant compared with the monitor.

Research Question 1: To what extent do self-report, observed performance and continuously monitored assessments of activity correspond?

Between Participants

Cross-sectional correlations (Spearman's rho) between measures of activity were calculated.

Directly Comparable Scores

Where observed and self-reported scores were available on identical measures (Shuttle Walking Test distance and time, time spent standing and sitting at the researcher's request), observed and self-reported scores were compared using Wilcoxon's Signed Ranks Test. The Sign Test was used to determine whether the number of participants over estimating was significantly different to the number under estimating.

The Sign Test was also used within participants to test whether participants reported activity in fewer half-hour blocks than did the monitor.

Within Participants

For each participant, correlations (Spearman's rho) were calculated between half-hourly measures of Step Amplitude, Step Number and Minutes Active. Significance levels are reported, but these levels will be inflated as the half-hourly activity scores are not independent.

Research Question 2: Are measures affected by biases of negative affectivity, social desirability and gender?

Cross-sectional correlations (Spearman's rho) between measures of social desirability or negative affectivity and the activity measures were performed. Correlations of Social Desirability and Negative Affectivity with Bias and Kappa were calculated with both Spearman's rho and Pearson's r prior to partial correlations being performed. If significant, correlations of Social Desirability and Negative Affectivity with Kappa (Agreement) would be recalculated with Bias partialled out.

Mann-Whitney U was used to test for differences in scores between male and female participants on self-reported and observed shuttle walking measures and on 24-hour activity assessments.

Research Question 3: Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Emotion and control cognitions measures were correlated (Spearman's rho) with activity measures. Correlations of emotion and perceived control measures with Bias and Kappa were calculated with both Spearman's rho and Pearson's r prior to partial correlations being performed (if bivariate correlations with Kappa were significant). The aim was to test whether correlations of psychological predictors with Kappa (Agreement) would persist on partialling out Bias. Partial correlations were to be performed where psychological predictors were associated with Kappa.

Results

Descriptives

Participants' mean age was 70.45 years (SD: 10.97), ranging from 50 to 87 years. 11 were female, 9 male. The sample seemed to be fairly highly educated: 6 participants left school at the minimum school leaving age, without qualifications (although 3 later completed professional qualifications), 3 had Standard/O-level qualifications, 2 had undergone further education (e.g. A-levels, NVQs) and 9 had higher education qualifications (3 of these had PhDs).

Participants' walking difficulties were caused by a range of conditions: 10 had osteoarthritis (accompanied by other conditions in 3 cases), 1 had rheumatoid arthritis, 3 had suffered strokes. Chronic degenerative disc disease, peripheral vascular disease, muscular sclerosis, peripheral neuropathy and 'tiredness of legs and weak ankle' (no GP's letter for this one participant) and cerebral palsy caused the walking difficulties of the remaining participants. Where participant and GP reports of the condition causing the walking difficulty differed (3 participants), GP reports were accepted.

Variation in N across measures is due to the following problems: One participant had difficulty with many of the self-report measures. Others had difficulty answering occasional items. Where data was missing, a participant's score for that scale was excluded. One participant's (P1) observed shuttle-walk time score is missing due to experimental error. As observed shuttle-walk time and distance would be expected to be highly associated, where a significant effect was found with one score and not the other, the analysis was re-run for shuttle-walk distance excluding P1's score to determine whether differences could be caused by P1 in this small sample. The low N for 'unaware observed' walking speed was due to a combination of experimental error (3 occasions), other people in the corridor triggering the stop clock mechanism (2 occasions) the participant leaving in a wheelchair (1), the participant stopping for a chat part way along (1) and no record of reason (1). No time 2 data was available for one participant who had been unable to return to the lab because of health problems.

The monitor and self-report Minutes Active data of 12 participants are shown in Table 4.2. These are the 12 participants for whom there was comparable monitor and self-report data. Monitor failure occurred for 4 participants and no self-report data was available for 2 participants (one was unable to return on the second laboratory day; the other had difficulty responding to the questionnaire). For two participants, some monitor data was missing. One participant's monitor repeatedly detached until the participant gave up reattaching it, leaving 3 hours with no monitor data. The other participant's monitor failed to record data for the first 3 hours. The data of these two participants will only be used for within-participant comparisons. In one instance, the monitor detached in the night and was re-attached when the participant awoke. As it is likely that participant and monitor records of 'lying' for this period are valid, this participant's data was retained.

Self-reported (Minutes Active) and monitor data were used for the 12 participants for whom both measures were available as the Minutes Active data was collected for comparison with the monitor data (and vice versa).

Descriptive statistics for all measures are shown in Table 4.2.

Scales were screened for outliers (z scores > 3.29, (Tabachnick & Fidell, 1996)). Only one outlying score was found, on Negative Affectivity (NA) – 24 hours. Analyses were performed with and without the outlier. As results were unchanged on excluding the outlier (in terms of significance at the $p < .10$ level (Bonetti et al., 2001)), reported results include the outlier.

Skew and kurtosis statistics are shown in Table 4.3. SPSS version 9 (1998) states that a skew value greater than 1 generally indicates that scores differ from a normal distribution. Of the 28 measures, 13 skew and 15 kurtosis scores were greater than 1. The following measures had particularly poor skew values (greater than 2): YPAS: 2.16 (compared with YPAS modified scoring: -.41); Estimated Time Sitting: 2.34; NA – 24 hours: 2.51. NA and YPAS were showing floor effects: participants reported low NA and low levels of activity.

Variable	Measure	Mean	SD	α	N
Activity	*Self-Report:*				
	- FLP	87.20	16.01		20
	- month (YPAS)	28.47	21.19		19
	- month (modified YPAS)	1078.11	220.33		19
	- shuttle-walk distance (m)	66.79	62.94		19
	- shuttle-walk time (s)	294.75	160.77		20
	- time sitting (s)	130.29	148.18		17
	- time standing (s)	90.28	86.51		18
	- 24-hour activity	1030.64	396.63		18
	- minutes active (24 hour)	227.27	163.09		12
	Observed Performance:				
	- Shuttle-walk score	99.50	87.75		20
	- Shuttle-walk distance (m)	134.59	78.20		20
	- Shuttle-walk time (s)	106.28	87.04		19
	- Unaware observed (s)	10.44	3.21		12
	Monitor:				
	- Step Number	7956.33	4458.68		12
	- Steps x Amplitude	187446.37	155205.53		12
	Self-report – Monitor relationship				
	Kappa (Agreement)	.27	.20		12
	Under-report Bias	.40	.26		12
Negative Affectivity	NA now – time 1	11.85	3.50	.83	20
	NA now – time 2	11.63	3.15	.83	19
	NA - past month	17.68	5.95	.79	19
	NA - 24 hours	12.68	4.45	.79	19
Social Desirability	Social Desirability	7.80	2.89	.69	20
Emotion	Anxiety	4.80	4.07	.81	20
	Depression	4.30	2.66	.63	20
	HADS	9.10	4.38	.74	20
Perceived Control	GSE	30.55	4.38	.77	20
	Perceived Control – walking task	21.68	5.26	.56	19
	Perceived Control – 24 hour activity.	20.79	5.74	.57	19

Table 4.2: Descriptive statistics. Mean, standard deviation, Cronbach's α (where appropriate) and N for each measure.

The following measures performed similarly poorly in terms of kurtosis (kurtosis > 2): YPAS: 5.44 (compared with modified scoring: -.50), Estimated Time Sitting: 5.95; NA – 24 hours: 7.57; Kappa: 3.58. Also with kurtosis scores above 2: Estimated Time Standing: 2.43, NA now – time 1: 2.07; NA now – time 2: 2.07; Anxiety: 2.25; GSE: 2.55.

It should be noted that, when later asked to recall for how long the researcher had asked participant to sit and stand at the start of the experiment, many participants had difficulty recalling this part of the study. This could have resulted in these data being anomalous, and data were only gained for 17 participants. For example, one participant reported having a 'blank spot' before estimating 10 minutes of sitting time (actual time: 45s). This participant (P15) was recognised as an extreme outlier on an SPSS boxplot although, at $z = 3.17$, it did not meet the $z > 3.29$ outlier criteria. A very high standard deviation (148.18) is seen for this measure (Table 4.2).

Variable	Measure	Skew	Kurtosis
Activity	*Self-Report:*		
	- FLP	.21	-.93
	- month (YPAS)	2.16	5.44
	- month (modified YPAS)	-.41	-.50
	- shuttle-walk distance (m)	.93	-.24
	- shuttle-walk time (s)	1.12	.79
	- time sitting (s)	2.34	5.95
	- time standing (s)	1.79	2.43
	- 24-hour activity	1.22	1.16
	- minutes active (24 hour)	1.10	1.26
	Observed Performance:		
	- Shuttle-walk score	.87	.18
	- Shuttle-walk distance (m)	.87	.29
	- Shuttle-walk time (s)	.01	-1.14
	- Unaware observed (s)	.82	.05
	Monitor:		
	- Step Number	.67	-1.31
	- Steps x Amplitude	1.21	1.92
	Self-report – Monitor relationship		
	Kappa (Agreement)	1.61	3.58
	Underestimation bias	.95	.24
Negative Affectivity	NA now – time 1	2.40	5.03
	NA now – time 2	1.80	2.07
	NA - past month	.54	-.46
	NA - 24 hours	2.51	7.57
Social Desirability	Social Desirability	-.86	1.05
Emotion	Anxiety	1.40	2.25
	Depression	.58	-.71
	HADS	.39	-.93
Perceived Control	GSE	-1.03	2.55
	Perceived Control – walking task	-.70	-.46
	Perceived Control – 24 hour activity.	-.37	-.84

Table 4.3: Skew and kurtosis scores.

As the YPAS modified scoring method yielded scores that performed better psychometrically and is more comparable with the adaptation of the YPAS for

24-hour activity than the YPAS (standard scoring), only the modified scores will be analysed further.

Non-parametric statistics were used where possible as some measures deviated from a normal distribution.

Reliability

Bryman and Cramer (1997) advise that Cronbach's α be at or greater than .8 for any measure. As can be seen from Table 4.4, the Negative Affectivity measures were at or approaching this value. At .69, Social Desirability showed moderate reliability. Anxiety's reliability (.81) was somewhat higher than Depression's disappointing score of .63; the overall HADS reliability was moderate (.69)

At .77, GSE's reliability approached the levels recommended by Bryman and Cramer (1997). However, the other Perceived Control measures fared poorly, with alpha values of .56 and .57 for Perceived Control – walking task and Perceived Control – 24 hour activity respectively. For these very similar measures, elimination of either item 1 or item 3 (see Box 1) lead to an increased alpha (from .56 to .65 and .67 for items 1 and 3 respectively, Perceived Control – Walking Task and from .57 to .71 and .61 for Perceived Control – 24 hour Activity). Item 3 was also found to reduce reliability in the Pilot Study.

In the Pilot Study, data was analysed without the reliability-reducing item. In this study, however, as 2 items would be removed, reducing this 5-item scale by almost half, they were retained. The consequent low internal reliability of this measure means that significant relationships between this and other measures are less likely to be detectable.

As a different order of questionnaire was used for the first 10 participants to that for the remainder, Mann-Whitney U was performed to detect whether the scores differed significantly between the first and last ten participants on those measures affected by the order change. As can be seen in Table 4.3, none of the measures differed significantly with questionnaire order. Mann-Whitney U for NA now –

time 1 does, however, approach significance (p = .08). Data suggest that scores for the participants with the original order of items were lower than with the second order. This scale was originally placed immediately after HADS whereas, in the second order, it was also preceded by GSE and Social Desirability. It could be that these measures caused participants to have a higher level of state negative affectivity. Johnston (1999) found participants completing the HADS following other questionnaires were significantly more anxious than those who completed the HADS first in the sequence. Analyses performed using NA now – time 1 were performed with the total measure and with scores from the original and secondly ordered questionnaires separately to ensure that any relationships affected by changing the order of questionnaire be explored.

Measure	Mann-Whitney U	P	N
NA now – time 1	26.0	.08	19
FLP	46.5	.80	20
YPAS	42.0	.84	19
SR month	38.0	.60	19
Social Desirability	49.5	.97	20
GSE	41.5	.53	20
Perceived Control - walking task	41.5	.78	19
NA – month	29.5	.21	19
Shuttle-walk score	39.0	.44	20
Shuttle-walk distance	32.0	.32	19
Shuttle-walk time	38.5	.39	20

Table 4.4: Mann-Whitney U tests comparing the first 10 and last 10 participants for whom the order of the above measures varied.

Question 1. To what extent do self-report, observed performance and continuously monitored assessments of activity correspond?

Between-participant correlations.

Intra-class correlations of self-report activity measures are shown in Table 4.5a. It can be seen that very few self-report measures correlate with each other. Higher shuttle-walk distance reports are associated with lower disability ratings on the FLP (rho = -.65, p = .00) and SR 24-hour activity is associated with Minutes Active over the same 24-hour period at trend level (rho = .52, p = .08). These results indicate that, even in the same measurement type domain, measures are assessing different aspects of activity.

The correlation between observed distance and time on the Shuttle-Walk Test was very high: rho = .96, p = .00. This would be expected, as the extent to which the two measures can differ is very limited.

	1	2	3	4	5
1. FLP					
2. SR month	-.29 (.24)				
3. SR 24-hour activity	.28 (.26)	.17 (.51)			
4. SR Minutes Active	.20 (.53)	.52 (.08)	-.01 (.97)		
5. SR shuttle-walk distance	-.65 (.00)	-.02 (.95)	-.12 (.65)	-.24 (.47)	
6. SR shuttle-walk time	.24 (.30)	-.13 (.60)	-.08 (.75)	.11 (.73)	-.37 (.12)

Table 4.5a: Correlations (rho) between Self-Report (SR) activity measures (significance level in parentheses).

Measures	Spearman's rho	P (2-tailed)	N
Self-Report (SR) – Observed Performance			
SR month – Shuttle-walk score	.06	.80	19
SR month – Shuttle walk distance	.05	.84	19
SR month – Shuttle-walk time	-.12	.65	18
FLP – Shuttle-walk score	-.62	.00	20
FLP – Shuttle-walk distance	-.60	.01	20
FLP – Shuttle-walk time	-.64	.00	19
SR distance – Shuttle-walk distance	.63	.00	19
SR time – Shuttle-walk time	-.16	.53	19
Observed Performance – Monitor			
Shuttle-walk score – Step Number	.23	.48	12
Shuttle-walk distance – Step Number	.27	.39	12
Shuttle-walk time – Step Number	.17	.63	11
Self-Report – Monitor			
SR 24 hour activity – Step Number	-.27	.40	12
SR Minutes Active – Step Number	.36	.25	12
SR month – Step Number	.86	.06	12
FLP – Step Number	.26	.42	12

Table 4.5b: Correlations between activity measures.

Table 4.5b shows that significant correlations were found between self-report FLP and shuttle-walk performance scores and between self-reported and observed shuttle-walk distance. Higher levels of FLP disability were associated with less distance walked and higher reported shuttle-walk distance was associated with higher observed shuttle-walk distance. The correlation between self-reported activity over the past month and monitor-rated step number neared significance (rho = .86; p = .06), although reports of activity over the 24-hour period for which the monitor was worn were not significantly associated with step number. Correlations between observed performance and monitor ratings were not significant.

Directly comparable scores

Observed and self-reported scores for distance walked in the shuttle-walking test, time walking in the test, time sitting and standing at the researcher's request were compared using Wilcoxon's Signed Ranks Test. As shown in Table 4.6, self-report and observed measures were found to be significantly different on all 4 measures (Z range: -1.95 to -3.00; P range .00 - .05). Figure 4.3 shows that the self-reported mean distance walked is less than that observed. Self-reported time spent performing activities was greater than that observed for time walking, time standing and time sitting.

	Distance walked	Time walking	Time standing	Time sitting
Z	-2.09	-2.98	-2.65	-1.95
P	.04	.00	.01	.05

Table 4.6: Wilcoxon's Signed Ranks Test statistics. Comparing self-report and observed measures of distance walked and time walking in the Shuttle Walking Test and standardised time standing and sitting.

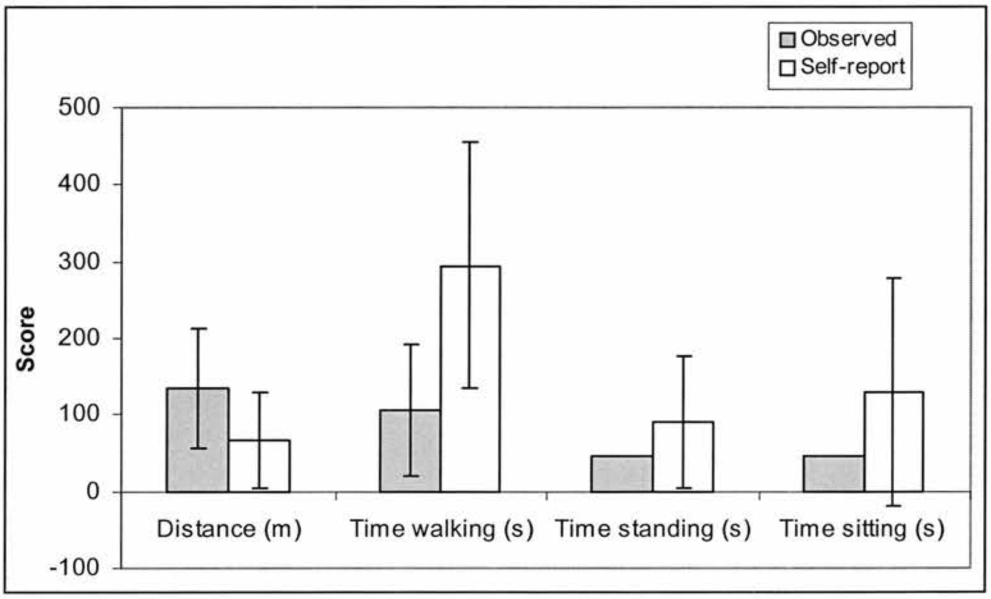


Figure 4.3: Mean (±SD) observed and self-report scores on measures of distance walked and time walking in the Shuttle Walking Test and standardised time standing and sitting.

Participants' estimates of time and distance walked in the Shuttle Walking Test and of standardised time sitting and standing were compared with actual scores on these items using the sign test. Frequencies of over-and under-reporting are shown in Figure 4.4. It can be seen that more participants under than over-reported distance walked in the Shuttle Walking Test but more participants over-than under-reported time spent walking, sitting and standing. However, using the sign test, these differences were not found to be significant, although significance was neared for the overestimation of time spent walking (p = .06) and time sitting (p = .08). Significance levels for distance and time standing were .36 and .24 respectively.

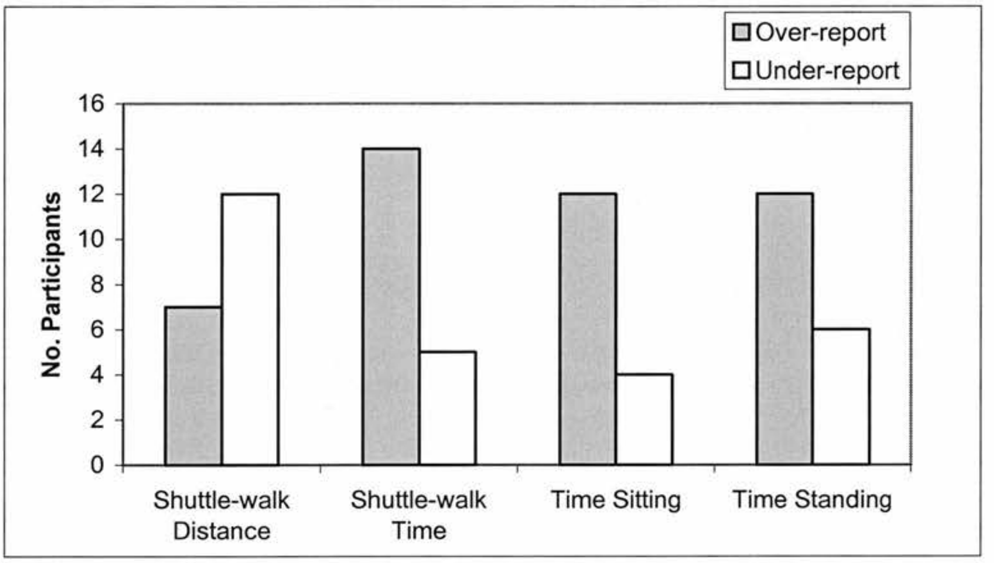


Figure 4.4: Number of participants over- and under-reporting distance and time in the Shuttle Walking Test and standardised time standing and sitting.

Within-participant correlations

Table 4.7 shows that, for all participants, Step number x Amplitude and Step Number measures correlate very highly (range: rho = .97 to 1.00). However, correlations between self-reported minutes active and monitor measures vary greatly, from .03 (P6) to .63 (P11). Kappa values also vary greatly, from .00 (P3, 11, 14) to .90 (P6). An apparently anomalous reversal of the placing of P6 and P11, with P6 having the lowest correlations but the highest kappa and vice versa with P11 is seen. Scatterplots (Figures 4.5 and 4.6) show that P6 reported very little variability in self-reported activity over blocks of time, removing the possibility of finding a correlation whereas, in contrast, P11 reported much more variable activity.

Bias scores also show variability between participants, ranging from .13 (P2 and P17) to .90 (P3). Lower scores reflect greater under-reporting of activity in half-hour blocks compared with monitor reports. Sign tests were also performed (see Table 4.7) to test whether participants significantly reported activity in fewer half-hour blocks than did the activity monitor. Results are shown in Table 4.4. It can be seen that all participants except for P3, P6 and P18 had significantly more half-hour blocks where activity was reported by the monitor but not in self-report than where activity was self-reported but not recorded by the monitor (see Appendix E.ii for self/monitor report frequency tables).

Participant	Measures	Spearman's rho	P (rho)	Kappa	P (kappa)	Bias	P (sign test)	No. blocks
P1	A	1.00	*.00*	.25	.05	.85	.01	48
	B	.39	*.01*					45
	C	.39	*.01*					45
P2	A	.98	*.00*	.13	.13	.13	.00	48
	B	.29	*.06*					46
	C	.30	*.04*					46
P3	A	.97	*.00*	.77	.00	.90	.38	48
	B	.61	*.00*					45
	C	.59	*.00*					45
P5	A	.98	*.00*	.28	.01	.45	.00	48
	B	.41	*.00*					47
	C	.41	*.00*					47
P6	A	1.00	*.00*	.02	.90	.43	.08	48
	B	.03	*.84*					45
	C	.03	*.84*					45
P8	A	1.00	*.00*	.40	.01	.70	.15	41
	B	.52	*.00*					40
	C	.52	*.00*					40
P11	A	.99	*.00*	.43	.00	.48	.00	48
	B	.61	*.00*					46
	C	.63	*.00*					46
P12	A	.99	*.00*	.18	.03	.29	.00	48
	B	.54	*.00*					46
	C	.52	*.00*					46
P13	A	.99	*.00*	.13	.13	.15	.00	36
	B	.43	*.01*					34
	C	.36	*.04*					34
P14	A	1.00	*.00*	.37	.00	.38	.00	48
	B	.60	*.00*					46
	C	.63	*.00*					46
P15	A	1.00	*.00*	.20	.03	.17	.00	48
	B	.46	*.00*					46
	C	.45	*.00*					46
P16	A	1.00	*.00*	.29	.02	.45	.01	48
	B	.36	*.01*					46
	C	.38	*.01*					46
P17	A	1.00	*.00*	.16	.05	.13	.00	48
	B	.34	*.02*					45
	C	.32	*.03*					45
P19	A	.99	*.00*	.14	.07	.18	.00	48
	B	.23	*.14*					45
	C	.23	*.12*					45

Table 4.7: Within participant correlations, kappa values, bias scores and sign test results. Number of half-hour blocks for which data was available is reported. Sections of monitor data were missing for participants S2P8 and S2P13 hence their data is not involved in between-participant analyses.

A = Steps x Amplitude – Step number

B = Steps x Amplitude – Minutes Active

C – Step number – Minutes active

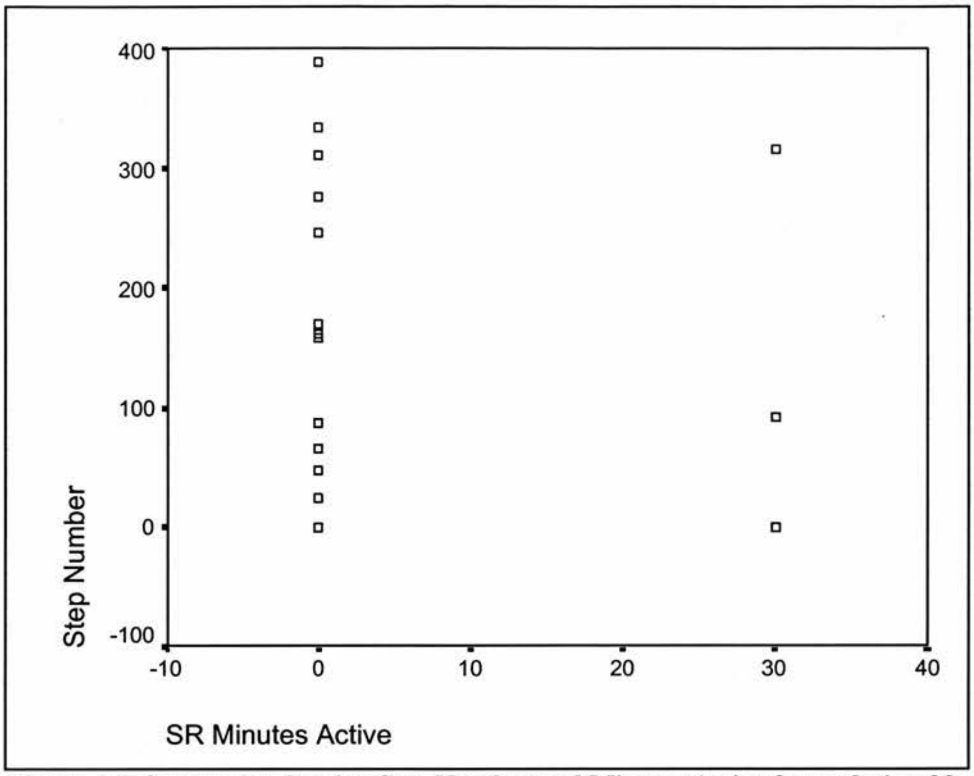


Figure 4.5: Scatterplot showing Step Number and Minutes Active for each time block (P6).

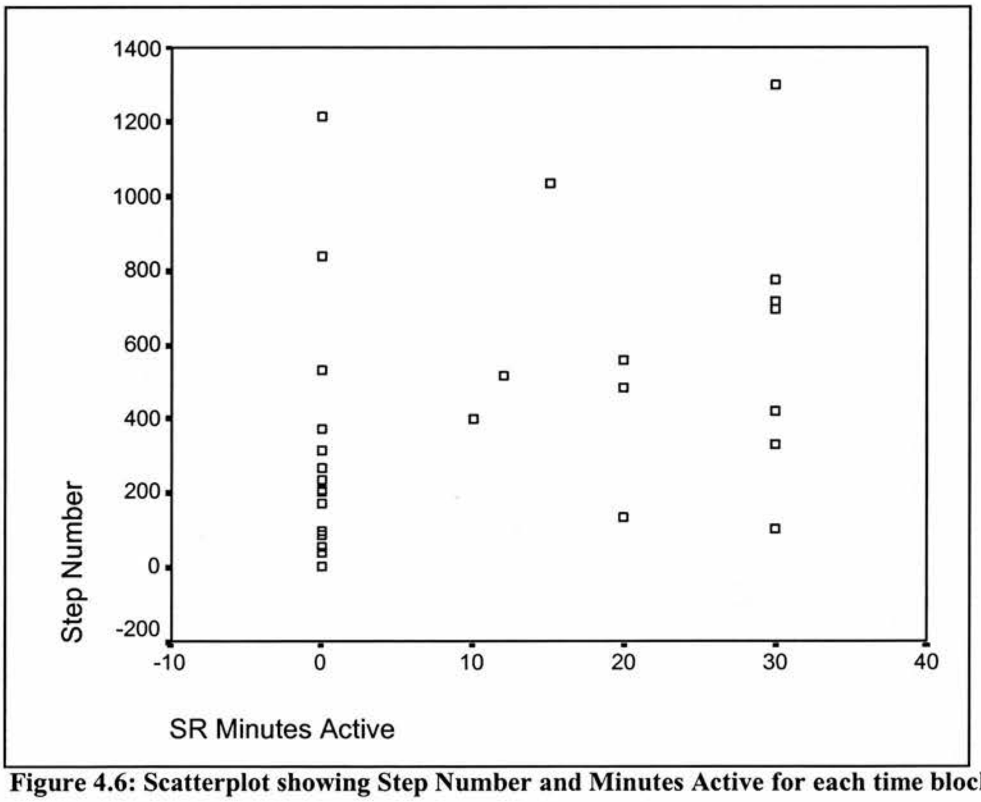


Figure 4.6: Scatterplot showing Step Number and Minutes Active for each time block (P11).

Question 2: a) Are measures affected by biases of negative affectivity and social desirability?

Of the correlations between Negative Affectivity and activity measures, only the correlations of NA 'now, time 1' with Self-Report shuttle-walk distance and Observed shuttle-walk time were significant (rho = -.52, p = .02; rho = -.50, p = .03, Table 4.8), with participants lower in Negative Affectivity reporting further distance walked and being observed to walk for a longer time. As an association was seen with observed time but not distance for the shuttle-walk test, the correlation between observed shuttle-walk distance and NA 'now, time 1) was re-run excluding P1 (missing from observed shuttle-walk time data). The correlation maintained non-significance (rho = -.34, p = .16). The correlation between NA – month and FLP approached significance (rho = .39, p = .1). It should be noted that, in Table 4.8, 22 correlations are reported. By chance, 2.2 of these would be expected to be significant at the p < .1 level. Three correlations were found to be significant at this level and so the findings should be treated with caution.

As the difference in NA 'now, time 1' scores between participants completing questionnaires in Order 1 and Order 2 had approached significance, the correlations involving NA 'now, time 1' were repeated separately for each order group (Table 4.9). Whilst none of the correlations approached significance with Order 1, two correlations neared significance when the questionnaire was in Order 2: FLP (rho = .50, p = .08) and SR shuttle-walk distance (rho = -.62, p = .07). Both are in the expected direction. The lack of significant correlations for the Order 1 group could, in part, be explained by the measure's distribution. As can be seen in Figures 4.7a and 4.7b, whilst floor effects are seen for both groups, for group Order 1, very little variation is seen at all, making any relationship between Negative Affectivity and activity impossible to detect.

Measures correlated	rho	P	N
NA now, time 1 with:			
SR – month	.37	.12	19
FLP	.30	.19	20
SR shuttle-walk distance	-.52	.02	19
SR shuttle-walk time	-.06	.81	20
Observed shuttle-walk distance	-.27	.25	20
Observed shuttle-walk time	-.50	.03	19
NA month with:			
SR – month	.03	.92	18
FLP	.39	.10	19
NA now, time 2 with:			
SR – 24 hours	.29	.25	18
SR Minutes Active	.-.30	.35	12
NA – 24 hours with:			
SR – 24 hours	.24	.34	18
SR Minutes Active	-.03	.92	12
Step Number	-.09	.79	12
Social desirability with:			
SR – month	-.15	.54	19
FLP	-.05	.83	20
SR shuttle-walk distance	-.14	.57	19
SR shuttle-walk time	-.17	.49	20
Observed shuttle-walk distance	-.10	.67	20
Observed shuttle-walk time	.07	.78	19
SR – 24 hours	.31	.21	18
SR minutes active	.09	.77	12
Step Number	-.14	.68	12

Table 4.8: Correlations of Social Desirability and Negative Affectivity with activity measures.

Measures correlated	rho	P	N
NA now, time 1, order 1 with:			
SR – month	.42	.23	10
FLP	-.12	.74	10
SR shuttle-walk distance	-.08	.83	10
SR shuttle-walk time	.00	1.00	10
Observed shuttle-walk distance	.27	.46	10
Observed shuttle-walk time	.00	1.00	9
NA now, time 1, order 2 with:			
SR – month	.38	.31	9
FLP	.50	.08	9
SR shuttle-walk distance	-.62	.07	9
SR shuttle-walk time	-.06	.88	10
Observed shuttle-walk distance	-.31	.39	10
Observed shuttle-walk time	-.49	.15	10

Table 4.9: Correlations of Negative Affectivity with activity measures according to questionnaire order.

Because all Negative Affectivity measures suffered from floor effects to some extent, correlations were re-calculated using the total HADS score as a measurement of negative affectivity. Whilst not encompassing all potential aspects of negative affectivity, HADS assesses two emotions, anxiety and depression, which are encompassed within the broader negative affectivity definition. Results are shown in Table 4.10. The only measure to correlate with HADS was Observed shuttle-walk time (rho = -.57, p = .01), indicating that, as distressed increased, time walking decreased. The correlation between observed shuttle-walk distance and HADS was re-run, without P1's data, to give a significant correlation: rho = -.56, p = .01. Similarly, as distress increased, distance walked decreased.

Measures correlated	rho	P	N
HADS with:			
SR – month	.02	.95	19
FLP	.25	.28	20
SR shuttle-walk distance	-.22	.37	19
SR shuttle-walk time	.31	.18	20
Observed shuttle-walk distance	-.37	.11	20
Observed shuttle-walk time	-.57	.01	19
SR – 24 hours	.31	.21	18
SR minutes active	-.28	.39	12
Step Number	-.31	.33	12

Table 4.10: Correlations of HADS with activity measures.

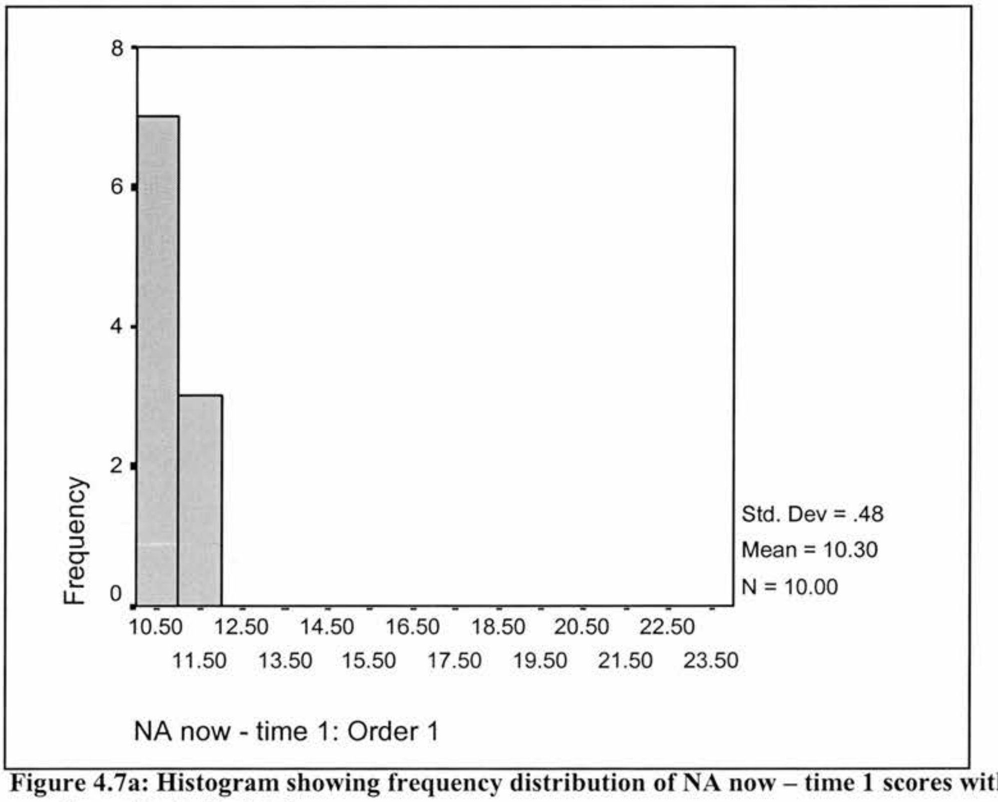


Figure 4.7a: Histogram showing frequency distribution of NA now – time 1 scores with questionnaire in Order 1.

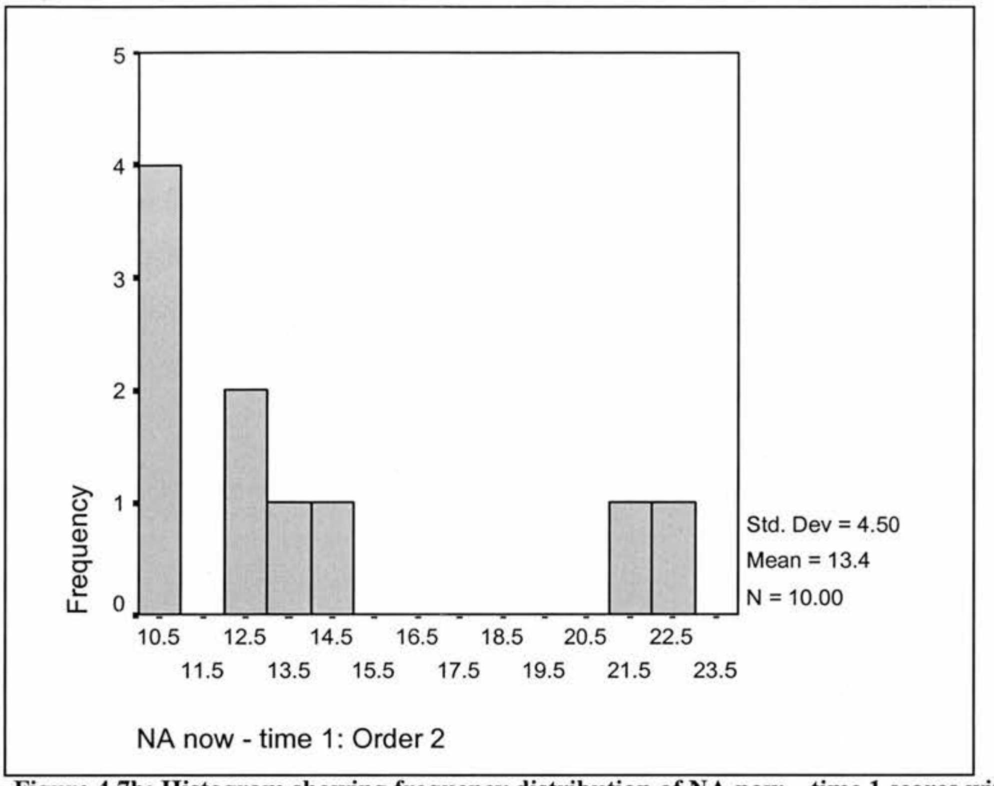


Figure 4.7b: Histogram showing frequency distribution of NA now – time 1 scores with questionnaire in Order 2.

To summarise the negative affectivity findings: few activity measures of any type significantly correlated with negative affectivity measures. These significant correlations were not only found with self-report measures; one was found with an observed measure of activity. Hence the negative affectivity bias hypothesis is not be supported by these data.

Correlations between Social Desirability and measures of activity are shown in Table 4.8. None approached significance and not all are in the hypothesised direction. There is not, therefore, support for the social desirability bias hypothesis in these data.

Partial Correlations.

The effects of Negative Affectivity and Social Desirability on agreement when Bias, the measure of under-reporting, was partialled out were to be investigated. The aim was to test whether any significant relationship with Agreement would persist with the effects of Bias partialled out. Prior to performing the partial correlations, Pearson's correlations were carried out to determine whether parametric correlations would be similar to those found using the non-parametric Spearman's rho. Correlations did not differ at the $p < .10$ level.

		Kappa	Bias	N
NA – 24 hours:	**r**	-.08 (.80)	-.11 (.73)	12
	rho	-.09 (.79)	-.04 (.91)	
Social Desirability:	**r**	.11 (.73)	.22 (.50)	12
	rho	-.08(.81)	.17 (.59)	

Table 4.11: Correlations of NA – 24 hours and Social Desirability with Kappa and Bias using Pearson's r and Spearman's rho. Significance levels are given in brackets.

However, Social Desirability and Negative Affectivity were found not to correlate with either Kappa or Bias, using either Pearson's r or Spearman's rho (Table 4.11). Partial correlation analyses were therefore not reported.

Question 2: b) Are measures affected by a bias of gender?

Activity scores were available with both self-report and non-self-report data for the shuttle walking test and 24-hour activity. Mean scores for comparable measures, according to gender, are found in Tables 4.12a – 4.12d. Figures 4.8a – 4.8d represent the data graphically. The gender bias hypothesis would predict that female participants are more likely to under-report activity than males and that males are more likely to over report activity than females. However, Figure 4.8a shows that both female and male participants under-reported distance walked in the shuttle-walking task, with the difference between self-reported and observed distance appearing to be slightly greater for male participants – the opposite effect to that predicted. Figure 4.8b shows that both male and female participants over-reported the time for which they were walking in the test, with the difference between self-reported and observed mean score being greater for female participants; again, this is the opposite effect to that expected. These gender differences were not significant, however (see Table 4.13).

For the 24-hour activity measures, as the self-reported and monitor scores can not be directly compared as under/over-estimates, the direction of score differences between male and female scores will be discussed. Figures 4.8c and 4.8d show that there is very little difference between male and female scores on all 24-hour measures. Any small differences are in the opposite directions to those predicted by the gender bias hypothesis: whilst the mean number of steps is slightly higher for male than female participants, the self-reported 24-hour Activity score is slightly lower for male than female participants and no gender difference is seen in self-reported Minutes Active. Also, if women were under-reporting activity more than men, the bias score for women would be expected to be lower than for men. Mean bias scores were .57 (SD = .28) for women and .28 (SD = .16) for men; Mann-Whitney U found female and male bias scores to not be significantly different (U = 7.5; p = .11). However, for comparisons including monitor results sample sizes are particularly small, consisting of 5 female and 7 male participants.

It should be noted that female and male scores were not significantly different on any of these measures (see Table 4.13); 'differences' were discussed above in order to describe the data and to emphasise that the lack of support for the gender bias hypothesis is not merely due to there being but a small effect.

Thus it would appear that no support for the gender bias hypothesis was found in these data.

	Mean observed distance	**SD**	**Mean SR distance**	**SD**	**N**
Female	97.45	93.37	60.63	71.51	11
Male	128.25	80.92	75.26	52.35	8
Total	110.42	87.37	66.79	62.94	19

Table 4.12a: Mean and standard deviations of shuttle-walk distance scores according to gender.

	Mean observed time	**SD**	**Mean SR time**	**SD**	**N**
Female	120.29	76.54	285.00	150.50	10
Male	150.48	81.41	298.33	187.72	9
Total	134.59	78.20	291.32	164.42	19

Table 4.12b: Mean and standard deviations of shuttle-walk time scores according to gender.

	Mean Step Number	**SD**	**SR – 24 hours**	**SD**	**N**
Female	7755.44	3017.07	999.78	272.48	5
Male	8099.82	5506.38	905.14	249.32	7
Total	7956.33	4458.68	944.58	251.55	12

Table 4.12c: Mean and standard deviations of monitor-recorded step number and self-reported 24-hour activity according to gender.

	Mean Step Number	SD	SR – Minutes Active	SD	N
Female	7755.44	3017.07	289.62	188.16	5
Male	8099.82	5506.38	182.73	140.01	7
Total	7956.33	4458.68	227.27	163.09	12

Table 4.12d: Mean and standard deviations of monitor-recorded step number and self-reported Minutes Active according to gender.

	Shuttle-walk distance (observed)	Shuttle-walk time (observed)	Shuttle-walk distance (SR)	Shuttle-walk time (SR)	24-hour activity (SR)	Minutes Active (SR)	Step Number
Mann-Whitney U	33.50	36.50	33.00	40.50	13.00	11.50	17.00
P (2-tailed)	.40	.50	.40	.72	.53	.34	1.00

Table 4.13: Results of Mann-Whitney U testing for gender differences in activity measures.

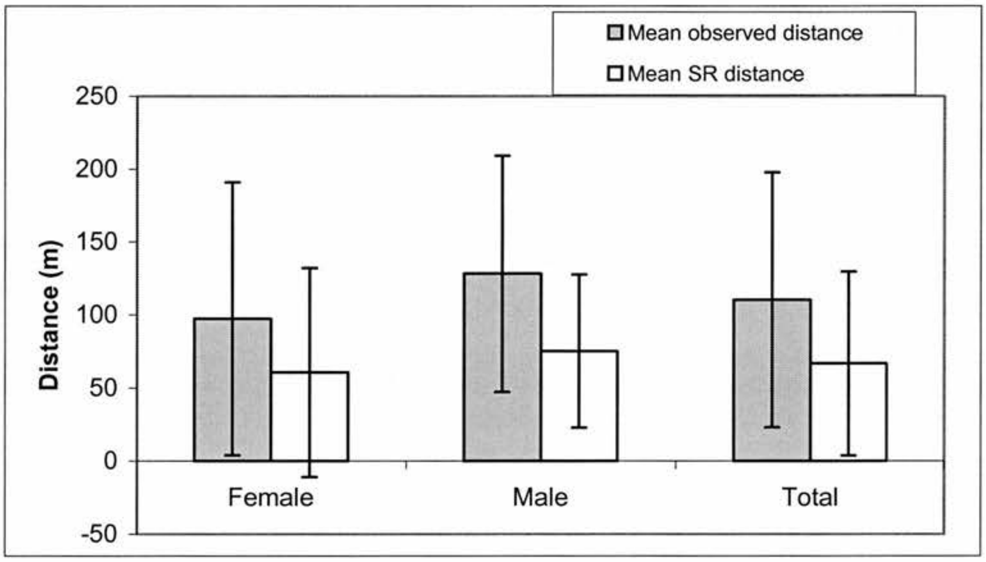


Figure 4.8a: Mean (± standard deviation) shuttle-walk observed and self-reported distance scores according to gender.

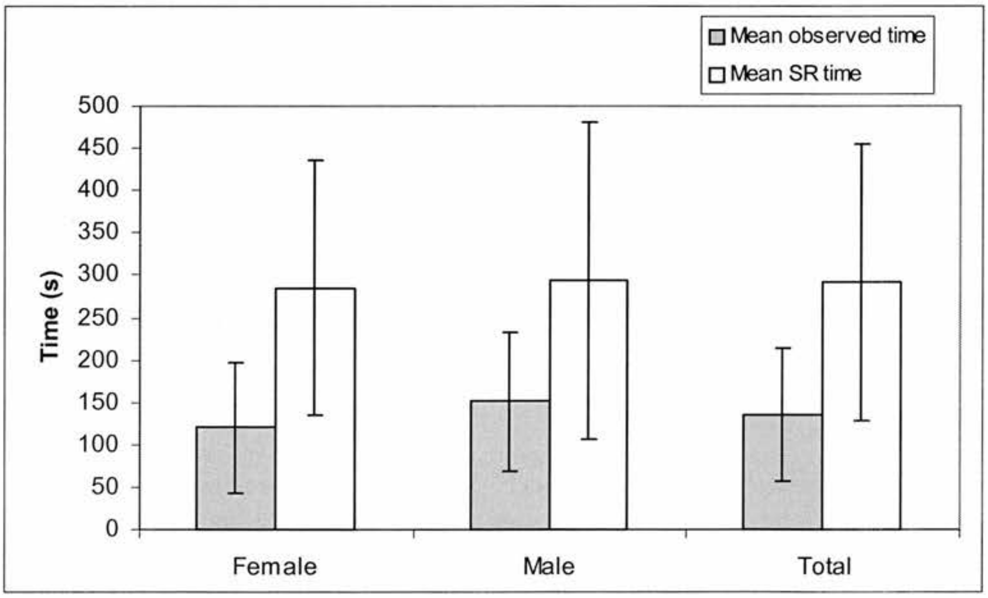


Figure 4.8b: Mean (± standard deviation) shuttle-walk observed and self-reported time scores according to gender.

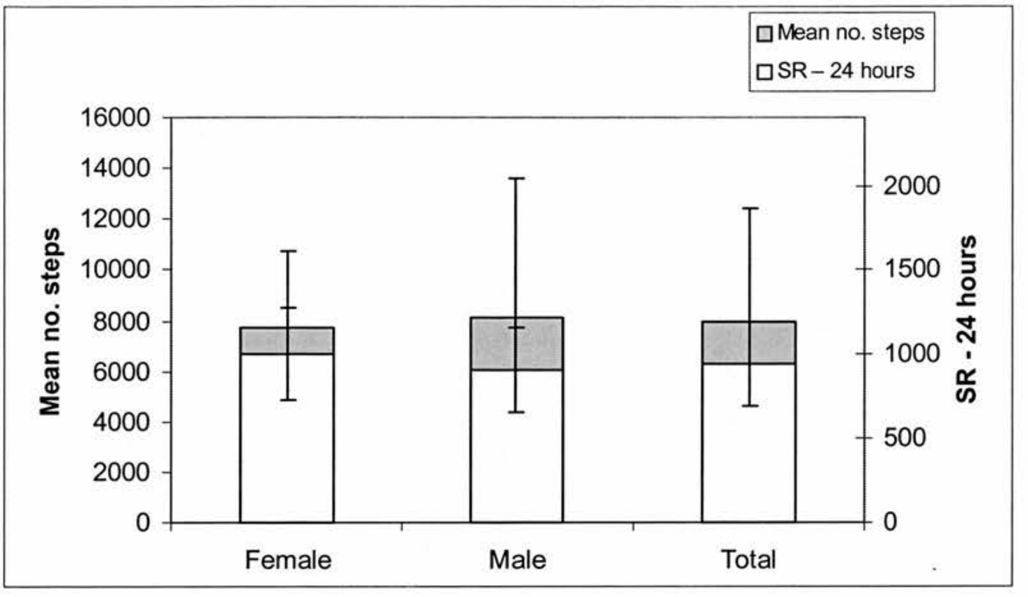


Figure 4.8c: Mean (± standard deviation) monitor-recorded Step Number and Self-reported 24-hour Activity according to gender.

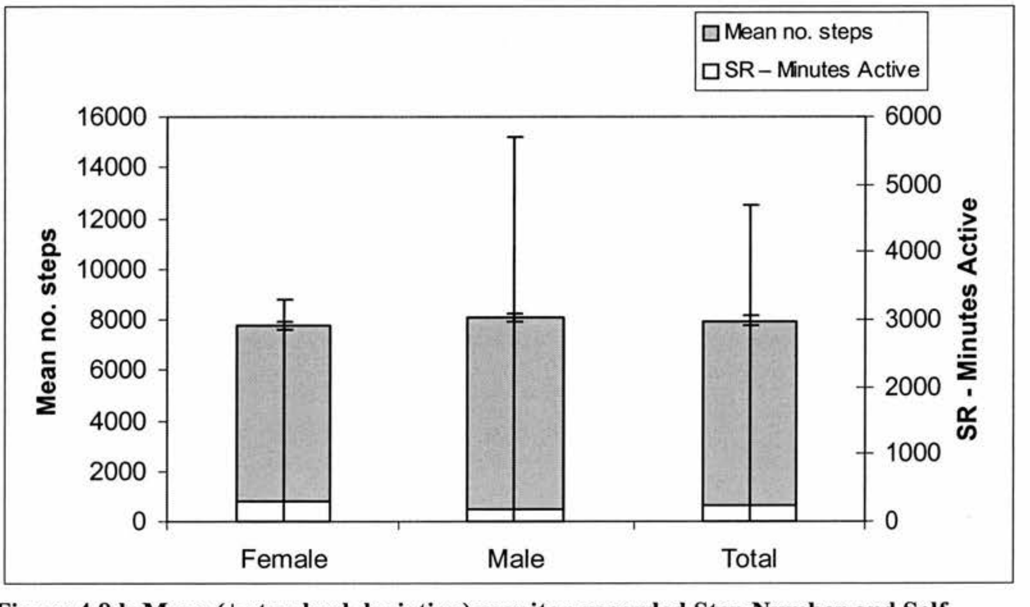


Figure 4.8d: Mean (± standard deviation) monitor-recorded Step Number and Self-reported Minutes Activity according to gender.

Question 3: Do the abilities of emotion and perceived control to predict activity depend on the measurement method used?

Activity Measures	Anxiety	Depression	GSE	PC: walking task	PC: 24-hr activity
Self-Report					
- SR- month	-.08 (.74)	.29 (.23)	.04 (.87)	---------------	---------------
- FLP	.16 (.50)	.03 (.91)	-.54 (.02)	---------------	---------------
- SR shuttle-walk distance	-.18 (.46)	-.09 (.72)	28 (.25)	.30 (.22)	---------------
- SR shuttle-walk time	.23 (.34)	.14 (.57)	-.46 (.04)	-.08 (.73)	---------------
- SR – 24 hours	.10 (.68)	.23 (.37)	-.42 (.09)	---------------	-.58 (.01)
- SR Minutes Active	-.59 (.05)	.31 (.69)	-.30 (.35)	---------------	.17 (.60)
Observed Performance					
- Observed shuttle-walk distance	-.28 (.23)	-.17 (.48)	.44 (.05)	.64 (.00)	-------------
- Observed shuttle-walk time	-.40 (.09)	-.40 (.09)	.53 (.02)	.63 (.01)	-------------
Continuous Monitoring					
- Step Number	-.06 (.84)	-.51 (.09)	-.12 (.72)	-------------	.03 (.93)

Table 4.14: Correlations (Spearman's rho) of Anxiety, Depression, GSE and specific perceived control measures with activity measures. Significance levels are shown in brackets.

Correlations of emotion and control cognitions variables with activity measures are shown in Table 4.14; correlations at or nearing significance (p < .1) are highlighted.

It can be seen that the pattern of significant findings appears to differ according to measure type and predictor variable. Considering self-report measures, Anxiety predicts only one measure (SR Minutes Active, rho = -.59, higher anxiety being associated with lower activity), and Depression predicts activity on no self-report measure. GSE predicts activity at or near significance on three self-report measures: FLP (rho = -.54, p = .02), SR shuttle-walk time (rho = .46, p = .04) and SR 24-hour activity (rho = -.42, p = .09). However, whilst the association with the FLP is in the predicted direction (higher GSE associated with lower disability), higher GSE predicts lower SR shuttle-walk time and 24-hour activity. This finding is supported by higher Perceived Control (PC) over 24-hour activity predicting less SR 24-hour activity (rho = -.58, p = .01). This apparently anomalous finding will be returned to below.

Using Observed Performance measures as the outcome variable, significant or near significant findings are seen for all psychological predictors, in the expected directions: higher Anxiety predicted less time walking (rho = -.40, p = .09), as did Depression (rho = -.40, p = .09). On excluding P1's data, the correlation of Anxiety with observed shuttle walking distance also neared significance (rho = -.40, p = .09) and the correlation of Depression with observed shuttle walking distance became stronger but still lacked significance (rho = -.37, p = .12). Higher GSE and PC over walking task predicted longer distances walked (rho = .44, p = .05 and rho = .64, p = .00) and more time walking (rho = .53, p = .02 and rho = .63, p = .01 respectively).

Only one psychological variable, Depression, neared significance in predicting Continuous Monitored Activity (rho = -.51, p = .09). This relationship is in the predicted direction, with lower depression being associated with higher activity levels.

It would, therefore, appear that Observed Performance measures of activity were most consistently predicted by measures of the psychological variables anxiety, depression and control cognitions.

As reported above, GSE predicted self-reported Shuttle Walk Test time and self-reported 24-hour activity and PC over 24-hour activity predicted self-reported 24-hour activity but in the unexpected direction: higher perceived control predicted lower activity. Scatter-plots of both self-report and observed performance measures of shuttle-walk time and GSE are shown in Figures 4.9a and 4.9b. Participants reported a much greater range of time spent walking that that observed; as reported earlier, the self-reported time spent walking was significantly higher than that observed. Given that the correlation between self-reported distance walked and GSE was, if low and non-significant, in the expected direction, it could be that the correlation between self-reported shuttle-walking time and GSE is confounded by participants' difficulty in estimating performance on this measure. If higher GSE is associated with more accurate reporting, lower GSE would be associated with increased over-reporting of activity and hence higher GSE would appear to predict lower self-reported activity. This hypothesis was tested by correlating GSE with Error in reporting Shuttle Walk Test time (SR shuttle walk time – observed shuttle walk time). A significant correlation was found: rho = -.59, p = .01: the lower the GSE score the greater the over-reporting of shuttle walk time. Hence it would appear that the relationship between GSE and SR shuttle walk time could be confounded by report accuracy. This was explored further by correlating GSE and SR shuttle walk time, partialling out Error. Table 4.15 shows both Spearman's rho and Pearson's r values for the correlations of GSE with SR shuttle walk time and Error. It can be seen that the correlation between GSE and Error is very similar with both parametric and non-parametric correlations; the correlation of GSE and SR shuttle walk time nears significance (r = -.38, p = .10) and is in the same direction with Pearson's parametric correlation as with Spearman's non-parametric method. On partialling out Error, the correlation between GSE and SR shuttle-walk time neared significance (r = .43, p = .07). The association is now in the predicted direction, with higher levels of self-efficacy being associated with longer time reported walking. This supports the hypothesis that accuracy confounded the bivariate association between GSE and self-reported shuttle-walk time.

Measures	**rho**	**p**	**r**	**p**
GSE – SR shuttle walk time	-.46	.04	-.38	.10
GSE - Error	-.59	.01	-.57	.01

Table 4.15: Correlations (Spearman's rho and Pearson's r) of GSE with SR shuttle walk time and Error.

It is possible that the same hypothesis explains the similar relationship between PC over 24-hour activity and SR 24-hour activity. Unfortunately this can not be tested with the same method as there is no non-self-report measure of 24-hour activity on a directly comparable measure.

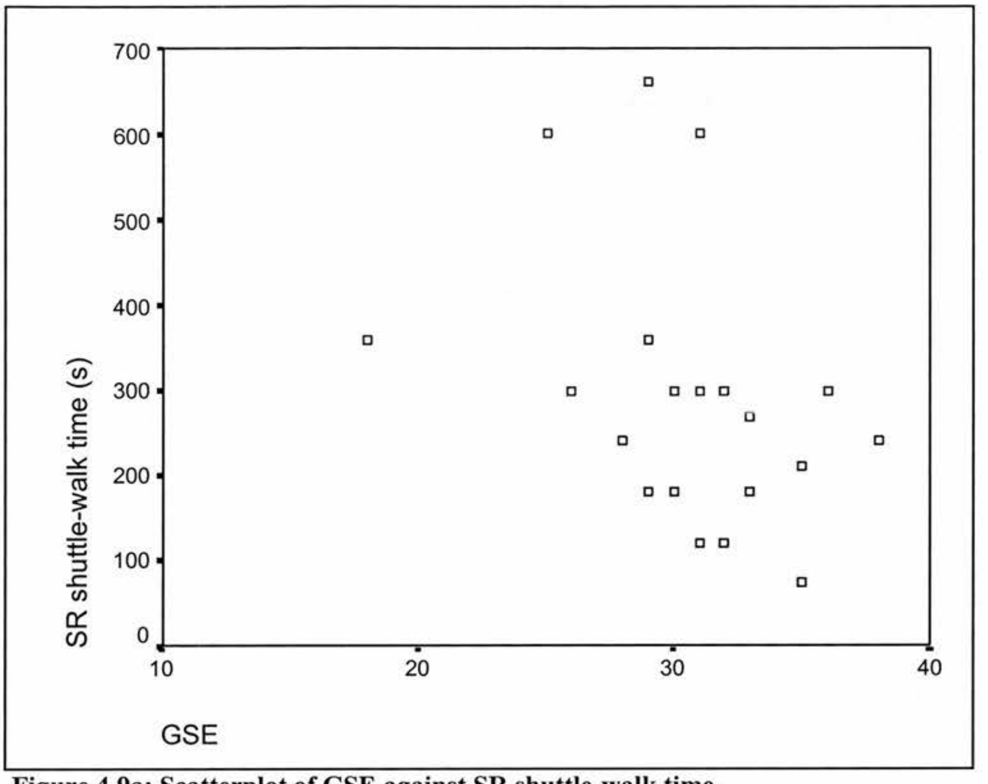


Figure 4.9a: Scatterplot of GSE against SR shuttle-walk time.

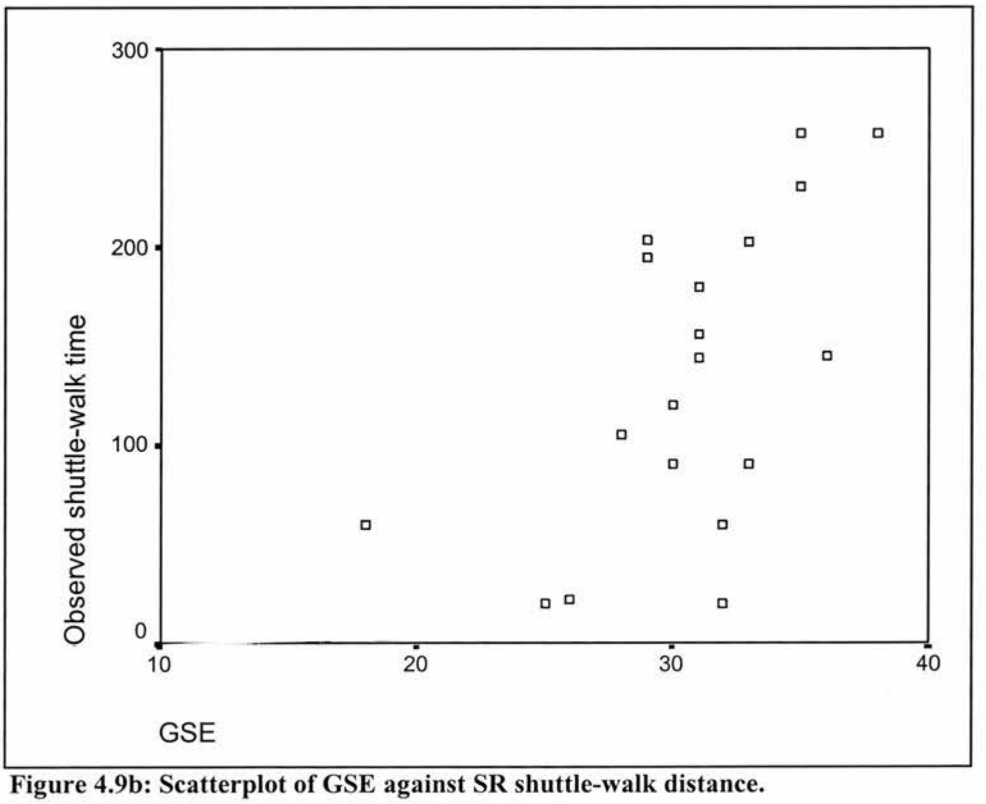


Figure 4.9b: Scatterplot of GSE against SR shuttle-walk distance.

Partial Correlations.

The effects of Anxiety, Depression, GSE and Perceived Control – 24 hours on Kappa (Agreement) when Bias, the measure of under-reporting, was partialled out were to be investigated. The aim was to test whether any significant relationship with Agreement would persist when any effects of bias had been partialled out. Prior to performing the partial correlations, Pearson's correlations were carried out to determine whether parametric correlations would be similar to those found using the non-parametric Spearman's rho.

		Agreement (Kappa)	Bias	N
Anxiety	**r**	-.20 (.54)	-.24 (.46)	12
	rho	-.10 (.76)	-.23 (.47)	
Depression	**r**	.13 (.69)	.62 (.03)	12
	rho	.06 (.85)	.45 (.13)	
GSE	**r**	-.12 (.70)	-.11 (.74)	12
	rho	-.07 (.84)	-.20 (.53)	
Perceived Control – 24 hour activity	**r**	-.04 (.92)	-.05 (.88)	12
	rho	.06 (.86)	-.01 (.97)	

Table 4.16: Correlations Anxiety, Depression, GSE and Perceived Control – 24 hour activity with Kappa and Bias using Pearson's r and Spearman's rho. Significance levels are given in brackets.

Table 4.16 shows that none of the psychological predictors were associated with Agreement with either Pearson's r or Spearman's rho and so partial correlations were not performed. Depression was found to correlate with Bias using Pearson's correlation coefficient although the Spearman's correlation does not quite reach trend level ($r = .62$, $p = .03$; $rho = .45$, $p = .13$). This positive association suggests that higher Depression is associated with less under-reporting, ie more accurate scores.

Discussion

Research Question 1: To what extent do self-report, observed performance and continuously monitored assessments of activity correspond?

Significant correlations between self-report and observed performance activity measures were found between activity limitation (FLP) and shuttle-walking test scores and between self-reported and observed shuttle-walk distance. One correlation between self-report and monitor measures neared significance: self-reported activity over the past month correlated with Step Number at rho = .86, p = .06. Observed performance and monitor ratings were not significantly related. These results contrast with the pilot study findings where the only association nearing significance was between monitor and self-reported 24-hour activity. Both these studies have limited power, however, because of the small sample sizes.

Nevertheless, it would appear that, the closer the domains of activity assessed, the more likely measures are to be associated. Even though the FLP asks participants to report what activity they *do*, the broad categories e.g. 'I do not walk at all', combined with the query as to whether that is due to the state of health, are more likely to reflect what people *are able* to do than the YPAS which asks people to recall what activity they have been performing. Taking the self-report and observed performance ratings, the self-report FLP, a measure likely to be related to what people *are able* to do was found to correlate with a task assessing how far and fast people *are able* to walk. Modified YPAS self-reported activity over the past month, a measure of what people report *actually doing* did not correlate with this task assessing what people *are able* to do. Monitor records, assessing what people *actually do* over 24 hours did not correlate with self-report or observed measures of what participants *were able* to do, but near significance with one measure of what people *actually do*: self-reported activity over the past month. These findings are paralleled by findings within the class of self-report measures: FLP correlated with self-reports of what participants *were able* to do on the walking task but not with other self-report measures. The only other correlation nearing significance within the self-report measure group was that between SR 24-hour activity and the 24-hour Minutes Active rating: measures rating activity over the same time period, in the same domain (self-report) were associated whilst measures assessing different aspects or time-frames of ability were not.

Monitor Step Number did not correlate with self-reports of activity over the same time period, however; this will be discussed further below.

Even where significant correlations were found, much variance remained unaccounted for, for example correlations between self-report FLP and shuttle-walk performance ranged from -.60 to -.64. It would seem that these measures are assessing different aspects of activity.

It seems that participants recalled distance and time very differently. Not only were self-reported and observed shuttle-walk distance scores found to correlate whilst shuttle-walk time scores did not but participants seemed to systematically underestimate distance walked whilst over-estimating time spent walking, sitting and standing. This replicates the findings of the Pilot Study but contradicts Klesges et al.'s (1985) results which found student participants to under-estimate sitting and moderate activity whilst over-estimating time in more active activity levels. It could be that the different activity instructions could affect recall: Klesges et al's participants were instructed to engage in any activity they wished whereas participants in the Pilot Study and the present study were given precise instructions as to what activity to report. Hence, whilst Klesges et al's participants may have perceived social pressure to have performed a high level of activity, spending less time in low-intensity activities, these chapters' participants' behaviour was more restricted by the researcher – no opportunity was given for activity more intensive than walking to be performed. Also, whilst a young adult sample might wish to have been highly active, people who volunteer to take part as 'adults with walking difficulties' may wish to demonstrate walking difficulties rather walking ability. Koivula (1996) found that participants under-estimated, rather than over-estimated short time intervals (15-75s). However, Koivula's (1996) participants knew that the task was to estimate time, which may have lead to finding a different effect. In a review of time estimation literature, Glicksohn (2001) reports perceived duration to be related to absorption such that the more absorbed people are, the shorter a time interval seems. It could be that Koivula's (1996) participants were absorbed in the task of time estimation, making the time intervals seem shorter than they actually were, whereas the participants in the present study were merely waiting to move on to the next stage of the study and so the time intervals seemed longer.

Significant relationships were not found between monitor Step Number and self-report activity assessed over the same time period. Within-participant correlations showed that the relationships between monitor activity records and self-reported activity varied greatly between participants. Either the monitors recorded activity differently for different participants, or different participants recalled activity in different ways. The latter seems plausible not only in terms of individual differences in 24-hour memory, but also because the activity measure 'Minutes Active' seemed to be interpreted differently by different participants. This limitation will be discussed in detail below. Most participants reported activity in significantly fewer half-hour blocks than the monitor recorded.

Self-reported activity may have been difficult to recall for this participant population because of the low-level type of activity they typically performed. Washburn, Jette and Janney (1990) found that reporting error was smallest for the most strenuous activity category but substantial for the least strenuous category. Hence the activity most likely to be carried out by their sample of older people, and perhaps also this sample of people with walking difficulties, was the type least accurately assessed.

Research Question 2: Are measures affected by biases of negative affectivity, social desirability and gender?

No support was found for reporting biases of negative affectivity, social desirability or gender.

Negative Affectivity was found to correlate with observed performance measures at least as strongly as with self-report measures, suggesting that the association between negative affectivity and activity limitations is not simply an artifact of self-report measures. However, as correlations were not seen with Step Number as assessed by continuous monitoring, it is possible that the observed performance measure is also biased by negative affectivity in as much as it may reflect what people feel *motivated* to do rather than what they actually *can* do. This issue will be returned to in the following chapter, where self-report and observed performance measures are compared with proxy-reported activity limitations. The Negative Affectivity measures were limited by floor effects in this sample, reducing the chances of finding correlations. However, when analyses were re-calculated using the less skewed Hospital Anxiety and Depression scale as a proxy measure of negative affectivity, an effect was still not found: HADS was only associated with observed performance measures of activity. This contradicts Watson and Pennebaker's (1989) suggestion that self-report measures reflect negative affectivity whereas more objective measures of health status are more likely to reflect actual health outcomes.

Social desirability was not found to be associated with any measure of activity limitations.

Not only was a gender effect not found, but small differences between male and female participants were in the opposite direction to that hypothesised: these results would suggest that men are more likely to under-report activity than women relative to observed performance or continuous monitoring assessments. However, sample sizes for gender comparisons were small, especially for comparisons involving activity monitors (5 female, 7 male).

Research Question 3: Do the abilities of control cognitions and gender to predict activity depend on the measurement method used?

Anxiety and Depression significantly predicted activity only on some measure types. Only one significant association was seen with a self report measure: Anxiety was associated with self-reported Minutes Active. Both Anxiety and

Depression neared significance in predicting an observed performance measure (Observed shuttle-walk time: rho = -.40, p = .09 for both anxiety and depression). The continuous monitoring measure (Step Number) was only predicted by Depression at trend level (rho = -.51, p = .09).

Some evidence was found to suggest that higher Depression scores were associated with less under-reporting (ie more accurate reporting) of activity. Evidence has been found to support the idea that depressives may be more realistic (see Williams, (1992)). It is surprising that the direction of error for less depressed participants was under-reporting rather than over-reporting however. Further investigation with a larger sample would be required to confirm this effect.

Some self-report activity measures were predicted by Generalised Self-Efficacy (GSE) and Perceived Control (PC) over 24-hour activity. However, apart from the association between FLP and GSE, these associations were in the opposite direction to what would have been expected: higher perceived control scores were related to lower activity reports. Further investigations suggested that low GSE was associated with more over-reporting error on shuttle-walk time and that this association confounded the association between control cognitions and self-reported activity. The association between GSE and SR shuttle-walk time reversed, nearing significance in the expected direction, when Error was partialled out. Koivula (1996) also found control perceptions to be associated with the report of time intervals. Participants with an internal locus of control reported time intervals more accurately than those with an external locus of control when a mental arithmetic task was being performed at the same time as time estimation. It was suggested that individuals with an internal locus of control search more actively for information in the environment, acquire more information and use it more adequately than those with an external locus of control. The 'attentional approach' to time estimation suggests that increased attentional load diverts effort from the task of time estimation, lessening the accuracy of time estimation (Glicksohn, 2001; Koivula, 1996). Therefore, when cognitive load exceeds processing capabilities, internals would be expected to better use information in the environment and be more accurate in time estimation than externals; this hypothesis was supported (Koivula, 1996). Four key differences exist between Koivula's study and this, however. First, the locus of control measures used by Koivula and the perceived control assessments used here may not give comparative indications of control perceptions. Second, Koivula's externals were underestimating time whereas the less accurate participants with lower GSE scores were over-reporting time. Koivula's participants knew that the purpose of the task was to test time estimation. Finally, Koivula's participants were completing a mental arithmetic task designed to provide cognitive load. In contrast, the Shuttle Walk test was designed to test physical activity but, given participants' unfamiliarity with the set-up was likely to have taken some concentration. Hence, whilst Koivula's (1996) study provides some support and theoretical framework for the association between control cognitions and error in reporting time, it can not be assumed that the processes underlying Koivula's findings are the same as in this study.

In contrast, observed performance measures were consistently predicted by perceived control variables in the expected direction: higher control scores predicted better performance (from rho = .44 to rho = .64). This is consistent with the findings of Lackner et al. (1996), Lackner et al. (1999) and Partridge and Johnston (1989). Neither GSE nor PC over 24-hour activity predicted continuous monitoring scores.

It would appear, therefore, that the ability of emotion and perceived control to predict activity did depend on the measurement method used. The observed performance outcome measures were the most consistently predicted followed by self-report measures.

Limitations.

Participants

The study was limited by size of the sample and by the number of participants for whom monitor data was obtained. Data on some measures deviated from the normal distribution. Both these factors restricted the range of analyses that could be performed on the data. The extent to which results from this highly-selected and highly educated sample can be generalised to a wider population is limited. However, despite these limitations, results were found which, if not sufficient to be conclusive of themselves, certainly suggest further investigation with larger samples would be merited.

Measures

In developing the Shuttle Walking Test, Singh, Morgan, Scott et al. (1992) found that COPD patients walked significantly further on 2nd and 3rd trials than on the 1st attempt. A practice walk was not given here as time was not available to do so without risking fatigue. During the laboratory sessions it was observed that many participants would have benefited from a trial walk to better learn to pace themselves and in gaining a clearer understanding of what the test involved. This was a drawback in terms of accurately comparing what participants might be really able to do with what they did in the self/monitor-reported 'real world', but should not be of importance in comparing self-reported walking test performance with observed performance. It is not clear what effect this might have had on the relationships between emotion and perceived control and performance.

As described fully in the Calibration Studies, the performance of the Numact activity monitor was disappointing. The poor results of the Calibration Studies meant that between-participants analyses were restricted to using the 'Step Number' monitor records, rather than being able to use 'Steps x Amplitude', a measure that would takes into account how energetically people walk as well as recording steps. However, the equipment was found to be acceptable to participants, despite being potentially cumbersome during activities such as dressing. Future studies with more reliable monitoring devices would, therefore, be a viable option.

The 'Minutes Active' over 24 hours measure seemed to be answered in different ways by different participants. Whilst some participants reported activity at a level of great detail, answering with a very anecdotal style such as 'I sat down for 5 minutes to have a cup of tea before standing and washing up for 20 minutes. Then I was moving around until 4 o'clock', others took their average activity for a half-hour period. Hence the measure will have been much more sensitive for some participants than for others. This would have the effect of increasing random variation in 'Minutes Active' scores and masking potential correlations of this measure. It could also have a great effect in those analyses using the bivariate scoring system where activity was either reported or not reported in a particular block. Whilst anecdotal reporters would have a high score on such a measure, with every half-hour block with any activity being recorded, the second group of participants would have systematically excluded half-hour periods where they were active for just a few minutes.

One problem considered with respect to comparing self-reported 24-hour activity with monitor data is whether or not to include night-time hours in analyses. During this time report of activity is likely to be relatively accurate: people can generally answer with some confidence when they went to bed and when they arose. As participants sometimes got up in the night, these data were not excluded. It is also questionable whether the ability to estimate time asleep is less important than ability to estimate time in other activities. Indeed, McDermott, Liu, O'Brien, Guralnik, Criqui, Martin and Greenland, (2000), using the Stanford 7-day Physical Activity Recall measure, calculated time spent in 'light' activity by subtracting time spent sleeping and in moderate, hard and very hard activity from 24 hours.

People need to have an understanding of the behaviour they are to perform in order to know how much effort to put into a task and, if people do not know what it is they are aiming for, there is little foundation from which to turn self-efficacy into appropriate effort (Bandura, 1986). In the context of this study, this would suggest that, unless participants had a good idea of what was expected of them in the Shuttle Walking Test, a relationship between control cognitions and test performance is unlikely. The Shuttle Walking Test is an unusual activity that participants have not tried before and, although the test was fully described to them and they indicated that they understood what they would be asked to do, it seems likely that they may not have had a clear idea of what it would be like to actually do the test. This could have reduced any relationship between control cognitions and test performance.

The specific perceived control measures (PC over walking task and PC over 24-hour activity) showed relatively poor reliability (Cronbach's α = .56 and .57 respectively). Elimination of either item 1 or item 3 (see Box 4.1) lead to an increased alpha (from .56 to .65 and .67 for items 1 and 3 respectively, Perceived Control – Walking Task and from .57 to .71 and .61 for Perceived Control – 24 hour Activity). Item 3 was also found to reduce reliability in the pilot study (Chapter 2). It is a potentially confusing question as it could be interpreted in 2 ways. It is also the one item in the scale where the response categories are reversed. It is unclear why item 1 also reduced reliability here. It does seem to be asking about a locus of control – *who* has control rather than *how much* control is perceived – it could be that, for this population, in this context, locus of control is perceived differently to other aspects of control.

1. Whether I do or do not perform well on the walking task / am or am not very active in the next 24 hours is entirely up to me.
1 = strongly disagree, 7 = strongly agree.

Strongly disagree							*Strongly agree*
	1	2	3	4	5	6	7

3. I would like to perform well on the walking task / be very active in the next 24 hours but I don't really know if I can.
1 = strongly disagree, 7 = strongly agree.

Strongly disagree							*Strongly agree*
	1	2	3	4	5	6	7

Box 4.1: Items 1 and 3 of Perceived Control – Walking Task and Perceived Control – 24-hour Activity

There is also a problem with the PC over 24-hours activity measure in that participants were asked to report on the control they perceived over being 'very active in the next 24 hours'. It could be that interpretations of 'very active' differed for people with different activity levels. For someone who mainly sits, being active could mean performing basic tasks around the house whereas, for someone who is more active, being very active could imply a greater, less attainable, level of activity.

Procedure

Overall, few procedural problems were found. The exception is in running the 'Unaware Observed' walking speed measure which did not work well in this context (see the Experimental Study for an account of it being with more success). The current study's participants' slow walking speed meant that there was more time in which other people could trigger the photo-electric detectors than for a quicker sample. The School of Psychology was a strange environment for most of the participants and they would need to be accompanied by the researcher on leaving down the corridor in which the photo-electric detectors were placed. Even though the researcher would walk behind the participant, on one occasion a participant stopped to chat to the researcher part way down the corridor, and the researcher's presence could also have influenced the walking speed of other participants. As a result of such problems, what could have been a useful measure was not analysed. It is difficult to see how these problems could be overcome in this particular workplace, but, as the technology was very effective and reliable, if another building had a quiet exit corridor down which participants could easily be directed without accompaniment by a researcher, a similar measure could be used more effectively.

Conclusions

Few significant correlations were seen between activity measures; associations found suggested that measures are more likely to concur when assessing more similar forms of activity, for example, what participants *can* do as opposed to what they *actually* do. Biases in the reports of time and distance walked were seen: participants under-reported distance and over-reported time on standardised measures. No support was seen for biases of gender, negative affectivity or social desirability. The ability of emotion and control cognitions to predict activity did appear to depend on the measurement method used, with the outcome measure most consistently predicted being observed performance. A major problem for this study was its size, with only 20 participants it became a second exploratory study rather than having the power to enable more concrete conclusions. These issues are examined in the Stroke Study with a larger patient sample.

CHAPTER 5:
COMPARING SELF-REPORT, PROXY-REPORT AND OBSERVED PERFORMANCE ACTIVITY MEASURES IN STROKE PATIENTS.

Abstract

Background

The Pilot and Walking Difficulties Studies did not find activity measures to be strongly associated with each other and biases of negative affectivity and gender were not found. Whether the predictive effects of emotion and control cognitions are dependent on activity measure type has not been satisfactorily answered as the Pilot Study suggested self-report and the Walking Difficulties Study suggested observed performance measures were most consistently predicted. However, both these studies were limited by small sample sizes. The present study addresses these issues with a large sample of stroke patients. Continuous monitoring of activity is not used; instead, a proxy-report measure is used alongside self-report and observed performance assessments of walking.

Research Questions

1. To what extent do self-report, proxy-report and observed performance assessments of activity correspond?
2. Are measures affected by biases of negative affectivity and gender?
3. Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Methods

101 stroke patients and proxies (their caregivers) were interviewed at 3 time points: within 2 weeks of discharge from hospital, 6 weeks later and 6 months after time 1. Self-report, proxy-report and observed performance measures of ambulatory activity were taken. Also included in the interview were assessments of negative activity, anxiety, depression and control cognitions.

Results and Conclusions

All measures correlated highly significantly with other measures at the same time point, with the highest correlations being between self (patient)- and proxy-reports but even in the best case correlation, much variance was unaccounted for. Different measures seem to access different phenomena. Evidence was not found to support hypotheses of bias caused by negative affectivity or gender. Self-reported activity ('Recovery' - residual when activity score from earlier time point was regressed on activity score at a later time point) seemed to be the measure most consistently predicted by psychological variables.

Introduction

The Pilot and Walking Difficulties Studies investigated to what extent activity measures give corresponding results, focusing on walking activity. Strong correlations between measures were not found overall, although the Walking Difficulties Study suggested that measures were more likely to correspond when assessing closer forms of activity – ie either both assessing what participants *are able* to do or both assessing what participants *actually* do. However, the sample sizes in both studies were small and associations may not have been found due to lack of power to detect effects. The present study investigates associations between self-report, proxy-report and observed performance ratings in a sample of stroke patients; the sample size is much larger than those of the Pilot and Walking Difficulties Studies (n = 101).

The negative affectivity bias hypothesis, which would predict negative affectivity to be more strongly related to self-report measures than other measures of activity, was not supported in the Pilot and Walking Difficulties Studies as higher correlations between self-report measures and activity than between observed performance measures and activity were not found. However, as mentioned in the Introduction (Chapter 1), it is possible that people high in negative affectivity put less effort into an observed performance test than people with lower negative affectivity and so the association with negative affectivity may not reflect the relationship between negative affectivity and 'normal' activity. Neither the Pilot nor Walking Difficulties Study found negative affectivity to be associated with continuously monitored activity. Alongside self-report and observed performance measures, the present study will examine the relationship between proxy-reported activity and negative affectivity. The negative affectivity hypothesis would predict that proxy-reported activity would be related to proxy negative affectivity but not to patient negative affectivity. In contrast, if the association between negative affectivity is real and not an artifact of assessment methodology, proxy-reported activity would be associated with negative affectivity similarly to self- and observer-rated activity.

Neither the Pilot nor the Walking Difficulties Study found gender to bias activity measurements. The possibility that gender may affect activity assessment (with men over-reporting and women under-reporting compared with observed performance measures) is considered again with the larger sample size of the present study.

The Pilot and Walking Difficulties Studies gave apparently contradictory results in assessing the relationship between the psychological variables emotion and perceived control and activity outcomes. The Pilot Study found only self-report measures to be predicted whereas the Walking Difficulties Study found observed performance to be more consistently predicted by the psychological variables. Whether the predictive effects of emotion and control cognitions depend on measure type are examined again in the present study, with the larger sample size giving the analysis more power to detect effects.

This study differs from the Pilot and Walking Difficulties Studies in four respects. The sample size was larger (n = 101) and data was available at 3 very separate time points: full data on two activity measures (self- and proxy-report) was obtained from stroke patients within 2 weeks of discharge from hospital, 6 weeks after this interview and 6 months after discharge, with an observed performance measure being included at times 2 and 3. Instead of being brought into the laboratory, participants were interviewed in their own homes. No activity monitor was used, but a proxy-report measure joins the self-report and observed performance assessments of walking activity.

Research Questions:

1. To what extent do self-report, proxy-report and observed performance assessments of activity correspond?

2. Are measures affected by biases of negative affectivity and gender?

3. Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Method

Design

The data described and analysed here were part of an existing data set[3] and were made available to me to address the present research questions. Study design was therefore determined by the requirements of Johnston, MacWalter, Morrison and Pollard (see footnote; details in Bonetti et al. 2001) and the measures used here are a selection from the wider range of measures utilised in the existing data set.

Stroke patients were assessed on three occasions following discharge from hospital: within 2 weeks of discharge, 6 weeks after interview 1 and 6 months after discharge. Self-report, proxy-report and observer assessed measures of disability, and measures of perceived control and emotion were used. Questions 1 and 2 were investigated cross-sectionally at each time point whilst, for question 3, a longitudinal, prospective design was used.

Participants

224 patients were recruited on discharge from Ninewells Hospital, Dundee following an acute stroke between 21 April 1998 and 8 May 2000. The data of the 101 patients (33 female, 68 male) with complete patient and proxy data on each activity measure at all 3 time points (T1, T2, T3) are included here (see Table 5.1). Mean age: 67.68 years, SD: 10.69; range: 39 – 87 years. Most (81) of the proxies were the patient's spouse (see Table 5.2).

Of the 52 patient/proxy pairs excluded, 50 had complete data for time 1. This baseline data was compared on all measures with the data of the 101 included participant pairs. No significant differences were found (Appendix E.iv).

---

[3] The Stroke Workbook Outcomes Trial was administered by Johnston, MacWalter, Morrison and Pollard and participants were enrolled and data collected for its purposes. I was an interviewer for this study over 5 weeks between November 1998 and February 1999 (entering data during this period) and also conducted interviews between January and March 2000.

	n
Total recruited	224
No proxy participating:	*71*
Excluded:	*52*
Withdrew due to health	*12*
P/C not available	*7*
Died	*5*
Data incomplete	*4*
GP withdrew	*2*
Moved address	*2*
P/C chose to withdraw for other reasons	*20*
Included participants (total with all 3 measures at all 3 time points):	101

Table 5.1: Summary of recruitment of participants.

Relationship	Frequency
Spouse	81
Daughter	8
Son	3
Sister	2
Daughter-in-law	1
Friend	1
Grandchild	1
Niece	1
Partner	1
Sister-in-law	1
Son-in-law	1
Total	**101**

Table 5.2: Relationship of proxy to patient

Measures

Activity limitations

a) Self-report: Modified Functional Limitations Profile (FLP) (Pollard & Johnston, 2001): Ambulation subscale (Appendix A.v)(T1, T2, T3).

This measure is described in full in the Walking Difficulties Study (Chapter 4).

b) Proxy-report: Modified Functional Limitations Profile (FLP) (Pollard & Johnston, 2001): Ambulation subscale (Appendix A.v)(T1, T2, T3)

The proxy was asked to report on the patient's level of activity limitation. The scale was administered and scored as for *a)* above.

c. Observer assessed: The Observer Assessed Disability Scale (OAD) (Partridge, Johnston, & Edwards, 1987): standing and walking items (items 10-15, T2, T3). Appendix A.viii.

The interviewer asked patients to perform standing and walking actions, for example, 'I would like you now to take 2 steps forward' and watched them perform the requested movement. Scoring was dichotomous: 0 = no, 1 = yes, can perform action. In testing the inter-rater reliability of the scoring method, ten pairs of physiotherapists assessed 10 patients; 100% agreement was found (Partridge et al., 1987). Item 15 was actually a self-report item: 'You have shown me your walking in the house, rather than go outside can you just tell me if you have been outside in the garden or street and have walked alone (no other persons help) over the last 2 days'. Initial analyses were performed both with (scale OAD6) and without (scale OAD5) this item. Results were very similar (see results section). Item 15 was therefore excluded in analyses so that the scale would only include observed items.

Bias Variables

Negative Affectivity: Hospital Anxiety and Depression Scale (HADS) (Zigmond & Snaith, 1983) (T1, T2, T3). (Appendix C.i)

HADS is described in the Pilot Study (Chapter 2). It was completed by both patients and proxies (proxies referred to themselves, not the patients, for this measure).

The scale contains 14 items, 7 to assess depression and 7 for anxiety. The Anxiety and Depression scores can either be summed separately, to create anxiety and depression scores, or all scores can be summed together, giving an overall measure of distress (Johnston et al., 2000). The overall distress score was used as a measure of negative affectivity as it was the measurement most closely reflecting Watson and Pennebaker's (1989) definition of negative affectivity (see Introduction) included in the stroke workbook project, with anxiety and depression being components of the broader negative affectivity construct. Tessler and Mechanic (1978) used the HADS was used in a similar way to measure negative affectivity.

Psychological Predictors: Emotion and Control Cognitions

Emotions: Anxiety and Depression: HADS (see under *'negative affectivity'*, above).

Anxiety and depression items were summed separately giving two emotion indices: 'Anxiety' and 'Depression'.

Perceived Control

a. Perceived control over recovery: Recovery Locus of Control Scale (RLOC) (Partridge & Johnston, 1989) (T1, T2) (Appendix C.v).

The RLOC consists of 9 items, derived from statements elicited from 34 stroke and 24 wrist fracture patients, about their perceptions of control. Five of the items describe internal perceptions of control (e.g. 'It's what I do to help myself that's really going to make all the difference') and 4 describe external perceptions (e.g. 'I have little or no control over my progress from now on'). Responses are given on 5-point scales (strongly agree, agree, uncertain, disagree, strongly disagree) and scored in the direction of internality. Partridge and Johnston (1989) found correlations of internal items with the internal mean to range from .49 to .75; correlations of external items with the external mean ranged from .49 to .80 for 20 stroke and 20 wrist patients. Internal items means correlated -.79 with external items' means (Partridge & Johnston, 1989). Johnston, Morrison, et al. (1999) found the measure to have internal consistencies of .64, .77 and .53 at 3 time points in a sample of 71 stroke patients.

b. Perceived control over walking: Perceived Control Index (PCI) (Bonetti, 2000) (T1, T2) (Appendix C.vi)

This scale combines two Perceived Behavioural Control items with one Self-Efficacy and one Locus of Control item. All items are derived from responses to the modified FLP. The descriptions below are based on a participant who agreed with the FLP statement 'I do not use stairs at all'.

i) Perceived Behavioural Control (PBC):

Two items ask about perceived control over the behaviour selected as the response to the ambulatory subscale of the modified FLP (see earlier).

b.1) 'How much control do you feel you have over whether you *will use stairs* over the next month?' and

b.2) 'How difficult will it be for you to *use stairs* over the next month?'

Responses are given on 5-point scales ranging from 'no control' to 'complete control' and from 'not at all difficult' to 'extremely difficult'. Item b.2) was reversed before the 2 items were summed; thus, higher scores reflect higher perceived control.

ii. Self-Efficacy (SE).

b. 3) 'How confident are you that you will *use stairs* over the next month?'

SE responses were rated on a 5-point scale ranging from 'not at all confident' to 'extremely confident'. Higher scores reflect higher self-efficacy.

iii. Locus of Control (LOC).

b.4) 'Do you think that it is entirely up to you whether you will *use stairs* over the next month'

LOC responses were rated on a 5-point scale ranging from 'strongly agree' to 'strongly disagree' were recorded, with higher scores reflecting higher internality.

Correlations between PBC, SE and LOC are shown in Tables 5.3a and 5.3b below. All correlations are moderately strong, ranging from r = .45 to r = .72, suggesting that the measures show concurrent validity but are also discriminative in assessing separate variance.

	T1 SE	T1 LOC
T1 PBC	.64**	.47**
T1 LOC	.55**	

Table 5.3a: Correlations between perceived control measures at time 1.
**** = p<.01**

	T2 SE	T2 LOC
T2 PBC	.72**	.47**
T2 LOC	.45**	

Table 5.3b: Correlations between perceived control measures at time 2.
**** = p<.01**

The 3 scores (PBC, SE and LOC) were converted to z-scores and summed to create PCI scores.

	Patient	Proxy
Time 1	FLP HADS RLOC PCI	FLP HADS (re. self, not patient)
Time 2	FLP OAD HADS RLOC PCI	FLP HADS (re. self, not patient)
Time 3	FLP OAD HADS	FLP HADS (re. self, not patient)

Table 5.4: Summary of measures completed by patients and proxies at each time point.

Procedure

Patients and proxies were simultaneously interviewed in their homes by 2 interviewers, usually in separate rooms, on each of the 3 occasions. Informed consent was obtained at time 1. Patients were screened for cognitive and communication problems that could affect questionnaire completion at the start of each interview. Interviewers administered the questionnaires orally and recorded the participants' responses. Visual aids - the display of response choices - were provided for the perceived control measures. Table 5.4 contains a summary of measures completed by patients and proxies at each time point.

Analyses

Research Question 1: To what extent do self-report, proxy-report and observed performance assessments of activity correspond?

Correlations between the activity measures at each time point were performed. T-tests were carried out to test whether the differences between self- and proxy-reported activity were significant. Further exploratory analyses were used in investigating the change in patient and proxy scores over time.

Research Question 2: Are measures affected by biases of negative affectivity and gender?

HADS scores were correlated with each activity measure at every time point.

Gender differences on self-report and observed measures were tested with independent t-tests. Correlations between self-report and observed measures for female and male participants were compared using Fisher's test (Howell, 1992).

Question 3: Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

The psychological variables Anxiety, Depression, RLOC and PCI were correlated with Recovery scores (residual activity on regressing activity score at the earlier time point on activity score at the later time point, see Bonetti (2000); Bonetti et al. (2001). Where both emotion and control cognition variables were found to predict a Recovery variable ($p < .15$, from Bendel & Afifi, 1977), hierarchical multiple regression equations were performed. First, emotion variables were entered in stage 1, followed by control cognitions at stage 2. In a second equation, control cognitions were entered at stage 1, followed by control cognitions at stage 2.

Results

Descriptive data for all measures are shown in Table 5.5. 101 participants with complete activity data were included in the analyses. Where the response to a single item was missing from a participant's data, the value of the group mean was substituted for the HADS and RLOC. As other scales contained fewer items, where a value was missing that participant's data was excluded for that measure. Two participants lacked HADS data at one time point; their HADS data was excluded across all time points so that the question of negative affectivity bias could be addressed looking at the same participants' data for each time point.

Tabachnik and Fidell (1996) state that, with larger samples (examples are given with 100 and 200 participants), the significance of skew and kurtosis values are not as important as the actual (undefined) size of the value and the visual appearance of the distribution. Visual examination of the distributions revealed that OAD measures were skewed, showing ceiling effects (see also high skew and kurtosis scores in Table 5.6). Scores were therefore subjected to square root transformations as used by Bonetti et al. (2001) (on the full-length OAD scale) where:

$$\boxed{\text{transformed score} = \sqrt{([\text{maximum score} + 1] - [\text{actual score}])}}$$

This transformation reduced skew from –2.36 to 1.73 (OAD6 T2), -2.78 to 1.87 (OAD6 T3), -2.62 to 2.01 (OAD5 T2) and –2.95 to 2.11 (OAD5 T3). Kurtosis reduced from 5.83 to 2.59 (OAD6 T2), 9.69 to 3.98 (OAD6 T3), 7.34 to 3.82 (OAD5 T2) and from 10.52 to 5.06 (OAD5 T3). Hence, from this point forwards, 'OAD' refers to transformed scores. Higher transformed OAD scores indicate higher activity limitations.

Variable	Measure, time	Mean	SD	α	N
Age	Age	67.78	10.69		101
Activity Limitations	Self-report T1 (FLP)	70.45	31.06		101
	Self-report T2	69.33	32.84		101
	Self-report T3	71.26	30.88		101
	Proxy-report T1 (FLP)	72.71	31.21		101
	Proxy-report T2	67.30	35.69		101
	Proxy-report T3	67.04	34.14		101
	Observer assessed:				
	OAD6 T2 (untransformed)	5.27	1.31	.77	101
	OAD5 T2 (untransformed)	4.48	1.04	.75	101
	OAD6 T3 (untransformed)	5.36	1.12	.70	101
	OAD5 T3 (untransformed)	4.51	.95	.70	101
	OAD6 T2 (transformed)	1.25	.40		101
	OAD5 T2 (transformed)	1.19	.34		101
	OAD6 T3 (transformed)	1.23	.35		101
	OAD5 T3 (transformed)	1.18	.31		101
Negative Affectivity	SHADS T1	12.26	7.91	.85	99
	SHADS T2	10.92	7.53	.85	99
	SHADS T3	9.33	7.28	.85	99
	PHADS T1	12.33	7.67	.88	99
	PHADS T2	11.48	8.04	.88	99
	PHADS T3	10.79	8.12	.90	99
Emotion	Sanx T1	5.64	4.32	.78	99
	Sdep T1	6.62	4.41	.76	99
	Sanx T2	4.85	4.70	.84	99
	Sdep T2	6.07	4.07	.75	99
	Sanx T3	4.28	4.59	.84	99
	Sdep T3	5.05	3.75	.73	99
Perceived Control	RLOC T1	35.00	4.40	.55	101
	RLOC T2	35.44	4.60	.62	101
	PCI T1	.03	-.19	.79	89
	PCI T2	.04	-.07	.78	86

Table 5.5: Mean, standard deviation and Cronbach´s alpha of each measure, at every time point.

OAD6/5 = Observer-assessed activity with/without self-report item
S/P HADS = Patient (Self)/Proxy distress
Sanx/dep = Patient (Self) scores on anxiety/depression subscales of HADS.
RLOC = Recovery Locus of Control scale
PCI = Perceived Control Index

Variable	Measure, time	Skew		Kurtosis	
		Statistic	SE	Statistic	SE
Activity Limitations	Self-report T1 (FLP)	-1.02	.24	.52	.48
	Self-report T2	-1.04	.24	.29	.48
	Self-report T3	-1.13	.24	.65	.48
	Proxy-report T1 (FLP)	-.80	.24	.16	.48
	Proxy-report T2	-.81	.24	-.53	.48
	Proxy-report T3	-.86	.24	-.28	.48
	Observer assessed:				
	OAD6 T2 (untransformed)	-2.36	.24	5.83	.48
	OAD5 T2 (untransformed)	-2.62	.24	7.34	.48
	OAD6 T3 (untransformed)	-2.78	.24	9.69	.48
	OAD5 T3 (untransformed)	-2.95	.24	10.52	.48
	OAD6 T2 (transformed)	1.73	.24	2.59	.48
	OAD5 T2 (transformed)	2.01	.24	3.82	.48
	OAD6 T3 (transformed)	1.87	.24	3.98	.48
	OAD5 T3 (transformed)	2.11	.24	5.06	.48
Negative Affectivity	SHADS T1	1.02	.24	.90	.48
	SHADS T2	.85	.24	.17	.48
	SHADS T3	1.05	.24	1.00	.48
	PHADS T1	.71	.24	-.12	.48
	PHADS T2	1.05	.24	1.33	.48
	PHADS T3	.83	.24	.33	.48
Emotion	Sanx T1	.90	.24	.74	.48
	Sdep T1	.97	.24	.54	.48
	Sanx T2	1.02	.24	.39	.48
	Sdep T2	.70	.24	-.21	.48
	Sanx T3	1.25	.24	-.65	.48
	Sdep T3	.78	.24	.36	.48
Perceived Control	RLOC T1	-.10	.24	-.21	.48
	RLOC T2	.00	.24	-.29	.48
	PCI T1	-.19	.26	-.80	.51
	PCI T2	-.07	.26	-.80	.51

Table 5.6: Skew and kurtosis statistics, with standard errors, of each measure, at every time point.
OAD6/5 = Observer-assessed activity with/without self-report item
S/P HADS = Patient(Self)/Proxy distress
Sanx/dep = Patient (Self) scores on anxiety/depression subscales of HADS.
RLOC = Recovery Locus of Control scale
PCI = Perceived Control Index

Internal reliability of measures

Cronbach's α was calculated for HADS, OAD, RLOC and PCI measures (Table 5.5). Bryman and Cramer (1997) advise that α be ≥.80. It can be seen that the RLOC scale did not show good reliability (.55 and .60), but HADS performed well (.85 - .90), and PCI gave reasonable results (.79 and .78). Including all 6 items, OAD had an α of .77 at time 2 and .70 at time 3. Tables 5.7a and 5.7b show that removing item 15, the self-report item, from OAD has little effect on the scale's internal reliability. Hence the OAD measure used in analyses below will include only items 10 – 14.

OAD item	α if item deleted
10	.77
11	.74
12	.74
13	.72
14	.72
15	.75

Table 5.7a: Cronbach's α if items individually deleted from OAD 2

OAD item	α if item deleted
10	.71
11	.62
12	.63
13	.66
14	.66
15	.70

Table 5.7b: Cronbach's α if items individually deleted from OAD 3

Outliers

Data were scanned for univariate outliers ($z > 3.29$, Tabachnik and Fidell, 1996). Two were found on OAD T2, three on OAD T3 and one on Depression T3. Analyses were performed both with and without the outliers. The majority of analyses were unchanged (in terms of significant results at $p = .05$) and the results presented include outliers. Where analyses did differ in outcome with the exclusion of outliers, results are presented alongside those including the outliers.

Baseline Comparisons.

T-tests were carried out comparing participants whose data was included with excluded participants (where time 1 data available). No significant differences found were found (Table 5.8).

Measure	t (df)		p	N (included)	N (excluded)
S age	-.45	(150)	.66	101	51
P age	-.44	(142)	.66	98	46
Self-report	-.24	(150)	.81	101	51
Proxy-report	1.48	(149)	.14	101	50
Sdep 1	.35	(149)	.73	100	51
Sanx 1	-.44	(149)	.66	100	51
SHADS 1	-.07	(149)	.94	100	51
PHADS 1	-.77	(149)	.44	100	51
RLOC	-1.14	(145)	.26	101	46
PBC	-.35	(128)	.73	89	41
PCI	-.89	(127)	.38	89	40

Table 5.8: Baseline comparisons: Independent t-tests.

Research question 1: To what extent do self-report, proxy-report and observed performance assessments of activity correspond?

Measures and time	Pearson's r	p (2-tailed)	n
SR1 – PR1	.60	.000	101
SR2 – PR2	.69	.000	101
SR2 – OAD 2	.39	000	101
PR2 – OAD 2	.39	.000	101
SR3 – PR3	.75	.000	101
SR3 – OAD 3	.41	.000	101
PR3 – OAD 3	.28	.004	101

Table 5.9: Correlations between activity measures.
SR = Self-Report, PR = Proxy-Report, OAD = Observer Assessed Disability

All measures correlated highly significantly with other measures at the same time point (Table 5.9). Patient and proxy reports were the most highly correlated ($r = .60$, .69 and .75 for each occasion), with correlations between observed limitation and self/proxy report being less strong ($r = .28$ to .41) (see Figure 5.1). William's test for testing the difference between 2 non-independent correlations was performed (Howell, 1992). At time 2, the correlation of the self- and proxy-reported measures ($r = .69$) is significantly different to that of either self- or proxy-reported activity and observer-rated activity ($r = .39$) at $t\ (98) = 3.30$ ($p<.01$). At time 3, the self/proxy correlation ($r = .75$) was significantly different to both the self/observer correlation ($t(98) = 4.13$, $p < .01$) and the proxy/observer correlation ($t(98) = .5.10$, $p < .01$). The difference between self/observer and proxy/observer correlations was borderline significant ($t(98) = 1.995$ (with $df = 60$, critical value for $p < .05 = 2.000$, for $df = 120$, critical value $= 1.980$). Thus, asking the same question of a patient and the patient's long term observer leads to the highest inter-measure correlations but, even in the best case correlation, much variance is not accounted for. The time 3 results also suggest that the observed performance measure is more closely related to self-report measures than proxy-report. This could reflect both observed and self-report measures being directly dependent on the performance (either physical or verbal) of patients.

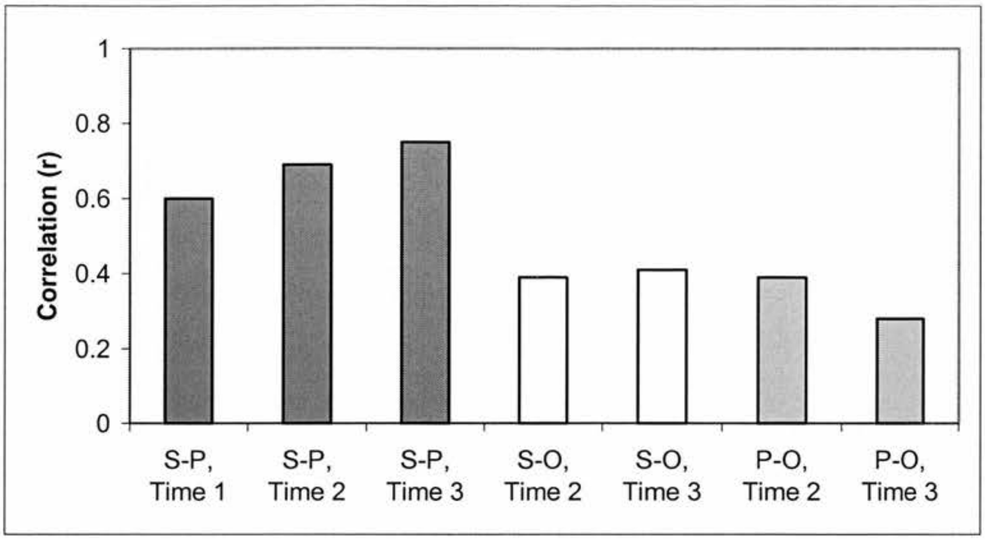


Figure 5.1: Correlations between measures.
S = Self-report; P = Proxy-report; O = Observed.
All correlations p ⩽.01.

Measures and time	t (df)		p	n
SR vs. PR, time 1	-.82	(100)	.42	101
SR vs. PR, time 2	.75	(100)	.46	101
SR vs. PR, time 3	1.84	(100)	.07	101

Table 5.10: Paired samples t-tests comparing self- and proxy-reported activity at each time point. SR = Self-Report, PR = Proxy-Report.

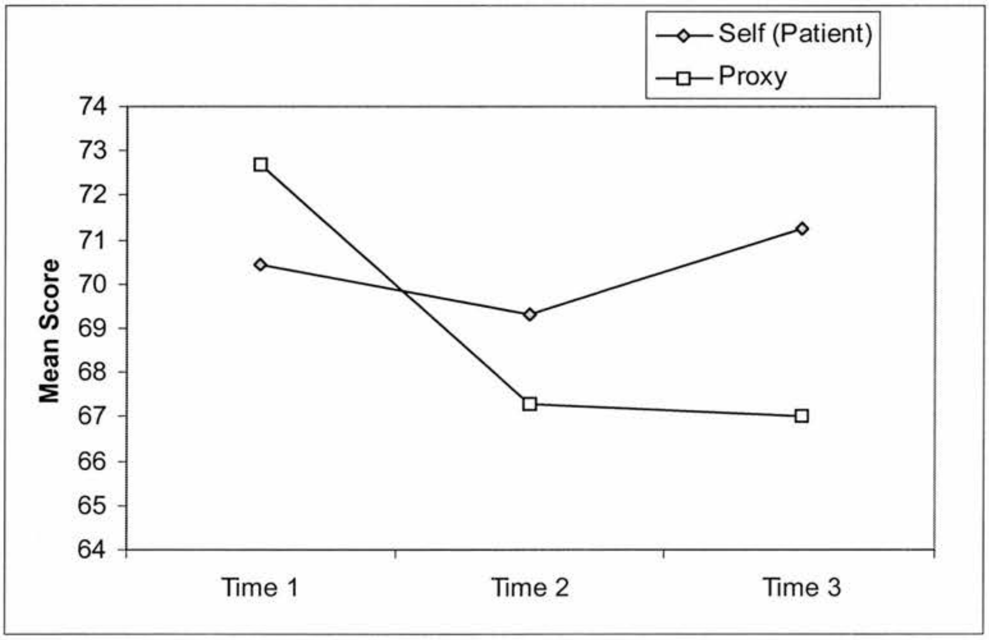


Figure 5.2: Self and proxy-reported activity (FLP), according to time point.

As seen in Table 5.10, paired-samples t-tests found no differences between self and proxy reports of activity to be significant, although a trend towards a difference (with patients reporting greater activity limitations) is seen at time 3 (t(100) = 1.84, p = .07). It should be noted that time 3 was also the time point at which self and proxy activity reports were the most highly correlated ($r = .75$, Table 5.9). A repeated measures ANOVA found effects of neither reporter nor time to be significant ($F(1, 100) = .57$, $p = .45$; $F(2,200) = 1.42$, $p = .24$ respectively), although (using the Greenhouse-Geisser correction for non-sphericity), an interaction came closer to trend level ($F(1.88, 188.41) = 1.97$, $p = .15$, Figure 5.2). Figure 5.2 shows that, after time 1, patients would appear to be reporting worse disability than proxies.

Repeated measures ANOVAs were performed on each of patient and proxy reports separately. Whilst the change over time was insignificant for patients ($F(2,200) = .28$, $p = .75$), an almost significant result was obtained for proxies ($F(2,200) = 2.94$, $p = .055$). Hence it would appear that proxies may be seeing an improvement in activity status whereas patients perceive no change.

Further analyses were conducted to explore this tendency for the discrepancy between patient and proxy ratings to increase over time, even though the correlation between the 2 measures remains strong at time 3 (Table 5.9), with proxy-reports reducing and self-reports remaining unchanged. It should be noted that, with the above findings not quite reaching significance and the hypotheses below being developed post-hoc, results are treated not as conclusive evidence but as possible explanations and potential avenues for further research.
The following hypotheses were raised:

a. *The absolute difference between self and proxy reported activity increases over time.*

b. *This discrepancy only arises in cases where the proxy thinks the patient has improved.*

c. *As activity limitations become less obvious, discrepancy increases.*

Hypothesis a). The absolute difference between self and proxy reported activity increases over time.

1-way ANOVAs were conducted to detect whether the mean differences between patient and proxy reports changed over time. Where the direction of the difference was included in the calculations (self-report – proxy-report), significance was not reached ($F(1.88, 188.42) = 1.97$, $p = .15$, Greenhouse-Geisser correction due to non-sphericity). When the direction of the difference was ignored (|self-report – proxy-report|), a trend effect of change over time was observed ($F(2,200) = 2.29$, $p = .10$). As can be seen in Figure 5.3, the absolute difference between scores decreases over time.

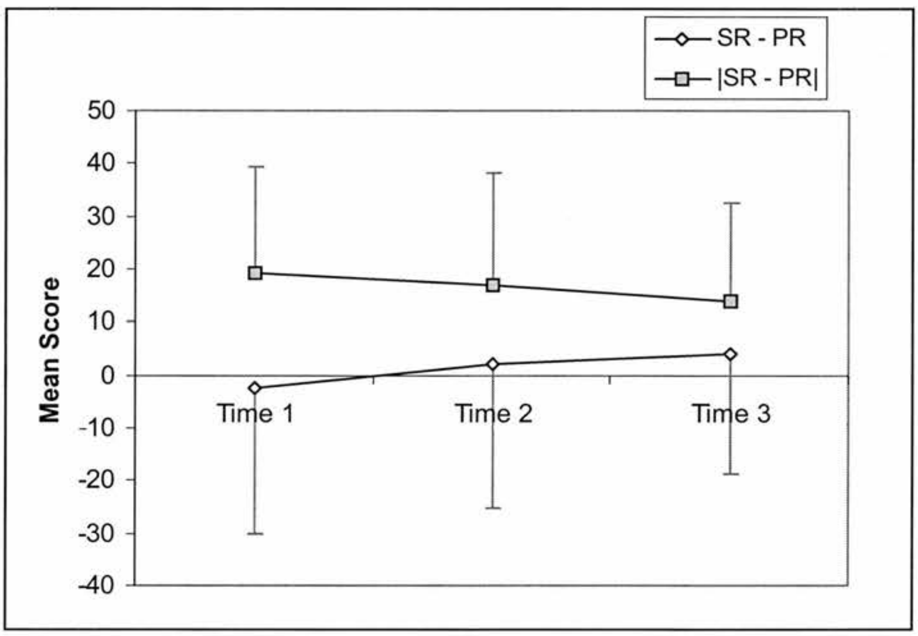


Figure 5.3 : Mean differences, and mean absolute differences between self and proxy-reported activity scores according to time (+/-SD).

Thus, the absolute discrepancy is not increasing over time; it is tending to decrease. However, given that there is a tendency for proxies to give increasingly lower ratings over time than the patient, it is possible that proxies who had scored higher than the patient have reduced their ratings to come in line with the patient. To test this hypothesis, proxies who initially gave higher scores than the patient were compared with proxies with lower initial scores than the patient, to test whether a greater reduction in scores over time is seen when initial scores are higher than the patients' than when the initial score is lower.

Grouping	PR (T1) – PR (T3)		N
	mean	SD	
1) PR (T1) – SR (T1) > 0	14.43	26.03	42
2) PR (T1) – SR (T1) = 0	4.83	21.08	29
3) PR (T1) – SR (T1) < 0	-5.77	26.31	30

Table 5.11: Mean scores (and standard deviations) of the changes in proxy-reported activity between time 1 and time 3. Results are grouped according to whether proxies initially reported higher activity limitations than patients at time 1 (group 1), the same (group 2) or lower initial activity limitations (group 3).
PR = proxy-report
SR = self (patient)-report.

It can be seen that, where proxies initially reported higher activity limitations than patients (group 1), proxy-reported activity limitations decrease between time 1 and time 3 (mean change = 14.43, Table 5.11). This change is significantly greater than 0 (t (41) = 3.59, p = .00). On the other hand, where proxies initially reported lower activity limitations than patients (group 3), proxy-reported limitations appear to increase over time (mean change = -5.77, Table 5.11). This mean is not significantly different to 0, however (t(29) = -1.20, p = .24). Where proxies and patients were initially in agreement (group 2), the mean change is also not significantly different to 0 (t(28) = 1.23, p = .23). It could be that the regression of scores to the mean is being seen. However, if the finding was spurious in this way, the correlation between patient and proxy scores would not be expected to have increased at each time point.

An independent t-test found the change scores of groups 1 and 3 to be significantly different (t(70) = 3.23, p = .00).

Thus it would appear that it is amongst proxies who initially report higher activity limitations than patients that the decrease in reported limitations occurs. Such proxies' scores decrease significantly more between times 1 and 3 than the scores of proxies who initially rate activity limitations as being lower than patients' ratings.

Hypothesis b). The discrepancy between patient and proxy reports only arises where the proxy thinks the patient has improved.

Given that the absolute discrepancy decreases over time whilst the gap between mean proxy and patient scores is greatest at time 3, the possibility that the mismatch only occurs amongst a subgroup of the population – cases where proxies perceive an improvement – was considered. This particular sub-group was selected because the one-way ANOVAs suggested that, whilst no change across time was seen in patients, an almost significant effect was seen with proxies, in the direction of reporting less disability. It seems likely that the gap between mean proxy and patient scores developed amongst participants whose ratings changed.

To examine this possibility, data was grouped as follows:

1. Proxy Report (PR) T1 – PR T3 > 0 (improvement)
2. PR T1 – PR T3 = 0 (no change)
3. PR T1 – PR T3 < 0 (worsened)

For each subgroup, paired samples t-tests were performed comparing patient and proxy reports at time 3. As can be seen in Table 5.12 (illustrated in Figure 5.4), the difference between patient and proxy reports was only significant where proxies perceived an improvement.

	t (df)		**p**	**SR3 mean**	**PR3 mean**	**SR SD**	**PR SD**	**N**
1. Improvement	2.75	34	.01	66.17	52.20	32.42	36.68	35
2. No change	.25	42	.81	73.12	72.49	33.70	34.80	43
3. Worsened	-1.21	22	.24	75.52	79.43	21.81	18.38	23

Table 5.12: T-tests comparing self and proxy-reported activity at time 3.
Cases were grouped according to proxy perceptions of patient activity change.

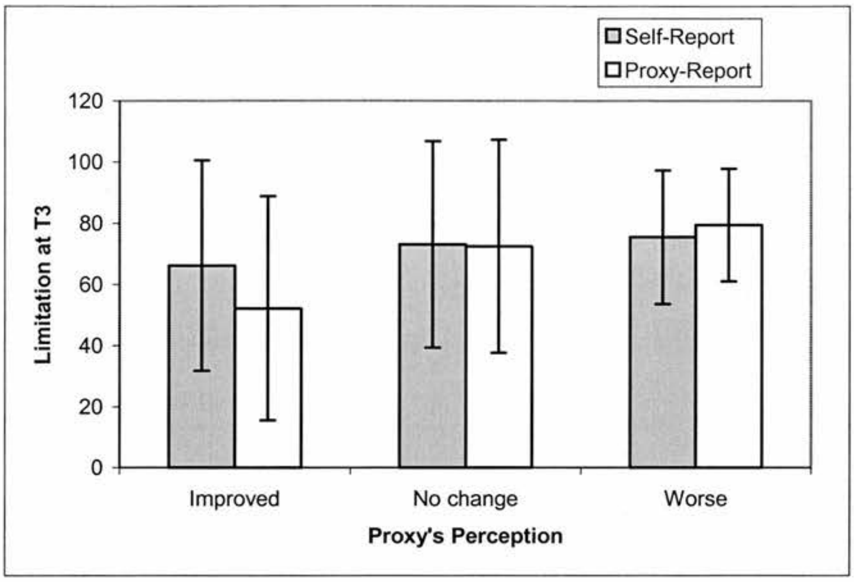


Figure 5.4: Time 3 activity limitation (FLP) as reported by selves (patients) and proxies grouped according to whether the proxies perceived patients' activity limitations to have improved, not changed or worsened.

Additionally, time 3 is the time point at which the lowest number of patients give better activity reports than the proxies (Table 5.13, Figure 5.5).

	T1	*T1 mean, SD*	*T2*	*T2 mean, SD*	*T3*	*T3 mean, SD*
S: better rating (SR < PR)	**42**	58.62, 32.87	**32**	61.97, 27.90	**23**	61.26, 29.80
Equal (SR = PR)	**29**	78.86, 32.35	**39**	64.46, 41.59	**43**	70.95, 36.54
S : worse rating (SR > PR)	**30**	78.87, 21.01	**30**	83.50, 17.90	**35**	78.20, 21.53

Table 5.13: Frequencies of patients rating activity as better/same/worse than proxies , with SR means and SDs.

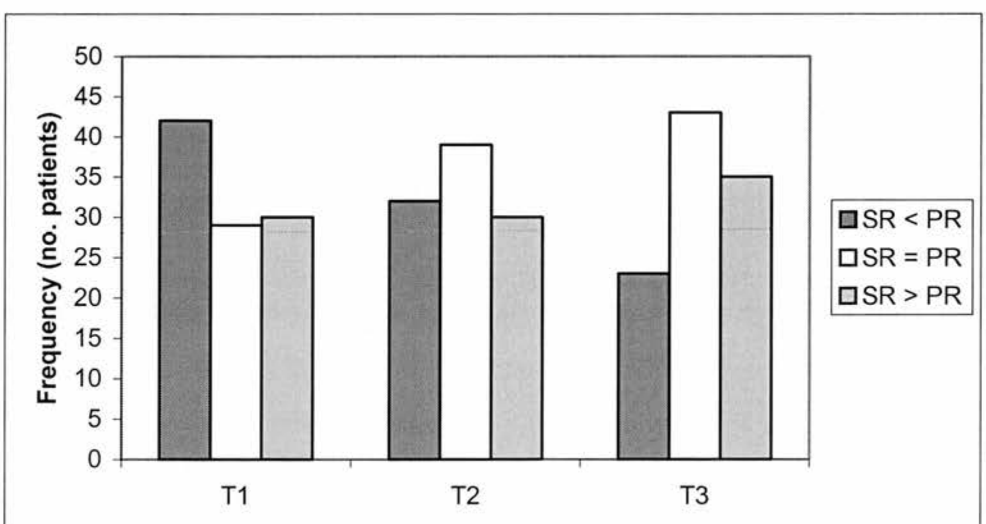


Figure 5.5: Frequencies of patients rating activity as better (SR < PR), the same (SR = PR) or worse (SR > PR) than proxies.

Hypothesis c). Discrepancy increases as activity limitations decrease, becoming less obvious.

Table 5.10 shows that, whilst the difference between self and proxy-reported disability is non-significant at times 1 and 2, a trend effect is seen at time 3. It is possible that, as patients recover and become less limited, activity limitations are less noticeable to the proxy reporter and discrepancies between patient and proxy ratings are greater.

Correlations were performed comparing the difference between patient and proxy reports with activity level as assessed with a) mean of patient and proxy scores and b) OAD score (times 2 and 3). Results are shown in table 5.14.

Time	a) (SR + PR) / 2	b) OAD
1	-.44**	---------
2	-.24*	-.23*
3	-.31*	.02 ns. No outliers: -.25*

Table 5.14: Correlations between patient-proxy absolute discrepancies (|SR-PR|) and activity levels determined by a) mean self/proxy scores and b) by observer ratings at each time point. (** = $p < .01$, * = $p < .05$, ns = non-significant)

It would appear that the absolute difference between patient and proxy measures is inversely related to disability severity. Discrepancies increase as severity decreases when severity is assessed both by averaging proxy and self (patient) reports and with the observed measure, apart from where time 3 discrepancy does not correlate with observed activity. On this latter occasion, on excluding outliers, the correlation increases to -.25, $p = .01$. Figures 5.6a and 5.6b illustrate the position of the outliers. It would therefore seem that, where activity limitations are less severe, and so less obvious, agreement between patients and proxies decreases.

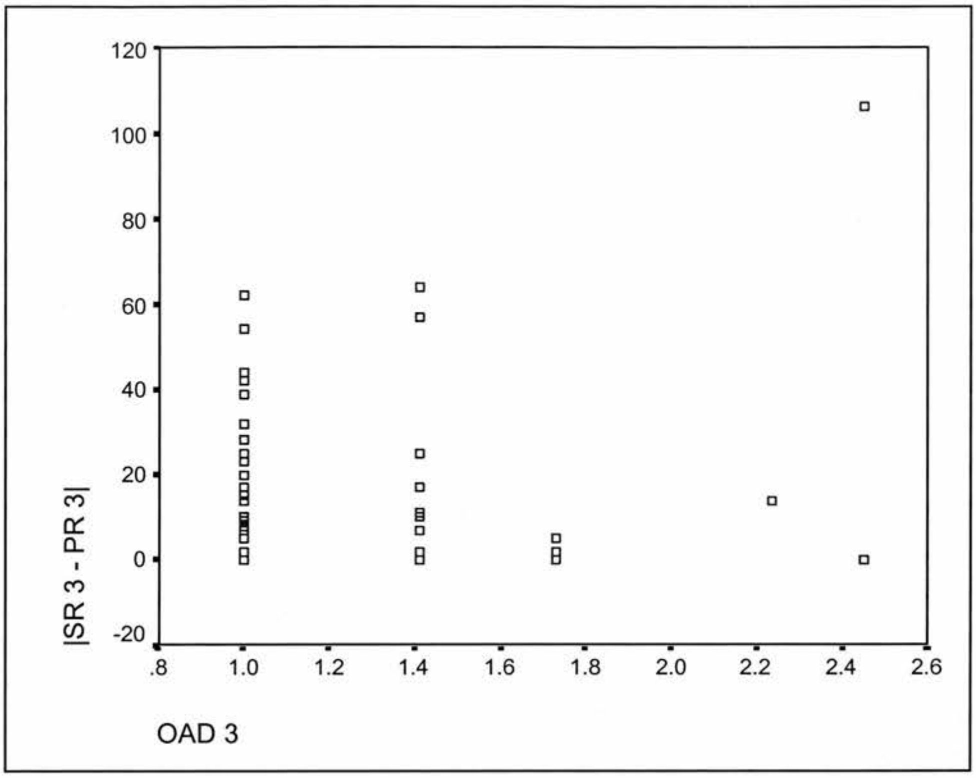


Figure 5.6a: Scattergram of absolute difference between patient and caregiver reports plotted against OAD at time 3. Three outliers included.

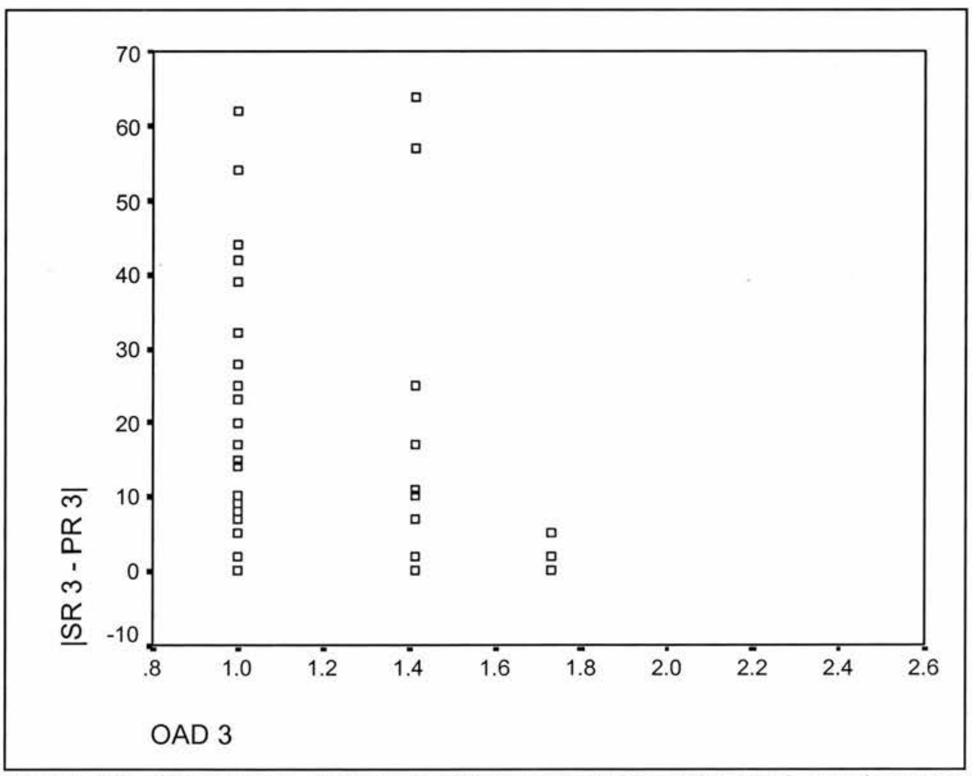


Figure 5.6b: Scattergram of absolute difference between patient and caregiver reports plotted against OAD at time 3. Three outliers excluded.

Research Question 1: Summary of Results.

All measures correlated highly significantly with other measures at the same time point, with the highest correlations being between self (patient)- and proxy-reports but even in the best case correlation, much variance is unaccounted for. Additionally, the time 3 results suggested that the observed performance measure was more closely related to self-report measures than proxy-report. At no time point were patient and proxy reports of activity found to be significantly different, although a trend towards a difference (with patients reporting greater activity limitations) was seen at time 3. Repeated measures ANOVAs suggested that proxies were seeing an improvement in activity status whereas patients perceived no change.

Further, exploratory, analyses suggested that the absolute discrepancy between patient and proxy scores was actually decreasing over time. It appears to be amongst proxies who initially report higher activity limitations than patients that the decrease in reported limitations occurs. This is supported by results indicating that a significant difference between patient and proxy scores at time 3 was only seen amongst proxies who perceived that patients had improved over time. Finally, it would appear that, the less severe the activity limitation, the greater the discrepancy between patient and proxy scores.

Research Question 2: Are measures affected by biases of negative affectivity and gender?

Negative Affectivity

Table 5.15 shows the correlations between activity and negative affectivity (HADS) measures taken at the same time points. The negative affectivity bias hypothesis would predict that correlations between patient Negative Affectivity and self (patient)-reported activity and proxy Negative Affectivity and proxy-reported activity (areas shaded in Table 5.15) would be higher than between proxy Negative Affectivity and self (patient)-reported activity or between patient Negative Affectivity and either proxy- or observer-rated activity. At time 1, even though there is a trend towards a relationship between patient Negative

Affectivity and patient-reported activity (r = .19, p = .06), only the correlation between patient negativity and proxy-reported activity is significant (r = .33, p = .00). Table 5.15 shows correlations between patient Negative Affectivity and patient-reported activity to be very similar to correlations between patient Negative Affectivity and proxy or observer-rated activity at times 2 and 3. In each case, the correlation between proxy Negative Affectivity and proxy-rated activity is the lowest. These results suggest that a bias hypothesis is insufficient to explain relationships between negative affectivity/distress and activity limitations.

Measure	Time	Patient HADS (n = 99)	Proxy HADS (n = 99)
Self-report	T1	.19 (p=.06)	
	T2	.29**	
	T3	.39**	
Proxy-report	T1	.33**	.07 (p = .50)
	T2	.31**	.16 (p = .13)
	T3	.35**	.22*
OAD	T1	--------	
	T2	.29**	
	T3	.39**	

Table 5.15: Pearson's correlations between distress (HADS) and activity limitations as assessed by self (patient), proxy and observer (OAD).
Shaded areas indicate the correlations that would be expected to be the highest, if the negative affectivity bias hypothesis were true. (* = $p < .05$, ** = $p < .01$).

Gender.

Mean self-report and observed activity limitation scores are shown in Table 5.16. In contrast to the Pilot and Walking Difficulties Studies where self-report and observed walking test performances could be directly compared, self-report and observed measures were not directly comparable here. However, independent t-tests were performed to test whether there were gender differences on measures (a difference on one measure only would suggest an effect of gender on measurement) and the correlations between self-report and observed ratings were compared for male and female participants.

Independent t-tests found no difference between female and male scores on any of the measures (Table 5.16).

Measure and Time Point		Mean	SD	t (df)		p
SR FLP (T1)	Female	68.79	31.26	.37	(63.28)	.71
	Male	71.25	31.16			
	Total	70.45	31.06			
SR FLP (T2)	Female	72.85	29.61	-.79	(72.70)	.43
	Male	67.62	34.38			
	Total	69.33	32.84			
SR FLP (T3)	Female	71.18	32.19	.02	(60.42)	.99
	Male	71.29	30.47			
	Total	71.26	30.88			
OAD (T2)	Female	1.19	.32	-.06	(69.57)	.95
	Male	1.19	.35			
	Total	1.19	.34			
OAD (T3)	Female	1.25	.41	-1.41	(43.94)	.12
	Male	1.14	.25			
	Total	1.18	.31			

Table 5.16: Self-report (SR) and observed (OAD) mean activity limitation scores for female and male participants, with results of independent t-tests. (Levene's test showed variances not always equal so equal variances not assumed for all analyses reported in Table 5.16)
N = 68 (female), 33 (male), 101 (total)

Correlations of self-report and observed scores for each gender are shown in Table 5.17. Fisher's test was performed to determine whether there were significant differences between the correlations for female and male participants (Howell, 1992). Neither the correlations at time 2 (z = .02, p = .98) nor time 3 (z = .22, p = .82) were significant (significance values given by SurfStat statistical tables, http://math.uc.edu/~brycw/classes/148/tables.htm#normal). Using the correlation values with outliers excluded at time 3, the difference between the correlations was still non-significant (z = 1.03, p = .31).

Measures correlated	Gender	r	p	N
T2: SR FLP – OAD	Female	.39	.03	33
T2: SR FLP – OAD	Male	.40	.00	68
T3: SR FLP – OAD	Female	.41 No outliers: .31	.02 No outliers: .09	33
T3: SR FLP – OAD	Male	.45 No outliers: .50	.00 No outliers: .00	68

Table 5.17: Pearson's correlations between self-report and observed activity limitation scores for male and female participants.

Research question 3: Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Table 5.18 shows correlations calculated between the psychological variables and activity measures. Residual activity 'recovery' scores were calculated where possible: ie for self and proxy-reported activity at times 2 and 3 and for observer-rated activity at time 3.

Thus, the activity score for self-reported time 2 activity is the residual score when time 2 activity is regressed against time 1 activity (rSR2)

For time 3 self and proxy-reported activity, residuals were calculated from regression equations with both time 1 and time 2 activity measures (r1SR3 and r2SR3). The residual calculated using the time 1 score was correlated with time 1 psychological variables; the residual calculated using the time 2 score was correlated with time 2 psychological variables.

As there was no observed time 1 assessment, following the same logic as for the other activity measures, cross-sectional relationships were investigated with time 2 OAD, and predictive relationships were examined with rOAD at time 3 (ie the predictive effects of time 1 psychological variables on OAD were not looked at as it was not possible to calculate time 2 residuals.

		Anxiety		Depression		RLOC		PCI	
		T1	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**
Activity limitations	**SR T1**	.13ns	---------	.22*	---------	-.24*	---------	-.32**	---------
	PR T1	.21*	---------	.38**	---------	-.31**	---------	-.30**	---------
	OAD T2	-------	.22*	-------	.29**	--------	-.02ns	---------	-.22* (-.15ns)
Recovery	**rSR T2**	.20*	.13ns	.32**	.26**	-.11ns	-.08ns	-.07ns	-.09ns
	rPR T2	.20*	.15ns	.21*	.19^	-.05ns	-.07ns	-.07ns	-.19^
	r1SR T3	.17ns	---------	.30*	---------	-.17^	---------	-.18^	---------
	r1PR T3	.11ns	---------	.07ns	---------	.01ns	---------	-.06ns	---------
	r2SR T3	-------	-.04ns	-------	.24*	---------	-.02ns	---------	-.13ns
	r2PR T3	-------	-.11ns	-------	-.05ns	---------	.04ns	---------	-.03ns
	r2OADT3	-------	.00ns	-------	.23* (.17^)	---------	-.09ns	---------	-.12ns

Table 5.18: Correlations between psychological variables and activity assessments.
^ = trend (p < .10), * = significant at .05, ** = significant at .01 (Pearson's r, 2-tailed).
E.g. rSR2 = the residual self-reported activity score at time 2 when time 1 self-reported activity regressed against time 2 self-reported activity. E.g. r1SR3 = the residual self-reported activity score at time 3 when time 1 self-reported activity regressed against time 3 self-reported activity (cf r2SR3).
Figures in brackets indicate results of analyses excluding outliers, where results differed from results including outliers in terms of significance at the .05 level.

Table 5.18 shows that, cross-sectionally, emotion and control cognitions are associated with activity on most measures. Correlations significant at $p < .10$ are highlighted. Anxiety, Depression and RLOC appear to be more strongly associated with proxy-report than with self-report activity. However, William's test found the correlations with Anxiety and RLOC to not differ significantly by measure ($t(96) = -.91$, $p < .40$; $t(98) = .81$, $p < .50$ respectively). The correlations of proxy and self-report time 1 activity with depression differed at trend level ($t(96) = -1.86$, $p < .10$). This finding is interesting with respect to the negative affectivity hypothesis which would predict self-report activity to have the stronger association with depression, whereas the trend is in the opposite direction.

Predictive correlations show time 2 Recovery derived from self-report measures to be predicted by Anxiety and Depression at time 1; self-report Recovery at time 3 was predicted by Depression at time 1 and time 2. Proxy-determined time 2 recovery was also predicted by Anxiety and Depression but time 3 Proxy Recovery was not predicted by any of the psychological variables. Time 2 recovery was predicted by time 1 Anxiety with equal strength on proxy and self-report scales ($r = .20$, $p < .05$) whilst time 1 Depression appeared to be more strongly associated with self-report measures ($r = .32$, $p < .01$) than proxy-report ($r = .21$, $p < .05$). This difference was not significant however (William's test: $t(96) = 1.26$, $p < .20$). Time 3 Observer-determined Recovery was predicted by time 2 Depression (only at trend level when outliers excluded); Anxiety did not predict time 3 Recovery on this measure.

Control cognitions were found to predict Recovery on few measures, and then only at trend level. Self-report derived Recovery at time 3 was, at trend level, predicted by both time 1 PCI and RLOC. Recovery derived from proxy-report at time 2 was predicted by time 2 PCI, again at trend level.

It would appear, therefore, that self-report Recovery was predicted more consistently by emotion and control cognitions than Recovery derived from proxy-report but that these findings do not reflect a self-report negative affectivity bias. The number of correlations involving observer-assessed Recovery was limited as the measure was only taken at the 2^{nd} and 3^{rd} time points. However, observer-assessed Recovery was predicted by Depression and cross-sectional correlations were seen between OAD and Anxiety, Depression and PCI.

Multiple Regression Equations

Hierarchical multiple regression equations were to be performed where both emotion and control cognitions were found to predict activity recovery at a later time point. This would enable whether emotion and control cognitions would independently contribute to activity as assessed by each measure type to be addressed. Correlations of $p < .15$ were required to bring variables into the multiple regression. However, only one recovery measure: time 3 self-report-derived recovery, was adequately predicted by both emotion and control cognition variables. Correlations between the independent variables did not exceed .80 (Table 5.19, recommendation from Bryman and Cramer(1997)) so multi-collinearity would not appear to be a problem.

Time 1 variables	**Anxiety**	**Depression**	**RLOC**	**PCI**
Anxiety	-------------	-------------	-------------	-------------
Depression	.64 **	-------------	-------------	-------------
RLOC	.10 ns	-.19 ns	-------------	-------------
PCI	-.20 ns	-.32 **	.07 ns	-------------

Table 5.19: Pearson's correlations between the multiple regression independent variables at Time 1.
RLOC = Recovery Locus of Control. PCI = Perceived Control Index. * = $p < .05$, ** = $p < .01$.

Two hierarchical multiple regression equations were run. First, the emotion variables (time 1 Anxiety and Depression) were entered at stage 1 and the control cognitions (time 1 RLOC and PCI) were entered at stage 2. Second, the order was reversed such that time 1 RLOC and PCI were entered at stage 1 and Anxiety and Depression at stage 2.

Multiple regression results are shown in Tables 5.20a and 5.20b. It can be seen that, on entering Anxiety and Depression first (Table 5.20a), Model 1 is significant ($F(2, 85) = 3.79$, $p = .03$). Entering RLOC and PCI in Model 2 did not significantly contribute to the model. RLOC and PCI were entered first in Table 5.20b. In this case, model 1 reached trend level significance ($p = .10$); F change on adding Anxiety and Depression also reached trend level ($p = .10$).

Model	β	p (β)	R^2 change	F	p (F)	Sig F Change
1: Anxiety	-.03	.83	.08	3.79	.03	.03
Depression	.31	.03				
2: Anxiety	.00	.98	.02	2.38	.06	.38
Depression	.23	.12				
RLOC	-.09	.41				
PCI	-.12	.27				

Table 5.20a: Hierarchical multiple regression results. DV = r1SR T3. IVs: time 1 Anxiety and Depression (stage 1), RLOC and PCI (stage 2).

Model	β	p (β)	R^2 change	F	p (F)	Sig F Change
1: RLOC	-.12	.28	.05	2.36	.10	.10
PCI	-.19	.07				
2: RLOC	-.09	.41	.05	2.38	.06	.10
PCI	-.12	.27				
Anxiety	.00	.98				
Depression	.23	.12				

Table 5.20b: Hierarchical multiple regression results. DV = r1SR T3. IVs: time 1 RLOC and PCI (stage 1), Anxiety and Depression (stage 2).

The first stages in each analysis suggest that Depression (Table 5.20a) and PCI (Table 5.20b) are taking up the variances of emotion and control cognitions respectively. To gain a simplified picture, regression equations were therefore repeated including only Depression and PCI. Results are shown in Tables 5.21a and 5.21b.

Model	β	p (β)	R^2 change	F	p (F)	Sig F Change
1: Depression	.29	.01	.08	7.62	.01	.01
2: Depression	.25	.03	.01	4.46	.01	.26
PCI	-.12	.26				

Table 5.21a: Hierarchical multiple regression results. DV = r1SR T3. IVs: time 1 Depression (stage 1), PCI (stage 2).

Model	β	p (β)	R^2 change	F	p (F)	Sig F Change
1: PCI	-.20	.06	.04	3.54	.06	.06
2: PCI	-.12	.26	.06	4.46	.01	.26
Depression	.25	.03				

Table 5.21b: Hierarchical multiple regression results. DV = r1SR T3. IVs: time 1 PCI (stage 1), Depression (stage 2).

Tables 5.21a and 5.21b show Depression to contribute to the equation both when entered as the first stage variable and in stage 2, alongside PCI. In contrast, PCI reaches trend level when entered as the first stage variable but is no longer significant when Depression is added to the equation. It would therefore appear that the emotion variable Depression, but not the control cognition variable PCI, is an independent predictor of self-reported Recovery at time 3.

Discussion

Research Question 1: To what extent do self-report, proxy-report and observed performance assessments of activity correspond?

All three activity measures, self-, proxy- and observer-rated scales correlated significantly with the other measures at the same time point. The highest correlations were found between self- and proxy-rated measures (correlations from r = .60 at time 1 to .75 at time 3 compared with correlations between self/proxy and observer which ranged from .28 to .41). It would seem, therefore, that the highest correlations are seen where a patient's score is compared with that of someone who spends a large amount of time with the patient and who uses the same rating scale. In contrast, the observer witnessed a restricted range of activities over a limited time period. However, even the highest correlation of .75 leaves over 40% of variance unaccounted for. Higher correlations between observer and self/proxy reports were seen by Peck et al. (1989): Self - spouse: r = .87, Self/spouse - observer: .89 and .85 respectively, all correlations p < .001. However, in contrast with the short observed measure used here, Peck et al.'s (1989) observers completed the same measure as index and proxy participants; their 'observed' measure consisted of detailed performancc and interview assessment of patients' abilities on each item. Thus it is likely that the observers gained a much more accurate and detailed report of patients limitations than the observers in the present study.

This study's time 3 results also suggested that the observed performance measure is more closely associated with the self-report measures than proxy-report. This could reflect that both observed and self-report measures rely on patient behaviour – either behaviour as in that physically performed or reporting behaviour. Both measures could, for example, be affected by factors such as negative affectivity (with high negative affectivity reducing activity performed and reported) or social desirability (high social desirability could lead to patients reporting more positively and putting more effort into performance). However, this study's results did not support such a negative affectivity hypothesis, with all measure types being similarly associated with negative affectivity and no measure of social desirability was included in the study.

A trend towards patients reporting greater limitations than did the proxies was seen; it appeared that proxies saw patients' activity status improving whereas patients did not. This is surprising given the literature which consistently reports proxies rating patients as more limited than patients rate themselves (see Introductory Chapter, Table 1.1). Further exploratory analyses suggested that the absolute discrepancy between proxy and patient reports decreased with time and that it was amongst proxies who initially reported higher limitations than the patients that scores decreased. It could be that these patients received less help from the proxy than they initially did and their limitation is more obvious to them than to the proxy. This is supported by the findings of Epstein et al. (1989): the more time proxies spent helping patients, the more impaired the proxies rated the patients, relative to the patients' own reports. Marteau and Johnston (1986) found that parents viewed their children's illness as being less serious as time went on. It could be that, as proxies become more accustomed to the stroke, it becomes less threatening and their perceptions of the patients' limitation changes.

It also seems that, the less severe the limitation, the greater the discrepancy between patient- and proxy rating. Sneeuw, Aaronson, deHaan et al. (1997) and Sneeuw et al. (1998) also found agreement to depend on patient's functioning level; Sneeuw, Aaronson, deHaan et al (1997) found discrepancies greater where patients were more limited whereas Sneeuw et al (1998) found poorer agreement at intermediate levels of functioning. The 1998 sample was reportedly more impaired than that of 1997. It is not clear how the impairment level of the present study's participants compared with those of Sneeuw et al, but it could be that this sample's less impaired participants are comparable with Sneeuw et al 1998's intermediate functioning participants; it would seem likely that, where limitations are either non existent or very severe (where need for assistance is clear to both parties), agreement is high and for the more ambiguous levels of ability concordance is lower.

Research Question 2: Are measures affected by biases of negative affectivity and gender?

The negative affectivity bias hypothesis was not supported; although correlations between Negative Affectivity and self-reported activity were seen they were no higher than between Negative Affectivity and other measures. Gender was also not found to affect activity reports although it was not possible to directly compare self-report and observed performance scores because of the different measures and units used.

Research Question 3: Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Cross-sectionally, most emotion and control cognition variables were associated with activity on all three measures. Whilst significant differences in correlations for self-and proxy-rated activity were not seen for control cognitions, Anxiety was associated with proxy- and observer-rated activity but not self-report activity. Additionally, the correlation of Depression with activity was significantly stronger for proxy- than self-reported activity. As anxiety and depression are components of negative affectivity, these findings further support the absence of a negative affectivity bias for self-report activity in this sample.

In contrast, self-reported Recovery measures were predicted more consistently than proxy-reported Recovery. Both time 2 and time 3 self-reported Recovery were significantly predicted by emotion variables (time 2 by Anxiety and Depression, time 3 by Depression only) whereas Anxiety and Depression predicted only time 2 proxy-reported Recovery. Time 3 observer-rated Recovery was predicted by time 2 Depression only, but analyses involving observer-rated measures were limited as the measure was not presented at time 1.

Control cognitions predicted Recovery only at trend level and, again, self-reported Recovery was the only measure predicted by control cognitions from an earlier time point, with time 3 self-reported Recovery being predicted by both RLOC and PCI at trend level.

These results would suggest, therefore, that self-reported Recovery is the recovery measure most consistently predicted by the psychological variables assessed.

Bonetti et al. (2001) found perceived control measures to predict both self-reported and observed recovery in a stroke patient sample. Anxiety and depression were also found to predict recovery, but only for self-report measures. The present study and measures are a sub-sample of those used by Bonetti et al; the current analyses are limited to activity measure items that only reflect ambulation whereas Bonetti et al. investigated disability over a range of activity limitations. It could be that restricting the number of items in activity measures reduced their reliability and made associations between measures more difficult to detect. However, it may be that these effects are specific to walking activity limitations. In a study of 93 stroke patients, assessed at 2 time points (a subsample of the same study sample as that from which the present study's participants were drawn), Bonetti (2000) found Recovery at time 2 to not be predicted by PCI at time 1. Nevertheless, Bonetti (2000) found that the Self-Efficacy component of PCI did predict Recovery (adj $R^2 = .04$, $\beta = .23$, $p < .05$).

Only one set of variables was entered into multiple regressions because of the small number of predictive correlations between psychological variables, especially control cognitions, and recovery measures. It is therefore not possible to compare multiple regression results across recovery measure types. However, the multiple regression equations showed that, whilst time 1 emotion variables (represented by Depression) contributed to time 3 self-reported Recovery, control cognitions (PCI) did not add to the variance accounted for. It would appear that, whilst control cognitions did not have an effect on self-reported Recovery independently of emotion, emotion did have an effect independent of control cognitions. This finding is contradictive to much of the literature: Lackner and Carosella (1999), Holm et al. (1998), Sullivan et al. (1998) and Johnston, Morrison et al. (1999) found control cognitions to predict activity limitations independently of emotion variables. In contrast, reports of independent effects of emotion are relatively sparse, although Härkäpää et al. (1991) found distress but not health locus of control to be associated with the performance of low back pain exercises. It seems likely that the present results reflect the lack of association between control cognitions and recovery (the highest bivariate correlations only reached trend level) rather than any possibility of control's effects on activity being mediated by emotion. Following Baron and Kenny's (1986) recommendations, emotion would only be shown to mediate the effects of perceived control if perceived control was found to affect emotion, emotion was found to affect recovery and perceived control was found to affect recovery. Correlations between control cognitions and emotion variables were mostly non-significant, ranging from .07 to -.32 (see Table 5.19) and associations between control cognitions and recovery did not exceed trend level.

The question of whether emotion and control cognitions independently affect activity is further addressed in the Experimental Study.

Hence, in this study, whether or not emotion and control cognitions predict activity similarly across activity measures is unclear as limited predictive effects were found. However, self-reported Recovery was predicted more consistently across measures than was Proxy-reported Recovery. Observer-rated Recovery (time 3) was only predicted by Depression but fewer comparisons could be performed with this outcome measure as OAD was not included at the time 1 interview. Results also suggested that differences are not due to simple self-report biases as, in cross-sectional comparisons, correlations of psychological variables with self-report activity limitations were not stronger than those with proxy-reported activity limitations.

Limitations

The Recovery Locus of Control scale was found to have low internal reliability (α = .55 at time 1 and .62 at time 2). In a similar sample of stroke patients, Johnston, Morrison et al. (1999) found the measure to have internal consistency of .64, .77 and .53 at 3 time points in a sample of 71 stroke patients. Hence the reliabilities here are not at odds with the literature, although it is unfortunate that they are at the lower end of the spectrum. As a result of this low reliability, it would have been harder to detect relationships between perceived control as assessed with the RLOC measure and activity measures and some associations may have been missed.

A major improvement of this study compared with some other proxy studies (see Chapter 1) is that, instead of simply requesting that patients and proxy reporters completed questionnaires independently, participants were interviewed in separate rooms, minimising influence. However, it is a limitation that the different types of proxy were not separated; Bond and Clark (2000) found differences in concordance between measures according to the type of relationship between proxy reporters. In this sample, data was not grouped according to proxy type because the vast majority (80.2%) were the patient's spouse.

Conclusions

It would seem that the different measures tap different phenomena; higher associations are seen between measures assessing an identical time span and activity range. No evidence was found to support negative affectivity or gender biases. Self-reported Recovery seemed to be the measure most consistently predicted by psychological variables. Contrary to much of the literature, multiple regression equations suggested that emotion, but not perceived control, independently predicted Recovery derived from self-reported activity limitations. The ideal method to investigate this question would be through using an experimental design; such a design was used in Chapter 6.

Acknowledgements

These data were collected as part of a study funded by a grant from the Scottish Executive Chief Scientist Office to Marie Johnston, Ron MacWalter, Val Morrison and Beth Pollard. Results are found in the report to the Chief Scientist Office (Bonetti et al., 2001) and a sub-sample is described in Debbie Bonetti's PhD thesis (Bonetti, 2000).

CHAPTER 6:
THE EFFECTS OF MOOD AND PERCEIVED CONTROL MANIPULATIONS ON WALKING SPEED
OR
MUSIC AND MAGIC[4]

Abstract

Background

The literature supports both emotion and control cognitions as predictors of activity limitations. However, it is not clear whether each variable has an independent effect on activity or whether the apparent predictive power of one is actually caused by the other. Experimental manipulations of both perceived control and emotion have affected observed performance activity (Fisher and Johnston, 1996a, 1996b) but as post-manipulation measures of the non-manipulated variable (perceived control/emotion) were not taken it is not clear by whether the effect of manipulating perceived control was mediated by changes in emotion and/or vice versa. This study's main aim, therefore, is to manipulate both emotion and perceived control to determine whether each has an independent effect on activity. To reduce demand effects, participants were unaware that the outcome activity measure was being taken.

Research Questions

1. Does manipulating mood and perceived control affect activity levels (walking speed) when participants are unaware that they are being observed?

2.i. To what extent do self-report, aware observed ('Baseline Observed') and 'Unaware Observed' assessments of walking speed correspond?

2.ii. Are measures affected by biases of negative affectivity and gender?

2. iii. Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Methods

112 healthy participants, mainly undergraduate students, completed a questionnaire including measures of self-reported normal walking speed, negative affectivity and baseline mood and perceived control assessments. Baseline walking speed was assessed by asking participants to walk down a corridor and back, at their normal walking speed, whilst counting their paces. Mood and perceived control manipulations were performed, followed by manipulation check measures of mood and perceived control. Participants were thanked, paid (£4) and directed out of the laboratory. An outcome measure of walking speed was taken as participants left, unaware that they were being observed. Participants were recalled, fully debriefed and consent for the use of walking speed data was requested.

---

[4] 'Music and Magic': the study's working title.

Results and Conclusions

The mood manipulation successfully influenced participants' mood but had no effect on walking speed. The perceived control manipulation did not affect either perceived control levels or walking speed. It could be that mood does not independently alter walking performance and that it is necessary to also affect control cognitions. Other explanations for the lack of effect on walking speed are discussed.

The three walking speed measures (Self-report, Baseline Observed and Unaware Observed) correlated significantly but not strongly. No evidence was found to support biases of either negative affectivity or gender. Predictive effects of control cognitions and emotion (anxiety and depression) were not found on any of the activity measures.

Introduction

Control cognitions and emotion have been seen to predict activity in the literature. For example, perceived control ratings have been found to predict recovery from activity limitations in stroke, wrist and low-back pain patients (Härkäpää et al., 1991; Johnston, Morrison et al., 1999; Partridge & Johnston, 1989). Mood has predicted disability with ALS/MND patients (Johnston, Earll et al., 1999) and Härkäpää et al. (1991) found increased distress to be associated with the poorer accomplishment of low-back pain exercises. However, it is not clear whether control cognitions or emotion independently affect activity or whether one mediates the other. Associations between control cognitions and emotion could also obscure which variable actually affects activity limitations.

As reviewed in the introduction chapter, a number of studies have supported an effect of control cognitions on activity limitations independently of mood. Lackner and Carosella (1999) and Holm et al. (1998) found self-efficacy but not distress (Lackner and Carosella) or depression (Holm et al.) to predict activity limitations. Sullivan et al. (1998) and Johnston, Morrison et al. (1999) found control cognitions to predict activity limitations after controlling for emotion variables. However, Härkäpää et al. (1991) found distress but not health locus of control to be associated with low back pain exercises and Langer (1983) found an intervention designed to increase perceived control to increase activity compared with a control condition but no change in perceived control was reported. Greater increases in happiness were seen in the experimental than control group however.

Manipulations of mood and perceived control with back pain patients have resulted in short-term effects on a lifting task performance (Fisher & Johnston, 1996a; 1996b). It is not clear, however, whether perceived control and mood influenced activity independently, or whether perceived control's effects were mediated by mood changes or vice versa. Fisher and Johnston (1996b) manipulated perceived control by asking participants to tell the researcher about times when they either felt in control of situations (increase perceived control) or about times when they felt out of control (decrease perceived control). For each group, performance changed in the predicted direction (improving for 'increase control' participants and worsening for 'decrease control' participants). However, post-manipulation measures did not assess whether the manipulation had affected mood: it could be that increasing perceived control improved participants' mood (Abramson et al., 1978; Bandura, 1986; Smarr et al.,1997) and it was mood that influenced lifting task performance.

Fisher and Johnston (1996a) manipulated anxiety by asking participants to talk about either upsetting events (increase anxiety condition) or good events (decrease anxiety condition). Again, performance changed in the expected direction – decreasing anxiety was associated with improved lifting performance and increasing anxiety lead to worsened performance. However, perceived control was not assessed after the manipulation. Thinking about upsetting or good events may have linked in with control cognitions – a lack of control could be characteristic of upsetting events. Riskind and Rholes (1985) argue that mood manipulations are mediated by cognitions. Hence it is possible that the effects of this mood manipulation were caused by changes in perceived control.

This study, therefore, aims to use manipulations of both mood and perceived control in the same experiment to determine whether each of mood and perceived control independently influences activity.

Also, it is unclear whether such manipulations will be effective when participants are unaware that their activity is being observed. Fisher and Johnston's (1996a; 1996b) participants knew their lifting performance was being recorded and so it is possible that demand effects may have occurred. Bargh, Chen and Burrows (1996) unobtrusively assessed the speed at which participants walked on leaving the laboratory. It is proposed that a similar measure be used here. Bargh et al. (1996) found that participants for whom an elderly stereotype was primed walked more slowly down the hallway on leaving the experiment than control participants. It was suggested that being exposed to a stereotype could influence behaviour non-consciously. It seems plausible, therefore, that manipulations of control cognitions and emotions could also influence this behaviour.

Manipulations

Perceived Control

The perceived control manipulation used by Fisher and Johnston (1996b) was based on the principle of rehearsal: participants were asked to tell the researcher about times when they either felt in control of situations (increase perceived control) or when they felt out of control (decrease perceived control). A similar manipulation was used in this study: participants were asked to write about times when they felt in control of their physical performance (increase perceived control) or times when they felt they were not in control of their performance (decrease perceived control). Wallston, Wallston, Smith and Dobbins (1989) recommend that measures should match a perceived control manipulation and that perceived control should be measured directly. Hence the control manipulation, the manipulation check and the behavioural measure were all activity-focused, rather than manipulating or measuring general control cognitions.

Emotion

Fisher and Johnston (1996a) used a similar manipulation for emotion: participants were asked to talk about either upsetting events (increase anxiety condition) or good events (decrease anxiety condition). A similar method was not used to induce emotion states in the present study because it would be difficult to ensure that control cognitions did not influence mood outcome. An upsetting event could also be an event over which the participant perceived little control. If this study is to have any possibility of disentangling the roles of cognitive perceived control perceptions and emotional distress in determining which aspect has an effect on activity, it is preferable to keep manipulations for the two as independent as is possible.

A method of inducing mood was therefore required which did not depend on cognitive processes, and so would be independent from the construct perceived control. Out of a number of mood induction methods in the literature (see e.g. Westermann, Spies, Stahl and Hesse, 1996), the method that would appear to be least reliant on cognitive processes is the musical mood induction technique. Straseske (1989) taped musical selections to induce mood differences but no differences were found between groups on measures of self-efficacy expectations. A review of the literature revealed that music has been used in a number of studies, usually with success, to produce the mood state required (see Table 6.1). The negative mood state addressed by most researchers is depression. However, evidence suggests that inductions cause multiple affective states rather than pure emotions (see Westerman et al., 1996; Albersnagel 1988)

Musical Mood Induction Procedures

In a review of mood manipulation studies, Clark (1983) reported that, whilst almost all participants responded to musical induction procedures, only 30-50% participants responded to the cognitive Velten mood induction procedure, a well-used method requiring participants to read statements off cards and trying to get into the mood described in the statements (Velten, 1968).

Two requirements for the music selected were that it had no lyrics (to minimise cognitive effects) and that it had been found to be successful when participants were unaware that the manipulation were being performed. Some studies (see Table 6.1) have not only told participants the purpose of the music but have also instructed them to get into the required mood state (e.g. Sutherland, Newman & Rachman, 1982). Such instructions risk demand effects, with participants reporting the required mood whether or not they feel to be in that mood state. Lenton and Martin (1991) found mood change instructions to be necessary and sufficient to produce the requested self-reported mood change. It is not clear whether this was due to participants effectively putting themselves into a mood, or demand effects. Other researchers (e.g. Chastain, Seibert & Ferraro, 1995, see Table 6.1) have found musical mood induction procedures to successfully induce mood without participants being aware of the music's purpose.

Kenealy (1988) found the effects of 'demand' and 'no demand' conditions not to differ on either self-report or behavioural measures of mood. Kenealy's 'demand' participants were told that people generally felt happy and exhilarated or sad and despondent after listening to the music, compared with Lenton and

Martin's (1991) or Clark's (1983) reviewed studies where participants were instructed to put themselves into the expected mood state.

To avoid any possibility that the effects of the musical mood induction procedure were due to perceived demands, participants in the current study were not informed of the purpose of the music until after the experiment had been completed. The pieces chosen were 10-minute selections, a time span fitting within the range of successful manipulations (3-20 minutes, see Table 6.1) and being a suitable length of time to allow a perceived control manipulation to match in time allowed.

Carter, Wilson, Lawson, and Bulik (1995) asked 24 women (12 bulimic, 12 controls) to rate 7 pieces of music according to how likely they would be to help them to lower their mood. Results suggested that the participants' responses varied, leading the authors to conclude that the same piece of music will not be equally effective at eliciting depressed mood for all participants. However, participants were not given the opportunity to fully experience the music as mood inductions (they heard only 1-1.5 minutes of each piece of music) and were not asked to rate their mood after hearing each piece. Although there may be variations in exactly how successful musical mood inductions are for different people, as the studies in Table 6.1 show, such methodology has been found to be successful across the board in a number of studies.

As can be seen in Table 6.1, a piece used to reduce mood in many studies (9 out of 21) was by Prokofiev: 'Russia under the Mongolian Yoke' from Alexander Nevsky, ½ speed. Extracts from Delibes' 'Coppelia' and pieces by Vivaldi and Mozart were popular choices to enhance mood. Neutral mood inductions, where used, differed from study to study. For the present study, 10-minute selections of music were compiled from the 15-minute selections used by Parrott and Sabini (1990). The reducing mood induction included Prokofiev's 'Russia under the Mongolian Yoke', ½ speed. The enhancing mood induction, upbeat Dixieland music, although not used in other studies reported here, was used where participants were not informed that the purpose of the music was to alter mood in contrast to Delibes' Coppelia.

Table 6.1: Summary of music selected for musical mood induction procedures. 'Effective?' column indicates measures on which manipulations were found to cause changes in the expected directions.
A = self-report assessments of mood.
B = count/number writing times
C = incentive ratings
D = memory retrieval.
X = no evidence manipulation effective. (e.g. AX, D = no effect on self-report mood assessments but effect on memory retrieval).
E, N, R = Enhancing, Neutral, Reducing.

Paper	Music used			Time	Aware of purpose?	Effective?
	Enhancing	Neutral	Reducing[5]			
Albersnagel (1988)	**Pilot:** Delibes: 'Coppelia'.	Debussy: 'Prelude l'Apres Midi d'un Faun'.	Sibelius: 'Swan of Tuonela' and (part) Dvorak's '9th Symphony'. 20 mins. *Anxiety-provoking:* Stravinsky: 'The Rite of Spring'.	20 mins	Not stated	
	Experiment: As pilot, but shortened tunes.	Faure: 'Ballad for Piano and Orchestra' (Op 19).	As pilot but without Dvorak. *Anxiety-provoking*: As pilot but shortened tunes.	7 mins	Yes	A (R only)
Brown and Mankowski (1993)	Hubert Law: Jazz version of Bach's Brandenburg Concerto, No. 3.	Not used.	Prokofiev: 'Russia under the Mongolian Yoke' from Alexander Nevsky, ½ speed.	10 mins	Not stated	A
Chastain, Seibert and Ferraro, (1995)	Mozart: 'Concerto for 2 pianos', 3rd movement followed by theme song from 'Chariots of Fire'.	Coyote Oldman: 'Lunar Symphony'.	Barber: 'Adagio for Strings'	Not stated	No	A (E differed to N and R)
Clark and Teasdale (1985)	Delibes: Extract from 'Coppelia'. Repeated.	Not used.	Prokofiev: 'Russia under the Mongolian Yoke' from Alexander Nevsky, ½ speed. Repeated.	7 mins	Yes	A, B, C.

[5] Reducing = sad / depressing unless otherwise stated.

Heatherton, Striepe and Wittenberg (1998)	Not used.	John Adams: 'Common Tones in Simple Time' from 'Chairman Dances'	Prokofiev: 'Russia under the Mongolian Yoke', 'Field of the Dead' from Alexander Nevsky (½ speed).	8 mins	No.	A
Hermans, De Houwer and Eelen (1996)	Vivaldi: 1st and 2nd movements from 'The Spring' , 'The Four Seasons'.	Not used.	Dvorak: Excerpts from 9th Symphony.	Continuous during task	No.	X
Kelvin, Goodyer, Teasdale, Brechin and Wittenberg (1999)	Not used	Kenny Gee: 'Songbird'	Prokofiev: 'Russia under the Mongolian Yoke' from Alexander Nevsky.	Neutral: 7 mins Reducing: 3 mins	Yes	A
Kenealy (1997)	Delibes: Selection from 'Coppelia' (Mazurka)	Not used.	Albinoni: 'Adagio in G minor'	Not stated.	Not explicitly stated.	A
Lewis, Dember, Schefft and Radenhausen (1995)	Expose: 'Come go with me'; Irene Cara: 'Fame'	Not used	Luther Vandross: 'Superstar'; Prince: 'Sometimes it Snows in April'	Enhancing: 9.18 mins Reducing: 9.33 mins	No	A
Martin and Metha (1997)	Vivaldi: 'Concerto No 3 (Autumn); Allegro'. Tchaikovsky's 'Swan Lake Ballet, op 20, Mazurka'. Mozart: 'Eine Kleine Nachtmusik; Allegro' and 'Flute concerto in D major: Allegro'	Reich: 'Variations for wind, strings and keyboards', Debussy: 'La Mer; from Dawn until Noon on the Sea'.	Albinoni: excerpts from 'Adagio', Beethoven: 3rd (Allegro Vivace) and 4th (Adagio-Allegro Vivace) symphonies, Tchaikovsky: 'Romeo and Juliet; Fantasy Overture'.	7 mins	Yes	AX, D

Parrott and Sabini (1990)	**Experiment 3:** Music selection **Experiments 4 & 5:** Upbeat Dixieland music	Not used Not used	Music selection Shostakovich: 15th Symphony, excerpt from 2nd movement (2/3 speed), Prokofiev: 'Russia under the Mongolian Yoke' from Alexander Nevsky (1/2 speed) and ending of Tchaikovsky's 6th Symphony, 4th movement (2/3 speed) (details of music obtained from Parrott 2000, personal communication)	8 mins 15 mins	Yes No	A, B A
Pignatiello, Camp, Elder, and Rasar, (1989)	Non-lyrical selection. Began with neutral selection, and became more elating.	Non-lyrical selection. Began with neutral selection, remained neutral.	Non-lyrical selection. Began with neutral selection, and became more depressing.	20 mins	No	A (E differed to N and R).
Pignatiello, Camp and Rasar (1986)	Non-lyrical selection. Began with neutral selection and became more elating.	Non-lyrical selection. Began with neutral selection, remained neutral.	Non-lyrical selection. Began with neutral selection and became more depressing.	20 mins.	No	A (E differed to R), BX.
Samsom and Rachman (1989)	Segment of 'Eine Kleine Nachtmusik' or Divertimento 136, both by Mozart	Not used.	Albinoni: 'Adagio in G minor' or Barber: 'Adagio pour Cordes'.	Played throughout session.	Yes	A, B.
Stein, Goldman and Del Boca (2000)	Delibes: Segment from 'Coppelia'	Faure: 'Ballad for Piano and Orchestra'.	Not used.	8 mins	Yes	A
Sutherland, Newman, and Rachman (1982)	Participants selected piece of music they felt would influence their mood (happy).	Not used.	Participants selected piece of music they felt would influence their mood (sad).	Not stated.	Yes	A (Change in R but not E significant)
Wenzlaff, Wegner and Klein (1991)	Selections from 'Beleze Tropical, Brazil Classics 1'compiled by David Byrne or Bach's Brandenburg Concerto, No. 3, jazz version by Hubert Law.	Not used.	Prokofiev: 'Russia under the Mongolian Yoke', 'Field of the Dead' from Alexander Nevsky or Keith Jarrett: 'Spheres', movements 6 and 7.	9 mins	No	A

Willner, Benton, Brown, Cheeta, Davies, Morgan and Morgan (1998)	Delibes: Extract from ‘Coppelia’.	Not used.	Prokofiev: ‘Russia under the Mongolian Yoke’ from Alexander Nevsky (½ speed)	3 mins	Yes	A
Willner and Jones (1996)	Delibes: selection from ‘Coppelia’ *Successful*	Not used.	Depressed: Prokofiev: ‘Russia under the Mongolian Yoke’ from Alexander Nevsky (½ speed).	Continued whilst doing task.	Yes	A
Wood, Saltzberg and Goldsamt (1990)	Jazz version of Bach’s Brandenburg Concerto, No. 3, by Hubert Law	1: Chopin: Waltzes No. 11 in G-flat and 12 in F minor. 2: ‘Aerial Boundaries’ played by Michael Hedges.	Prokofiev: ‘Russia under the Mongolian Yoke’ from Alexander Nevsky (½ speed)	10 mins	Not stated	A

Neutral music was not included in the present study. As seen in Table 6.1, where researchers did used 3 music conditions (enhance mood, neutral and reduce mood), mood outcome measures tended not to show differences between all three conditions (e.g. Chastain et al., 1995). Research seems to have concentrated on finding music that will change mood rather than music that will keep moods constant. To control for time, 10 minutes of silence was considered, but what effect sitting in silence for 10 minutes might have on participants' emotions and cognitions was unclear. Hence the neutral mood condition consisted of perceived control manipulations only.

It was possible that the elating music would be more energising than the depressing music and that any relationship between mood condition and walking speed change would be due to perceived energy levels, not mood state. The questionnaire therefore included 'arousal' items so that, should an effect of mood condition be found, this hypothesis could be tested.

The main aim of the present study was to examine the effects of experimental manipulations on walking speed, determining whether emotion and perceived control manipulations have independent effects on this activity. However, questionnaires were included that allowed the research questions considered throughout this thesis to be further investigated. Manipulation check measures assessed emotion and perceived control. Once more, three different methods of assessing activity were used: self-reported walking speed, 'aware observed' walking speed (the baseline walking speed measurement) and 'unaware observed' walking speed (walking speed assessed as participants believed they were leaving the study). Thus it would be possible to not only compare self-reported and knowingly observed activity, but also a measure of activity that should be free from demand effects.

Research Questions

1. Does manipulating mood and perceived control affect activity levels (walking speed) when participants are unaware that they are being observed?

   It is predicted that change in walking speed will be greater when both mood and perceived control levels are manipulated than with either manipulation alone. Change in walking speed is hypothesised to be positive when mood and/or perceived control levels are improved and negative when mood and/or perceived control levels are reduced.

2.i. To what extent do self-report, aware observed ('Baseline Observed') and 'Unaware Observed' assessments of walking speed correspond?

2.ii. Are measures affected by biases of negative affectivity and gender?

2. iii. Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Methods

The study was approved by the University of St Andrews School of Psychology Ethics Committee (see Appendix G).

Design

Research Question 1:

An experimental, between subjects design was employed. After completing baseline measures of mood, perceived control and walking speed, emotion and perceived control, manipulations of mood and perceived control were performed. Participants were randomly allocated to experimental conditions. The dependent variable was the change in walking speed between the baseline measurement and participants' walking speed on leaving the laboratory.

There were 7 experimental conditions (see Table 6.2).
1: enhance mood and perceived control.
2: enhance mood, neutral perceived control.
3: neutral mood, enhance perceived control.
4: neutral mood and perceived control.
5: reduce mood, neutral perceived control.
6: neutral mood, reduce perceived control.
7: reduce mood and perceived control.

		Mood		
		Enhance	*Neutral*	*Reduce*
Perceived control	*Enhance*	1	3	
	Neutral	2	4	6
	Reduce		5	7

Table 6.2: Experimental design. Numbers correspond to conditions listed above; shaded cells indicate conditions not included.

This is essentially a 3 x 3 design (increase, neutral and decrease manipulations for perceived control and mood), only conditions were not included for increasing mood and decreasing perceived control or decreasing mood and increasing perceived control because there would not have been a theoretical prediction for these cells. Manipulation checks were performed to determine the manipulations' effectiveness.

Research Question 2:

Activity measures were compared cross-sectionally to determine the extent to which they were associated.

Measures of negative affectivity, emotion and control cognitions were taken at the same laboratory session as walking speed measures; all except for the perceived control manipulation check (PCS) measure (because of experimental constraints) preceded the walking speed measures in order to assess the predictive effects of emotion and control cognitions.

Participants

Power Calculations

During the planning stages of this project, data analysis was discussed in terms of analysing the data either as a 1-way 7-level ANOVA or as a 2-way 3 x 3 7-celled ANOVA. Power calculations (GPOWER, Faul & Erdfelder, 1992) were performed for a 1-way 7-level ANOVA as fewer participants would be required for the same power level in a factorial ANOVA. It was calculated that, to detect a medium effect ($f = .25$), $n = 231$; to detect a large effect ($f = .4$), $n = 98$. Taking these values and practical constraints into consideration, it was decided to aim for 112 participants, 16 in each experimental condition. In an ordinary (9-celled) 3 x 3 ANOVA, 16 participants in each group gives a power of .89 for main effects and .81 for interaction effect ($\alpha = .05$, effect size $f = .30$) (SOLO, 1992).

112 participants (79 female, mean age 20.91 years,) signed up to take part in the study, responding to a poster placed in the foyer of the psychology department. The poster indicated that participants were required for a study investigating the relationship between beliefs and activity (Appendix F.iii). Most (106) were undergraduate students, 4 were postgraduate students and 2 worked outside the

department. £4, the standard recommended departmental hourly rate, was paid to each participant.

Experimental Manipulations

Mood:

Participants were either asked to listen to music for 10 minutes (the musical mood induction procedure) or no mood task was included (neutral condition). Participants were not informed that the music was designed to alter their mood; they were simply asked to listen to the music and were told that they would be asked some questions about it at a later point. The possibility of asking participants to sit in silence for 10 minutes as a neutral control condition was considered but the idea was discarded because it was unclear what effect sitting in silence for 10 minutes might have on participants' mood. On the few occasions when, at some point before the full debriefing, participants asked why the music was played, they were told that experiment time was being controlled across different experimental groups and that the music was played to stop their minds from wandering.

At debriefing, 27 participants (42.2% of those given a music manipulation; 13 participants in decrease and 14 in increase mood conditions) said that they suspected that the music's purpose had been to affect their mood. Of the 37 remaining participants who had received a music manipulation, 11 (9 in decrease and 2 in increase mood conditions) thought the music was designed to be relaxing and 6(all in increase mood condition) thought its purpose was to be energising.

Enhance mood condition: Participants listened to elating music lasting for 10 minutes. The music was selected from a 15-minute recording used by Parrott and Sabini, (1990) (see Table 6.1):

That's a Plenty performed by Eli's Chosen Six (3:15)
W&L Swing by Louis Armstrong (3:33)
Original Dixieland One Step performed by Eli's Chosen Six (2:48)

Neutral condition: no music.

Reduce mood condition: Participants listened to 10 minutes of depressing music. The music selected has been used with success by Parrott and Sabini, (1990) (see Table 6.1):

Extract from ending of *Symphony No. 6, 'Pathetique', op 74, 4th movement (finale: adagio lamentoso)* by Tschaikovsky, played at 2/3 speed. (2:45) *Russia under the Mongolian Yoke* from *Alexander Nevsky*, by Prokofiev, played at ½ speed (7:09).

The latter track has been used to depress mood by a number of researchers, for example Brown and Mankowski (1993), Clark and Teasdale (1985), Kelvin et al. (1999), Wenzlaff et al. (1991) and Wood et al., (1990)

Perceived Control:

The perceived control manipulation was presented as the second part of the questionnaire. Potential items were piloted on 7 people (6 postgraduate students and one research fellow) and the items selected were those rated as being those where participants easily understood what the item was requiring them to do.

Participants were told that they had 10 minutes to spend on the single item. On debriefing, 9 participants (8.0%) suspected that this questionnaire section was designed to affect how they felt or thought. Six of these were in the neutral condition.

Enhance perceived control condition:

'Can you give examples of times when you performed well physically and felt you were in control of your performance? Describe your thoughts at these times. You have 10 minutes to spend on this task.'

Neutral condition:

'Write about times when people are active in their everyday lives. You have 10 minutes to spend on this task.'

Reduce perceived control condition:

'Can you give examples of times when you performed poorly physically and felt you were not in control of your performance? Describe your thoughts at these times. You have 10 minutes to spend on this task.'

Measures

Walking speed measures

1. Baseline, 'aware' observation of walking speed ('Baseline Observed')

Participants were asked to walk down the corridor outside the laboratory and back. The time taken to walk 24m was recorded by the experimenter with a stopwatch (see Figure 6.1: A → C → A). Participants were asked to walk at their normal walking speed and were asked to count their steps as they went. They were informed that this was because we wished to obtain a measure of gait, but the actual intention was to distract them from their walking speed so that as natural an estimate of walking speed as possible could be acquired.

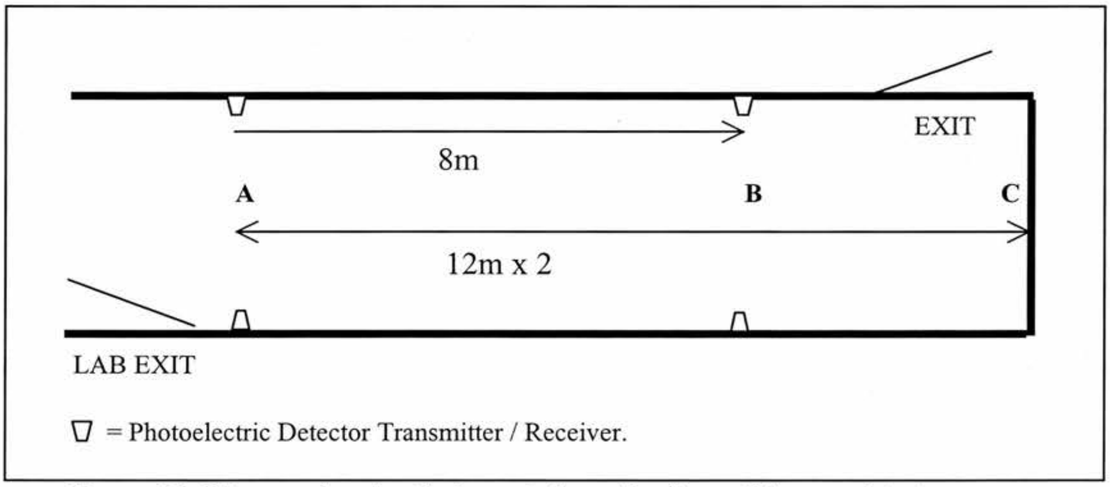


Figure 6.1: Diagram showing the layout of corridor for walking speed tests

2. 'Unaware' observation of walking speed ('Unaware Observed')

Participants walking speed as they left the laboratory was measured (A→ B, Figure 6.1). At this point, participants believed that the experiment was over and were unaware that their walking speed was being measured. Two Photoelectric Detector transmitter/receiver pairs were attached to the corridor walls at ankle-level (at A and B, Figure 6.1). These devices sent beams of infrared light from the transmitter to receiver sides. When the light beam was broken, a small electrical pulse was given out, triggering a stop-clock situated in the laboratory. The system was designed, by Lesley Noble, so that the stop-clock would only start when the devices at A were passed, and would then only be stopped on passing B. This helped to reduce the number of occasions on which the clock was triggered by other people walking down the corridor.

3. Self-reported walking speed.

This measure was taken using an item from the Paffenbarger Physical Activity Questionnaire (Paffenbarger et al., (1978); see Pilot Study (Chapter 2) for further details). The item requested that the participant report their usual walking speed:

What is your usual pace of walking? (please check one)	
a. Casual or strolling (less than 2mph) ☐	b. Average or normal (2-3mph) ☐
c. Fairly brisk (3-4mph) ☐	d. Brisk or striding (4mph or faster) ☐

Mid-point walking speeds were taken for each category (a → 1mph, b → 2.5mph, c→ 3.5mph, d → 4mph) and scores were converted to metre/s to be comparable with other walking speed measures. Due to the categorical nature of this scale, non-parametric analyses were used where possible.

Manipulation checks

Manipulation check scales were designed to gauge participants' perceived control and mood levels post-manipulation at the same time as taking as little time as possible so that participants would leave the laboratory whilst any effects on perceived control or mood were active.

1. Perceived Control Over Activity Visual Analogue Scales: Perceived Control Score (PCS)

Participants were given 3 visual analogue items (see Appendix C.vii) to be completed both before the manipulations were performed and at the end of the session, before the final walking speed measurement was taken. Participants were requested to mark a 100 mm line marked 'agree' on the left hand side and 'disagree' on the right hand side to indicate the extent to which they agreed with the following statements designed to test perceived control over activity levels:

1. I am in control of being as active as I would wish to be.
2. It is difficult to be as active as I would wish.
3. I am confident in my ability to be as active as I would wish.

These items were selected after potential items (based on (Conner & Norman, 1996), see also Chapter 1) were piloted on 7 people (6 post-graduate students from a range of disciplines and 1 research fellow) who were requested to rate items according to ease of comprehension. Scores were obtained by measuring the distance along the line to the participant's mark, in the direction of higher scores indicating greater Perceived Control. The PCS was obtained by summing the scores of the three items. Cronbach's α s were found to be .76 before ($n = 111$) and .73 ($n = 112$) after the manipulation (see Table 6.3)

2. *Mood & Arousal Visual Analogue Scales*

This measure consisted of items assessing participants' mood and arousal. Arousal items were included so that, should the musical mood manipulation be effective in influencing walking speed, it would be possible to check whether it was changes in mood or arousal that were responsible for the effect. The Arousal scores were not analysed.

Participants were asked to rate, on 7 x 100mm visual analogue scales labelled 'not at all' to 'extremely' (see Appendix C.viii) the extent to which they were feeling happy, energetic, distressed, content, sluggish, sad and sleepy. The items 'energetic', 'sluggish' and 'sleepy' were taken from the 'arousal' dimension of Mackay et al.'s Mood Test (Mackay, Cox, Burrows, & Lazzerini, 1978). 'Distressed' and 'contented' were also taken from Mackay et al's Mood Test (from the 'stress' dimension), being the 2 items that mostly fitted the 'depressing' and 'elating' purposes of the musical mood induction procedure. To increase reliability, 2 further 'mood' adjectives were added: 'happy' and 'sad'. Matthews, Jones and Chamberlain (1990) found the adjectives 'happy' and 'sad' to load on the same factor ('hedonic tone') as 'contented' and 'depressed' in evaluating the UWIST Mood Adjective Checklist.

Thus, 'Mood Score' consisted of responses to the items: 'distressed', 'contented', 'sad' and 'happy'. Scores were the measured distances along the lines to the participant's mark in the direction of higher scores indicating more positive mood. The 4 scores were summed. Cronbach's α was .80 before and .76 after the mood/perceived control manipulations (n = 112).

The 'Arousal Score' consisted of responses to the items: 'energetic', 'sluggish' and 'sleepy'. Again, scores were the distances along the lines to the participant's mark, with higher scores indicating higher arousal. The 3 scores were summed. Cronbach's αs, pre- and post-manipulations, were .63 (n = 111) and .72 (n = 112).

Potential Bias Source

Negative Affectivity: The Positive and Negative Affect Schedule (Watson et al., 1988) (PANAS) (Appendix B.i)
PANAS is described in the Pilot Study (Chapter 2). Time frame: 'right now, that is, at the present moment'.

Psychological Predictors: Emotion and Perceived Control

Emotion: Anxiety and Depression: Hospital Anxiety and Depression Scale (HADS) (Zigmond & Snaith, 1983) (Appendix C.i)
This scale is described in the Pilot Study (Chapter 2).

Generalised Self-Efficacy: Generalised Self-Efficacy Scale (GSE) (a short version of this scale is outlined in Jerusalem and Schwarzer (1992); for further details see Johnston, Wright and Weinman, 1995). (Appendix C.ii)

The scale is described in the Pilot Study (Chapter 2). The scale has been included in this study as a general level perceived control measure.

Music questions.

As participants had been informed that they would be asked questions about the music listened to, 2 music items were included post-manipulation for participants in music conditions:

1. Of what origin was the music?
    a. British
    b. American
    c. Russian
    d. Spanish
2. How familiar are you with this type of music?
    a. Very familiar
    b. Moderately familiar
    c. Slightly familiar
    d. Unfamiliar

These data were not analysed.

Procedure

1. Participants were tested individually in the Wolfson Health Psychology Laboratories, in the School of Psychology. They read an information sheet and were asked to sign a consent form (Appendix F.vi). Participants were not aware at this stage that mood and perceived control manipulations were to be performed or that their walking speed would be recorded on leaving the laboratory.
2. The experimenter gave participants the first part of the questionnaire (HADS, PANAS, GSE, PPAQ, Perceived Control, Mood and Arousal Visual Analogue Scales). Participants were left to complete this in their own time and were asked to inform the experimenter, who waited in an adjoining room, when they had finished.
3. Participants were taken to the corridor outside the laboratories and the 'Baseline Observed' walking speed measurement was taken.
4. Mood and perceived control manipulations were performed, the order of presentation being counterbalanced. In each case, the participant was left with the manipulation materials, the experimenter returning with the next task after 10 minutes.
5. The final part of the questionnaire was given. This consisted of the music questions and Perceived Control, Mood and Arousal visual analogue scales.

Participants were asked to complete this section in their own time and to inform the experimenter on completion.

6. Participants were thanked and received payment. The experimenter walked to the laboratory exit with them, directing them down the corridor in order to ensure that they walked past the photoelectric detector beams. This triggered the stop-clock, recording the 'Unaware Observed' walking speed measurement. On debriefing, only 2 participants (1.79%) reported suspecting the unaware observed measurement was being taken.

7. The experimenter called back participants for debriefing. They were informed that the second, 'Unaware Observed' walking speed measurement had been taken and permission to use this data was requested. No participant refused to give permission; all of those where a measurement had been successfully taken signed forms giving consent for this data to be used (see Appendix F.v). Participants were asked if they had been aware that any of the tasks might have been designed to affect how they felt or thought, and whether they were aware that their walking speed was being measured as they left the laboratory. The mood and perceived control manipulations were explained, and participants were asked to not mention the experimental methods used to other people who might wish to participate. During debriefing, the music from the positive manipulation was played, regardless of initial experimental condition, to undo any potential negative mood effects of either mood or perceived control manipulations.

Analysis

Research Question 1: Does manipulating mood and perceived control affect activity levels (walking speed) when participants are unaware of observation?

Data were analysed using a 2-way Univariate ANOVA described by SPSS (1999) as Type IV. This analysis is designed for use where cells are missing from the conventional 9 cell (3 x 3) situation (SPSS, 1999). The DV was change in walking speed between pre- and post-manipulation measurements.

Similar ANOVAs were performed with PCS and Mood Score change as the dependent variables.

Correlations were performed between change in walking speed and change in PCS and Mood Scores to test whether changes in walking speed were seen where there were changes in Mood Score and PCS, regardless of experimental condition.

Hypotheses:

a. Enhancing/reducing mood will affect walking speed.

b. Enhancing/reducing perceived control will affect walking speed.

c. Enhancing both mood and perceived control will have greater effect than each individually.

d. Reducing both mood and perceived control will have greater effect than each individually.

Research Question 2i: To what extent do self-report, aware observed ('Baseline Observed') and Unaware Observed assessments of walking speed correspond?

Spearman's rho was used to test the correlations between Self-report, Baseline observed and Unaware observed walking speed. The non-parametric test was used because, although the observed walking speed was interval data, the self-report measure required participants to choose one of four walking speed categories (see Measures).

Research Question 2ii: Are measures affected by biases of negative affectivity and gender?

Negative Affectivity

Correlations (Spearman's rho) of Negative Affectivity with each of the 3 walking speed measure types were calculated. If the negative affectivity hypothesis were true, the highest correlation would be with self-reported walking speed.

Gender

A measure x gender repeated-measures ANOVA was performed.

Research Question 2iii. Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Correlations (Spearman's rho) between perceived control (PCS, GSE), Anxiety and Depression and each measure of walking speed were performed. Should associations be found ($p < .15$, from Bendel and Afifi, 1977), multiple regressions would be run to test whether emotion and perceived control accounted for different or shared variance: IVs: Anxiety, Depression, GSE and PCS; DV: 1. Self-reported walking speed, 2. Baseline observed walking speed, 3. Unaware observed walking speed.

Results

Descriptive statistics are shown for all measures in Table 6.3. Table 6.4 details reasons for missing data where a full data set (N = 112) was not available for a measure.

Tabachnik and Fidell (1996) state that, with larger samples (examples are given with 100 and 200 participants), the significance of skew and kurtosis values are not as important as the actual (undefined) size of the value and the visual appearance of the distribution.
Histograms were inspected for normality; skew and kurtosis scores are also presented in Table 6.5. Depression and Negative Affectivity were found to show skew values greater than 2 (skew = 2.54 and 2.14 respectively). Visual inspection of the distributions suggested that they were being distorted by an outlier (Participant 69, $z = 5.50$ for Depression, 4.86 for Negative Affectivity). Excluding this outlier reduced the skews to 1.81 (Depression) and 1.69 (Negative Affectivity). Kurtosis values also decreased from 10.36 to 6.22 and from 6.05 to 3.33 respectively. Nevertheless, this participant's data were included in analyses reported because a) analyses involving Depression and Negative Affectivity were non-parametric and b) Participant 69's data was treated like any other outlier (see below): analyses were performed both with and without outlier data and any altered results were noted.

Data were scanned for univariate outliers ($z > 3.29$, Tabachnik and Fidell, 1996). One outlier was found on each of Mood Score 1, Mood Score 2, Mood Score Change, and GSE and two were found on Negative Affectivity and Depression. Analyses were performed both with and without outlier data. In most cases, no changes were seen (in terms of significance at $p < .05$) and results presented include outliers. Where a difference was found, results excluding outliers are presented alongside the whole sample's results.

Variable	Measure, time	Mean	SD	α	N
Age	Age	20.91	4.40		110
Walking	Baseline observed speed (m/s)	1.44	.16		110
	Unaware observed speed (m/s)	1.44	.16		99
	Speed change (m/s)	-.03	.17		98
	Number of steps counted	43.25	4.41		112
	Self-reported usual walking speed	1.37	.30		112
Emotion	Mood Score 1	298.63	60.18	.80	112
	Mood Score 2	298.67	59.78	.76	112
	Mood Score change	.04	44.14		112
	Arousal Score 1	191.80	51.05	.63	111
	Arousal Score 2	185.56	56.52	.72	112
	Arousal Score change	4.18	48.18		111
	Anxiety (HADS)	6.86	3.52	.78	112
	Depression (HADS)	3.26	2.86	.71	112
	Depression (no P 69)	3.12	2.44	.60	111
	Distress (HADS)	10.12	5.52	.82	112
	Distress (no P 69)	9.87	4.91	.77	111
Perceived Control	Perceived Control Score 1	159.81	70.37	.76	111
	Perceived Control Score 2	163.50	64.24	.73	112
	PCS change	4.73	24.18		111
	GSE	31.03	3.64	.77	111
Negative Affectivity	Negative Affectivity	13.72	4.17	.78	111
	Negative Affectivity (no P 69)	13.54	3.71	.74	110

Table 6.3: Descriptive statistics. Mean, standard deviation, Cronbach's α (where appropriate) and N for each measure.

Measure	No. participants missing data	Reason
Baseline Observed speed	2	Experimenter error
Unaware Observed speed	13	5 Participants walked wrong direction 6 Photoelectric beam blocked/triggered by passers-by 2 Experimenter error
Arousal 1	1	Participant error
PCS1	1	Participant error
GSE	1	Missing data
Negative Affectivity	1	Participant error

Table 6.4: Causes of missing data.

Variable	Measure, time	Skew		Kurtosis		N
		Stat	SE	Stat	SE	
Walking	Baseline Observed speed (m/s)	.17	.23	-.32	.46	110
	Unaware Observed speed (m/s)	.22	.24	.11	.48	99
	Speed change (m/s)	.09	.24	.17	.48	98
	Number of steps counted	-.06	.23	.04	.45	112
	Self-reported usual walking speed	-.94	.23	1.01	.45	112
Emotion	Mood Score 1	-1.03	.23	1.16	.45	112
	Mood Score 2	-1.03	.23	1.72	.46	112
	Mood Score change	-.49	.23	7.80	.46	112
	Arousal Score 1	-.25	.23	-.58	.46	111
	Arousal Score 2	-.21	.23	-.68	.45	112
	Arousal Score change	.72	.23	.86	.46	111
	Anxiety (HADS)	.51	.23	-.04	.45	112
	Depression (HADS)	2.54	.23	10.36	.45	112
	Depression (no P 69)	1.81	.23	6.22	.46	111
	Distress (HADS)	1.47	.23	4.96	.45	112
	Distress (no P 69)	.73	.23	1.37	.46	111
Perceived Control	Perceived Control Score 1	.21	.23	-.43	.46	111
	Perceived Control Score 2	.19	.23	-.39	.45	112
	PCS change	-.47	.23	1.48	.46	111
	GSE	-.56	.23	1.00	.46	111
Negative Affectivity	Negative Affectivity	2.14	.23	6.05	.45	111
	Negative Affectivity (no P 69)	1.70	.23	3.33	.46	110

Table 6.5: Skew and kurtosis statistics and standard errors for each measure.

F_{max} was calculated for change in walking speed, mood and perceived control grouped according to condition. Tabachnik and Fidell (1996) report F_{max} to be the ratio of the largest cell variance to the smallest. An F_{max} value up to 10 suggests that data have sufficient homogeneity of variance. F_{max} scores were found to be 2, 8.87 and 3.15 for change in speed, mood and perceived control respectively.

Research Question 1: Do manipulations of mood and perceived control affect activity levels (walking speed) when participants are unaware of observation?

Baseline walking speeds, Perceived Control and Mood scores for each experimental group are shown in Table 6.6. One-way ANOVAs found no differences between the seven groups on any of the three measures (baseline walking speed: $F(6, 104) = .67$, $p = .67$; Mood Score: $F(6, 105) = .60$, $p = .73$; PCS: $F(6, 104) = .62$, $p = .72$).

Condition		Baseline Observed walking speed (m/s)	PCS 1	Mood Score 1
1	**Mean**	**17.24**	**152.09**	**292.63**
	SD	1.72	64.82	76.08
	N	16	16	16
2	**Mean**	**16.82**	**166.16**	**319.31**
	SD	1.87	69.94	39.70
	N	16	16	16
3	**Mean**	**16.69**	**179.20**	**303.01**
	SD	2.46	63.79	50.20
	N	16	15	16
4	**Mean**	**17.43**	**155.16**	**298.00**
	SD	3.26	80.05	54.58
	N	16	16	16
5	**Mean**	**17.45**	**174.09**	**304.97**
	SD	1.76	69.83	71.78
	N	16	16	16
6	**Mean**	**16.25**	**138.25**	**285.06**
	SD	1.70	83.59	77.19
	N	15	16	16
7	**Mean**	**16.73**	**154.94**	**287.44**
	SD	1.66	62.49	44.04
	N	16	16	16

Table 6.6: Mean Baseline Observed walking speed, PCS and Mood Scores according to experimental condition. Key to conditions:
1 = both mood and control enhanced
2 = mood enhanced
3 = perceived control enhanced
4 = neutral
5 = mood reduced
6 = perceived control reduced
7 = both mood and perceived control reduced

Type IV 2-way Univariate ANOVA

Speed change scores for all seven experimental conditions are shown in Table 6.6. Using change in walking speed as the dependent variable, main effects of neither mood nor perceived control condition were found ($F(2, 91) = .08$, $p = .92$; $F(2, 91) = .325$, $p = .72$). The effect of an interaction between mood and perceived control in predicting change in walking speed was also non-significant ($F(2,91) = .49$, $p = .61$). As shown in Figure 6.2, whilst none of the manipulations increased walking speed, the most negative effect on walking speed appears to have occurred in condition 1, where both mood and perceived control were enhanced. This non-significant effect is in the opposite direction to that expected.

Condition		Speed change (m/s)	PCS change	Mood Score change
1	**Mean**	**-.07**	**14.26**	**19.09**
	SD	.17	30.38	62.36
	N	16	16	16
2	**Mean**	**-.03**	**9.50**	**13.88**
	SD	.18	28.56	20.94
	N	14	16	16
3	**Mean**	**-.00**	**-.23**	**-3.85**
	SD	.15	22.56	42.64
	N	14	15	16
4	**Mean**	**-.04**	**.44**	**11.29**
	SD	.15	22.51	23.93
	N	13	16	16
5	**Mean**	**-.05**	**-.91**	**-29.81**
	SD	.19	24.65	40.66
	N	14	16	16
6	**Mean**	**-.03**	**1.72**	**-8.66**
	SD	.12	21.15	46.35
	N	13	16	16
7	**Mean**	**-.01**	**8.00**	**-1.64**
	SD	.20	17.11	42.84
	N	14	16	16

Table 6.7: Mean walking speed, PCS and Mood Change scores according to experimental condition. Key to conditions:
1 = both mood and control enhanced
2 = mood enhanced
3 = perceived control enhanced
4 = neutral
5 = mood reduced
6 = perceived control reduced
7 = both mood and perceived control reduced

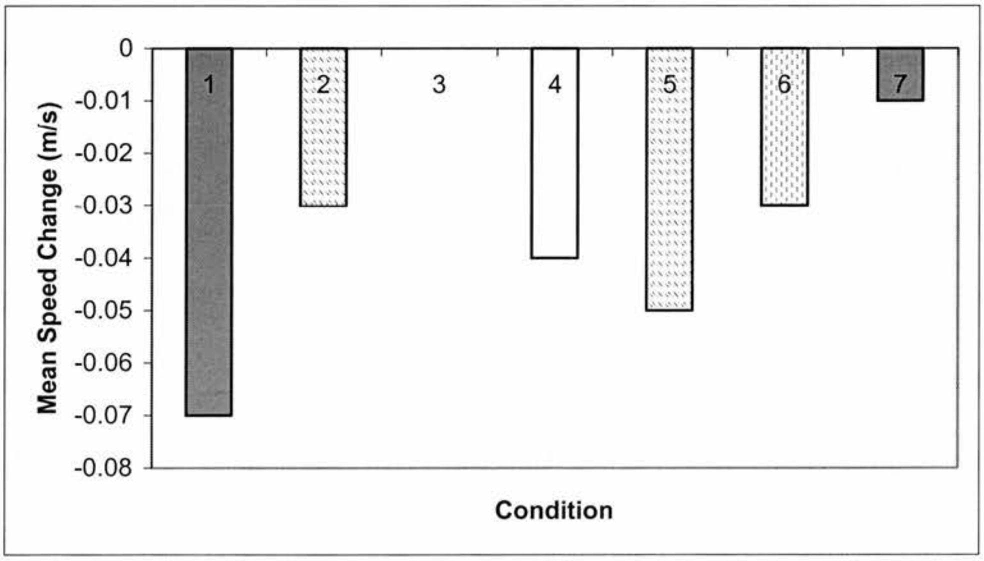


Figure 6.2: Mean Walking Speed Change scores for each experimental condition.
1 = both mood and control enhanced
2 = mood enhanced
3 = perceived control enhanced
4 = neutral
5 = mood reduced
6 = perceived control reduced
7 = both mood and perceived control reduced

ANOVAS were performed as manipulation checks, to detect whether there were changes in either PCS or mood score across experimental conditions.

With change in PCS as the dependent variable, a type IV 2-way ANOVA revealed neither main effects of mood and perceived control nor an interaction effect ($F(2, 104) = 1.14$, $p = .32$; $F(2, 104) = .36$, $p = .70$; $F(2,104) = .24$, $p = .79$). As shown in Table 6.8 and Figure 6.3, whilst the non-significant change in PCS was in the expected direction for the enhance perceived control group, perceived control also increased in the reduce perceived control and neutral groups.

Condition	**PCS change mean**	**PCS change SD**	**N**
Enhance perceived control	7.25	27.45	31
Neutral	3.01	25.25	48
Reduce perceived control	4.73	19.19	32

Table 6.8: Mean and SD of changes in PCS according to whether participants received enhancing, neutral or reducing perceived control manipulations.

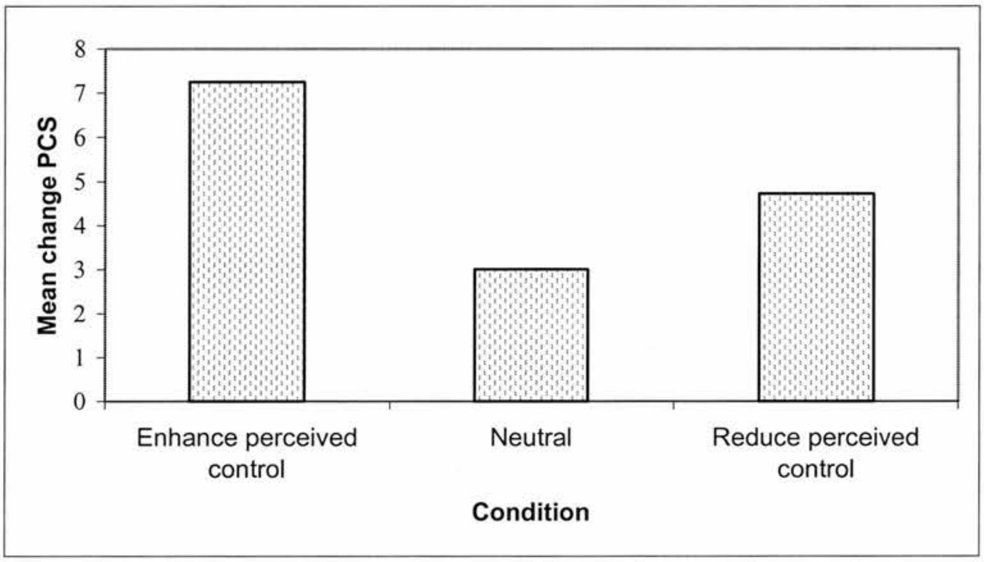


Figure 6.3: Mean changes in PCS according to whether participants received enhancing, neutral or reducing perceived control manipulations.

The final type IV 2-way ANOVA used change in Mood Score as the dependent variable. A main effect of mood manipulation condition was seen ($F(2, 105) = 4.42$, $p = .01$); as shown in Table 6.9 and Figure 6.4, the mood scores improved for participants in the enhance mood condition and worsened for those in the reduce mood condition. The post-hoc test Dunnett T3 (used because of unequal variance, see SD in Table 6.7) showed that the difference in Mood Score Change between enhancing and reducing mood conditions was significant ($p = .02$), with those in the enhancing mood condition reporting a more positive change in mood than the reducing mood group. The changes between enhancing and neutral and neutral and reducing mood conditions were not significant ($p = .26$ and $p = .31$ respectively).

Condition	Mood score change mean	Mood score change SD	N
Enhance mood	16.48	45.84	32
Neutral	-.41	40.18	48
Reduce mood	-15.73	43.50	32

Table 6.9: Mean and SD of changes in Mood Score according to whether participants received manipulations to enhance, reduce or not affect mood.

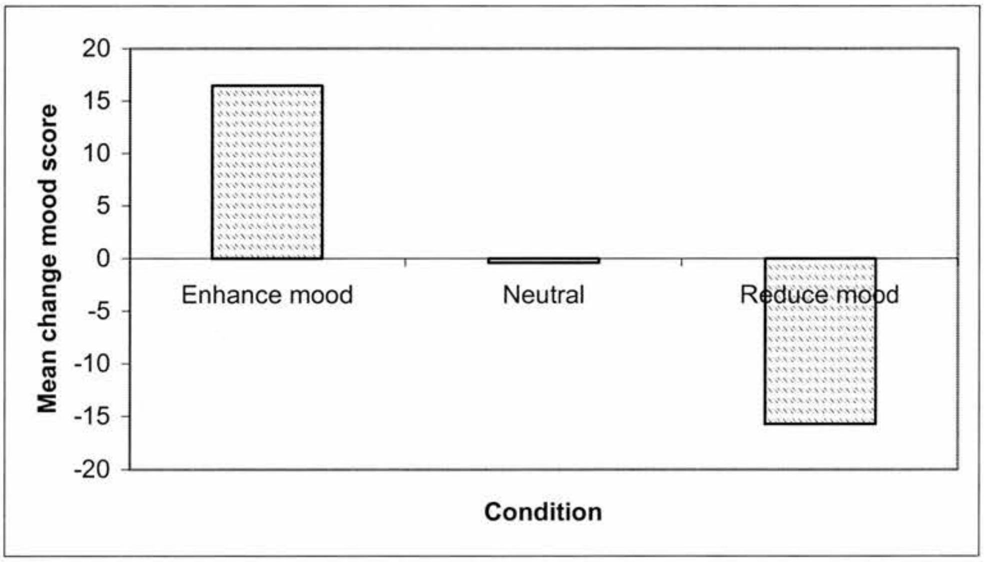


Figure 6.4: Mean changes in Mood Score according to whether participants received manipulations to enhance, reduce or not affect mood.

A main effect of perceived control manipulation on mood was not found ($F(2, 105) = .19$, $p = .82$). The interaction of mood and perceived control condition was non-significant, but at trend level ($F(2,105) = 2.65$; $p = .08$). On removing the outlier, the interaction reached significance ($F(2,104) = 3.46$, $p = .04$). Figure 6.5 would suggest that the combination of perceived control and mood manipulations together had less of an effect on Mood Score change than did mood inductions alone – an unexpected effect.

These results would suggest that the mood manipulation was effective in altering self-reported mood but the changes in mood were not mirrored by changes in walking speed. The perceived control manipulation did not appear to be effective either in altering self-reported perceived control or in altering walking speed. As main effects, mood manipulations do not appear to have affected perceived control levels nor vice versa.

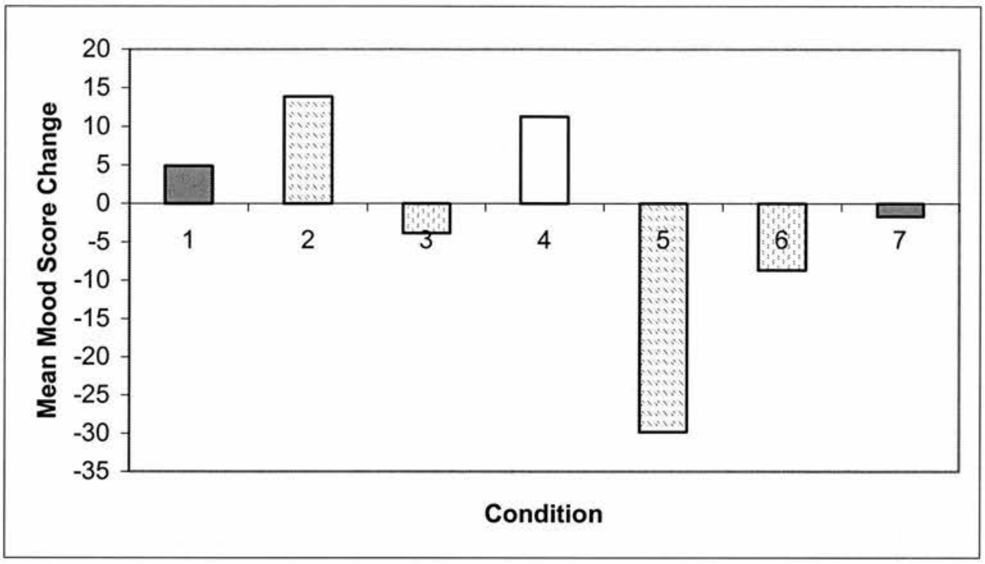


Figure 6.5: Mean Mood Score Change scores for each experimental condition (outlier excluded).
1 = both mood and control enhanced
2 = mood enhanced
3 = perceived control enhanced
4 = neutral
5 = mood reduced
6 = perceived control reduced
7 = both mood and perceived control reduced

Pearson's correlations were performed to test whether changes in walking speed occurred with changes in Mood or PCS, across all conditions.
Change in walking speed was not significantly correlated either with change in PCS or with change in Mood Score ($r = -.09$, $p = .37$; $r = -.14$, $p = .179$ respectively).

Change in PCS and change in Mood Score were found to correlate ($r = .29$, $p = .005$). However, on removal of the Mood Score Change outlier, Participant 85, the correlation was no longer significant ($r = .10$, $p = .30$). Figure 6.6 illustrates the effect Participant 85 (P85) had on the association.

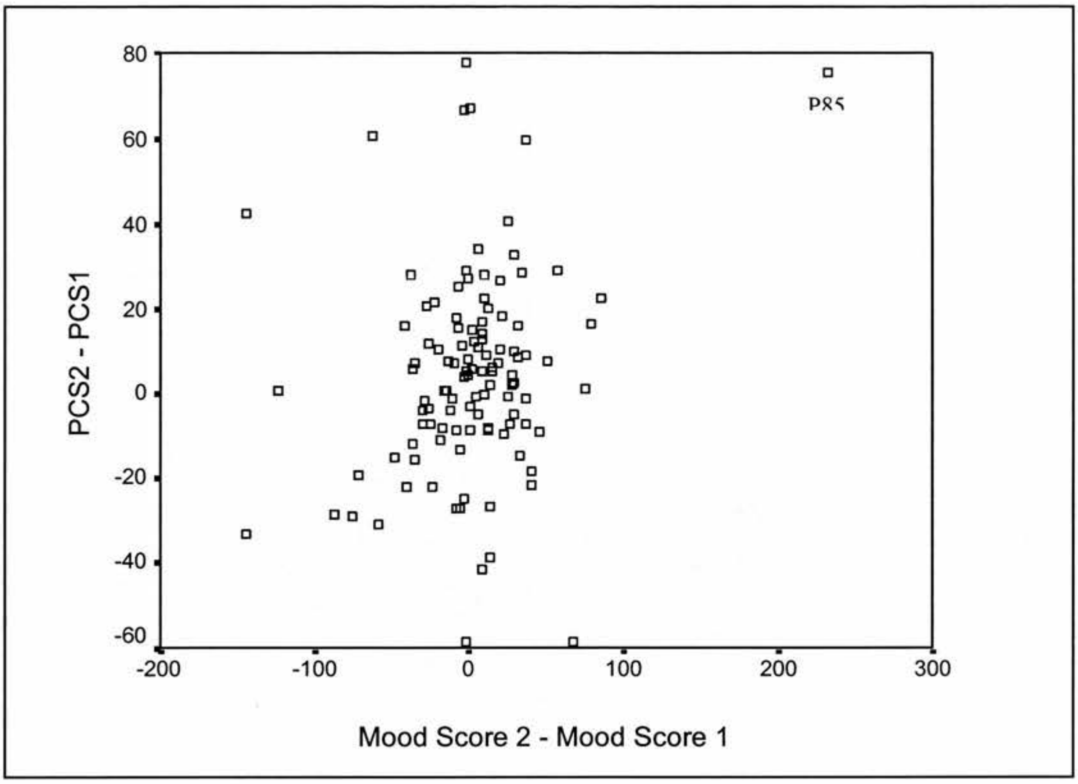


Figure 6.6: Scattergram showing change in PCS plotted against change in Mood Score.

Although change in PCS and Mood Score do not appear to be reliably correlated, cross-sectional PCS and Mood Scores were found to correlate, both pre- and post-manipulation ($r = .32$, $p = .00$; $r = .36$, $p = .00$) (see Figures 6.7a and 6.7b).

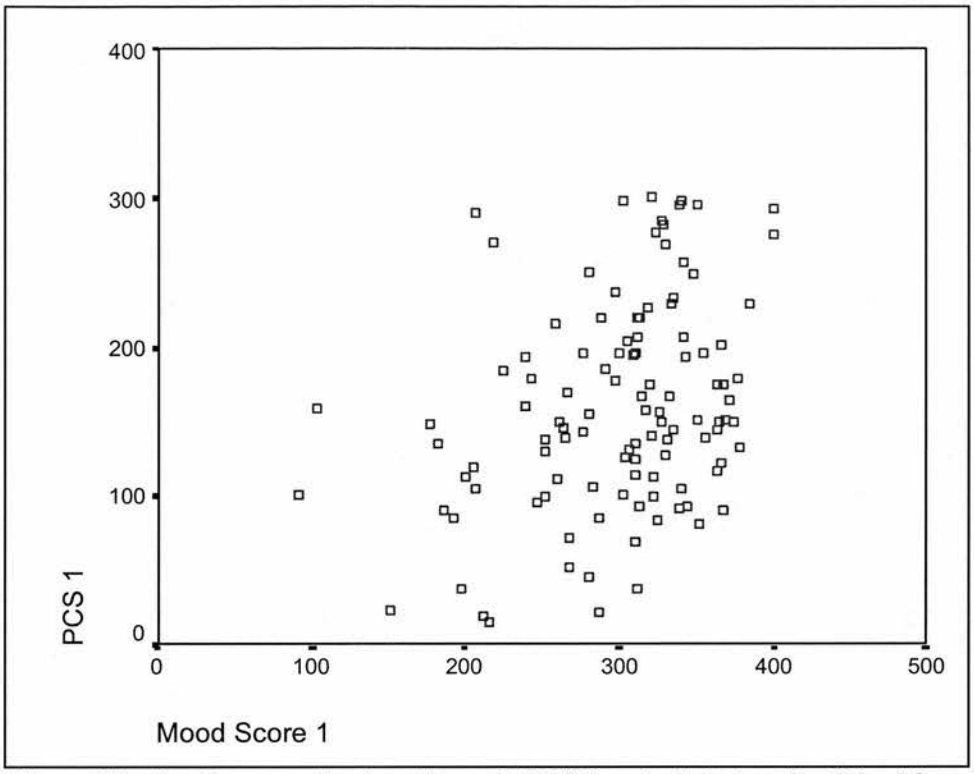


Figure 6.7a: Scattergram showing change in PCS time 1 plotted against Mood Score time 1.

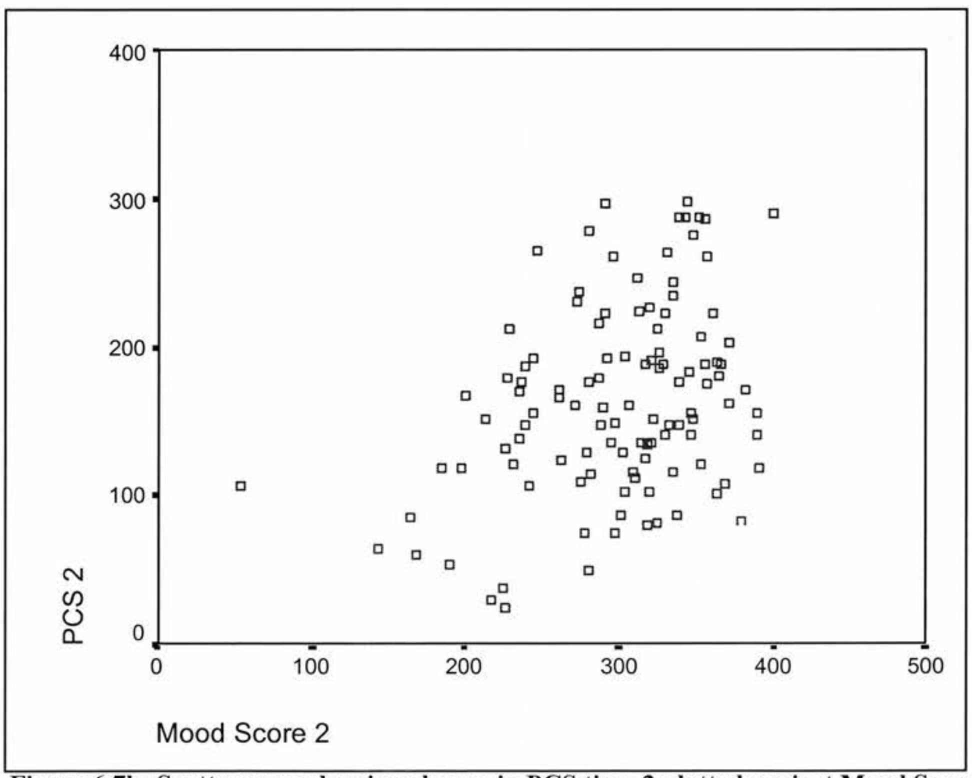


Figure 6.7b: Scattergram showing change in PCS time 2 plotted against Mood Score time 2.

Research Question 2i: To what extent do self-report, baseline observed and unaware observed assessments of walking speed correspond?

All further analyses involve the comparison of 3 walking speed measures. Therefore, only the data of participants with complete data for each of the three measure types (self-report, Baseline Observed and Unaware Observed) has been used in the analyses below (N = 98).

The mean and standard deviation walking speed using each of the three measure types are shown in Table 6.10. Table 6.11 shows that the measure types correlate highly significantly, but with rho ranging from .38 to .45, there is much unaccounted variance.

Measure		Mean (m/s)	SD	N
Self-report	Total	1.38	.29	98
	Female	1.33	.31	69
	Male	1.48	.22	29
Baseline Observed	Total	1.44	.16	98
	Female	1.43	.15	69
	Male	1.47	.17	29
Unaware Observed	Total	1.41	.16	98
	Female	1.38	.14	69
	Male	1.47	.18	29

Table 6.10: Mean and standard deviation of Self-reported, Baseline Observed and Unaware Observed walking speed.

	Baseline Observed	Unaware Observed
Self-report	.45**	.38**
Baseline Observed		.42**

Table 6.11: Correlations (Spearman's rho) between the three walking speed measures. (** = $p < .01$).

As seen in Figure 6.8, the differences between mean walking speed as assessed with each measure type appear small. However, using Wilcoxon Signed Ranks test, Self-Report and Baseline Observed walking speed are significantly different ($Z = -.25$; $p = .03$). The difference between Self-Report and Unaware Observed measures is not significant ($Z = -.87$, $p = .39$) and the difference between Baseline Observed and Unaware Observed measures reaches trend level significance ($Z = -1.82$, $p = .07$) (Baseline Observed tending to be a faster score than Unaware Observed). It would appear, therefore, that participants demonstrate a quicker walking speed than they report when they know they are being observed but this difference is not significant when they are not aware that their walking speed is being recorded.

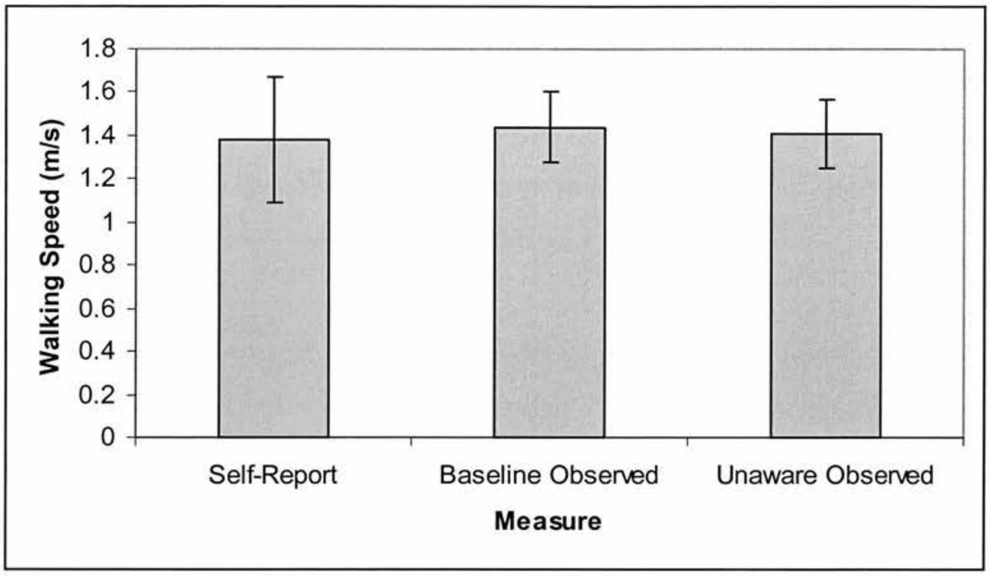


Figure 6.8: Mean walking speed (± SD) on self-report, Baseline Observed and Unaware Observed measures.

Research Question 2ii: Are measures affected by biases of negative affectivity and gender?

Negative Affectivity

If the negative affectivity bias hypothesis were true, it would be expected that the correlation between Self-reported speed and Negative Affectivity would be stronger than between Negative Affectivity and either of the observed measures. However, that does not appear to be the case: as Table 6.12 shows, moderate and significant correlations were seen between Negative Affectivity and Baseline Observed and Unaware Observed walking speed (rho = -.29, p = .00; rho = -.27, p =.01), with higher negative affectivity being associated with slower speed, whereas the correlation between Negative Affectivity and Self-reported walking speed was small and non-significant (rho = .08, p = .41).

Measure	Correlation with NA (rho) (N = 97)
Self-report walking speed	.08 ns
Baseline Observed walking speed	-.29**
Unaware Observed walking speed	-.27**

Table 6.12: Correlations (Spearman's rho) between distress (HADS) and activity limitations as assessed by self (patient), proxy and observer (OAD).
Shaded area indicates the correlations that would be expected to be the highest, if the NA bias hypothesis were true. (ns = non-significant, * = $p < .05$, ** = $p < .01$).

Gender

Mean and standard deviations of walking speed measures according to gender are seen in Table 6.10 and Figure 6.9.

A 3x2 factorial ANOVA was performed comparing the within-participant variable walking speed as assessed by Self-report, Baseline Observed and Unaware Observed measures by gender. Mauchly's Test of Sphericity was significant; the Greenhouse-Geisser correction was therefore used to correct for non-sphericity. A significant between-subjects effect was seen, with male participants walking more quickly than female ($F(1,96) = 7.61$, $p = .01$). However, the within-subjects effects of measure type and interaction between measure and gender were not significant ($F(1.53,146.67) = 1.38$, $p = .25$; $F(1.53,146.67) = 1.97$, $p = .15$). The gender bias hypothesis was not, therefore, supported.

Figure 6.9: Mean walking speeds of female and male participants, as assessed with each measure type.

Research Question 2iii. Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

The correlations between control cognitions (PCS, GSE), Anxiety and Depression and each measure of walking speed are shown in Table 6.13. None were significant.

	Anxiety		Depression		PCS time 1		GSE	
	rho	p	rho	p	rho	p	rho	p
Self-Report	**-.02**	.89	**.11**	.29	**.10**	.32	**-.13**	.22
Baseline Observed	**.01**	.91	**.18**	.86	**.14**	.17	**.02**	.87
Unaware Observed	**-.03**	76	**.08**	.46	**.12**	.25	**.03**	.75
N	98		98		97		97	

Table 6.13: Correlations (Spearman's rho) of Anxiety, Depression, PCS 1 and GSE with walking speed measures.

Summary of Results

Research Question1: Does manipulating mood and perceived control affect activity levels (walking speed) when participants are unaware that they are being observed?

Neither perceived control nor mood manipulations were found to affect walking speed. Manipulations were also not found to have an effect on Perceived Control manipulation checks but a main effect of the mood manipulation on Mood was seen, with the increasing mood group reporting a more positive Mood change than the decreasing mood group. In correlations, neither change in Mood nor Perceived Control was correlated with change in walking speed.

Participants' change in Mood was not correlated with change in Perceived Control on the removal of an outlier, although cross-sectional correlations between mood and perceived control before and after the manipulation were significant.

Research Question 2.i. To what extent do self-report, aware observed ('Baseline Observed') and 'Unaware Observed' assessments of walking speed correspond?

The three measures were significantly but not highly correlated (self-report – Unaware Observed: rho = .38, Self-report - Baseline Observed: rho = .45, Unaware-Baseline Observed: rho = .42, all $p < .01$). Baseline Observed walking speed was significantly quicker than self-report speed and almost significantly quicker than Unaware - Observed scores. The difference between self-report and Unaware Observed measures was not significant. It appears that participants demonstrate a quicker walking speed than they report when they know they are being observed but this difference is not significant when they are not aware that their walking speed is being recorded.

Research Question 2.ii. Are measures affected by biases of negative affectivity and gender?

No support was found for the negative affectivity hypothesis. An interaction between gender and measure type was not found, suggesting that gender did not bias measurement methods.

Research Question 2. iii. Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

No significant associations were found between control cognitions or emotion and any of the walking speed measures.

Discussion

Research Question 1: Does manipulating mood and perceived control affect activity levels (walking speed) when participants are unaware that they are being observed?

The only manipulation found to be effective in any way was that of mood. Mood effects did not seem to impact on walking speed, however.

It is possible that manipulation effects did not last long enough to have an effect on walking speed level. Clark (1983) found the effects of mood induction procedures to be short-lived. In the present study, the researcher attempted to end the procedure and send participants out as quickly as possible but this was at times difficult as participants believed the experiment to be over and some had questions to ask about the study. This could explain why the mood manipulation was successful according to manipulation check measures but had no effect on the later measure of walking speed.

Second, the task performed by this healthy adult sample – walking down a corridor – may have been too easy to be affected by mood. Fisher and Johnston (1996a) found manipulations of emotional distress to affect back pain patients' performance on a lifting task; the combination of health condition and task type are likely to have made the task challenging. If mood influences performance by affecting for how long and with how much effort people persevere on difficult tasks, it may be necessary for participants to find a task challenging for an effect to be seen. 'Normal' walking would not have come close to pushing healthy adults to physical activity limits.

Finally, it could be that mood alone is not enough to influence activity. Fisher and Johnston (1996a) manipulated the mood state of anxiety by asking participants to talk about either upsetting or good events. As well as bringing about positive or negative moods, this manipulation is also likely to have lead to participants having positive or negative thoughts. It could be that such thoughts, some of which could be control cognitions, were responsible for affecting activity performance, rather than mood per se.

As the perceived control manipulation had no effect on the perceived control manipulation check, it seems likely that the manipulation was ineffective in altering participants' perceived control levels. Further work needs to be done to find a perceived control manipulation that would be effective with this population in order to repeat the study. It is particularly odd that the perceived control manipulation appears to have lessened the effects of the mood induction on Mood Score when researchers such as Smarr et al. (1997) would have predicted the opposite effect. As seen in the introduction, a large body of literature supports the effects of perceived control on activity and so it would seem possible that a more effective manipulation method could affect walking speed.

However, it is possible that a perceived control manipulation could fail to affect walking speed in this population even if it were effective. Bargh et al. (1996) found priming an elderly stereotype to influence participants' walking speed, suggesting that cognitions can automatically influence physical performance. However, the cognitions primed by Bargh et al. (1996) are very different to those this experiment attempted to promote. Walking slowly seems to be part of the elderly stereotype that was primed. The present study attempted to affect control over physical activity, a concept that may not include walking, especially walking down a university corridor, in a healthy adult population. Bandura (1986) recommends that the level of behaviour should match the level of measure; from this it could be extrapolated that the level of the manipulation should also match. However, Fisher and Johnston (1996b) did find a general level perceived control manipulation (telling the researchers about times when they felt in control or out of control of situations) to affect performance on a specific lifting task.

In contrast with Bargh et al.'s (1996) subconscious priming of constructs, Ajzen (1991) proposed that behavioural intention, (a construct through which perceived behavioural control was to act, as well as acting on behaviour directly) would only have an impact on volitional behaviours. Whilst the overall act of walking down the corridor would be volitional, participants would not have consciously been thinking about the movement processes that make up walking, ie those behavioural actions that determine walking speed.

Bandura (1986) suggests that the stronger people's self-efficacy, the more effort they put into accomplishing a task. Walking down a corridor is unlikely to be perceived as challenging by a healthy student population and so self-efficacy could have limited input into effecting variance.

Research Question 2.i. To what extent do self-report, aware observed ('Baseline Observed') and Unaware Observed assessments of walking speed correspond?

The three measure types were only found to moderately correlate. That correlations are not higher is particularly interesting in the case of Unaware Observed and Baseline Observed measures (rho = .42, $p < .01$). Here two measures assess an almost identical behaviour, the difference being that participants are observed on one measure (whilst counting paces) and not on the other. A main effect of measure was not found to be significant in the gender ANOVA (see Results section; Research Question 2ii), but it is unfortunate that no measure of social desirability was included in the questionnaire to test whether stronger associations existed between social desirability and Baseline Observed walking speed than with Unaware Observed walking speed. Alternatively, it could be that the placement of the measures within context affected the results. The Baseline Observed was controlled not only in terms of behaviour observed, but, very likely, also with respect to cognitive processes. All participants were counting their steps as they walked and all would be returning to the laboratory to continue the experiment. In contrast, at the Unaware Observed measure, no cognitive task had been given and participants were continuing with their own lives; some may have been hurrying to lectures or meetings, others dawdling whilst thinking about where to go and what to do next. Finally, it could be that the measures themselves were not reliable. Test-re-test reliabilities for neither measure had been tested.

Baseline Observed walking speed was significantly faster than self-reported usual walking speed, and almost significantly faster than Unaware-Observed speed. The difference between Self-Report and Unaware Observed measures was not significant. It is possible that this reflects demand effects. Participants walked more quickly when observed by a researcher than they reported walking normally or than when they were not aware they were observed. The pressure of having someone observe walking could have speeded up walking pace. However, it would be expected that such demand effects would be associated with a social desirability effect, in which case participants would also be expected to self-report quicker walking speeds. Alternatively, it could simply be that counting steps whilst walking (as in Baseline Observed) leads to a quicker walking pace as participants are thinking about their paces and perhaps walking with more of a rhythm than in more ordinary walking, such as in the Unaware Observed measure.

Research Question 2.ii. Are measures affected by biases of negative affectivity and gender?

Consistent with the other studies in this thesis, no support was found for the negative affectivity bias hypothesis. More surprisingly, self-reported walking speed was the only measure to *not* correlate with negative affectivity. It could be that the self-report measure was not sufficiently sensitive for this purpose: participants chose one of four walking speeds, compared with the continuous measures of Baseline Observed and Unaware Observed walking speed. Nevertheless, in the Stroke Study, patients' self-reported activity limitations were not found to significantly correlate with negative affectivity assessed with HADS at the first interview so the present finding is not completely anomalous.

Evidence was not found to suggest gender biases measurements. The absence of an effect here is also consistent with the findings from earlier chapters.

Research Question 2. iii. Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

No significant associations were found between emotion or control cognitions and any of the walking speed measures. It could be that 'normal' walking speed is too familiar and automatic to be affected by emotional state in a healthy adult population. Similarly, the lack of an effect of control could be explained by the behaviour, 'normal' walking speed, being insufficiently volitional and requiring too little effort in this population, as discussed above in relation to the lack of effect of manipulations on walking speed.

Limitations

A major limitation in this study was the failure of the perceived control manipulation to affect levels of perceived control. The manipulation was modelled on that used by Fisher and Johnston (1996b) (telling the researchers about times when they felt in control or out of control of situations) which was found, on manipulation checks, to have the intended effect on reported perceived control. The manipulation was converted to a written task, and the wording was changed to make it more orientated towards physical activity. However, whilst the manipulation questions were piloted to make sure that participants would be able to understand the items and give the appropriate response, pilot tests were not performed to test whether the manipulations would actually alter perceived control. It is possible that the items simply did not sufficiently prime the concept of control. Participants were asked: 'Can you give examples of times when you performed well (poorly) physically and felt you were (not) in control of your performance?' Participants were first asked for examples of times they performed well/poorly, and only second was the concept of being in control introduced. For further studies, it would seem advisable to thoroughly pilot the perceived control manipulations and to focus manipulations on control.

The main research question, whether perceived control and mood manipulations would affect change in walking speed, was analysed using a factorial ANOVA.

However, a more efficient way of analysing the hypotheses would have been to use planned comparisons according to the four sub-hypotheses:

a. Enhancing/reducing mood will affect walking speed.
b. Enhancing/reducing perceived control will affect walking speed.
c. Enhancing both mood and perceived control will have greater effect than each individually.
d. Reducing both mood and perceived control will have greater effect than each individually.

Nevertheless, given the poor significance levels of the ANOVA performed (main effect of mood: $p = .92$, main effect of perceived control: $p = .72$, interaction effect: $p = .61$), it seems unlikely that significant effects would have been found.

Conclusion

It was not possible, from this experiment, to reach firm conclusions about the mechanism by which perceived control and emotion affect activity as the perceived control manipulation was ineffective. Future research should focus on developing effective perceived control manipulations. No evidence was found to support biases of either negative affectivity or gender. Predictive effects of control cognitions and emotion (anxiety and depression) were not found on any of the activity measures.

Acknowledgements

I am grateful to Professor Gerod Parrott, Georgetown University, Washington DC for sending me the mood induction music casettes used by Parrott and Sabini (1990) and to Lesley Noble for designing and setting up the photoelectric detector transmitter / receiver system.

CHAPTER 7
GENERAL DISCUSSION

Research Question 1: To what extent do the different measure types correspond?

Activity scores were compared across measure types in the Pilot, Walking Difficulties, Stroke and Experimental Studies. Correlations between measurement types ranged from low to moderate. In the Pilot Study, only the correlation between continuous monitoring and self-reported 24-hour activity neared significance. The Walking Difficulties Study reports three correlations being at or near significance: self-report FLP and shuttle-walk performance, self-reported distance and observed distance on the Shuttle Walking Test and the correlation between self-reported activity over the past month and monitor-recorded Step Number neared significance. The highest inter-measure correlations were seen in the Stroke Study. At time 3, the correlation between self- and proxy-reports reached .75. Correlations between self/proxy-reports and observer-assessed mobility were lower, with the correlation between self-report and observed activity being stronger than that between proxy and observer-reported activity at time 3. Finally, correlations between self-reported, Baseline Observed and Unaware Observed walking speed for the Experimental Study's healthy adult sample were all significant, but not high, ranging from rho = .38 to rho = .45.

The small sample sizes of the Pilot and Walking Difficulties Studies limit the certainty with which conclusions can be reached, but it would seem that, overall, the closer the activity assessed, the stronger the association between measures. In the Pilot Study, the strongest association was between self-reported and monitored activity over the same time period. In the Walking Difficulties Study, a self-report measure, which is likely to have reflected what participants were *able* to do, correlated with a performance task also measuring what they were *able* to do. Even though the time span differed, self reported activity over the past month (activity participants remembered *actually* doing) correlated with activity detected by the monitor being *actually* performed. Self-reported and observed activity on the Shuttle Walking Test – an identical activity – correlated. In the Stroke Study, the strongest correlations were seen between self- and proxy-reports: assessments on identical measures made either by the patient her/himself or by someone who spent much time with the patient. The association between self-report and observer-report was stronger than between proxy-report and observer-report: a direct report from a participant was more strongly associated with observed performance than a report from a third party.

These findings are consistent with the literature, which also seems to suggest that, the closer the activities assessed across measures, the stronger the associations between measures. For example, Myers et al. (1993) found that the more closely self-report and performance items matched, the closer the relationship between scores. However, given that the present studies aimed to test the extent to which measures were associated whilst keeping the activities assessed as comparable as possible over measurement types, further discussion as to why more and stronger associations were not found is necessary.

In the Pilot Study, only one inter-measure correlation neared significance. The absence of further associations could, in this study, be due at least in part to the small sample size; sample sizes for these analyses ranged from n = 9 to n = 14. The Walking Difficulties Study was also limited by its small sample size. However, it is notable that, whilst the correlation of monitor-reported 24-hour activity with self-reported activity over the past month neared significance, the association of 24-hour monitor reports with self-reported activity over the same 24-hour period did not. The hypothesis that the more similar the activities assessed are, the stronger the association between measures should be would have predicted the latter analyses to have given the more significant results. It is not clear why this was not found; replication with a larger sample size may help to clarify the results. Voorrips et al. (1991) argued that the reference time frame of activity measures should not be too long with older populations as short-term memories may be weaker; further work could take this hypothesis on board. All three measures of walking speed used in the Experimental Study were found to significantly correlate, but not at high levels. One suggestion is that social desirability may have differentially affected measures; for example, participants with high levels of social desirability may have given higher scores on self-report and Baseline Observed measures than on Unaware Observed walking speed. Unfortunately, a measure of social desirability was not taken. Additionally, although all three measures assessed a very specific activity: walking speed, they differed in context. In self-report, participants reported on their usual walking speed. The Baseline Observed was taken in a formal experiment context, whilst participants were counting their steps. As participants believed themselves to be leaving the experiment and going on to continue with their own business, the Unaware Observed measure was taken. Thus, even though the activity assessed was superficially identical, the contextual differences may have been significant.

Perhaps of most importance is the highest correlation found between any two measures: the time 3 correlation between self- and proxy-report (Stroke Study), $r = .75$. Here, activity on an identical measure was assessed by people very well acquainted with the patients' activity: the patients themselves and the patients' caregivers. Yet over 40% of the variance was still not accounted for.

Where possible (where two measures assessed the same activity, in the same units), measures scores were compared directly. This led to some surprising results. It was found in both the Pilot and Walking Difficulties Studies that participants systematically under-reported distance walked and over-reported time spent lying (Pilot Study), sitting, standing and walking (Walking Difficulties Study). In contrast, Klesges et al. (1985) found student participants to under-estimate sitting and moderate activity whilst over-estimating time being more active; a finding supported by Klesges et al. (1990). As discussed in the Walking Difficulties Study, it is possible that different instructions may have affected participant report. Klesges et al.'s participants were able to choose their activities whereas participants in the present studies were given no choice. Klesges et al.'s participants may have been vulnerable to demand effects, feeling that they should have performed a high level of activity, leaving less of the time allocation (total time = 1 hour) for lower intensity activities, whereas the present studies' participants were not given the opportunity to perform such high level activity. However, this still does not explain why participants consistently over-estimated time. Glicksohn (2001) reported perceived duration to be related to absorption such that the more absorbed people are, the shorter a time interval seems. It could be that the times spent performing lab tasks seemed longer than they actually were because the activities were not engaging – participants were asked to perform straight-forward physical activities which may have been quite boring. This hypothesis could be tested in further studies reproducing the basic methodology of these studies but giving some participants cognitive tasks to perform at the same time as the physical activities. It would be predictive that the 'cognitively enhanced' tasks would be perceived to be of shorter duration than the standard tasks.

It is reported in the Experimental Study that Baseline Observed walking speed was significantly faster than self-reported usual walking speed, and almost significantly faster than Unaware Observed speed. As discussed in the Experimental Study, this could have been caused by demand effects or as an artifact of the Baseline Observed walking task. In view of the findings reported in the Pilot and Walking Difficulties Studies, an alternative explanation would be that participants were under-reporting their usual walking speed. Participants in the earlier chapters under-reported distance and over-reported time. According to the equation 'speed = distance/time', such errors would predict that participants would also underestimate walking speed.

The literature consistently reports proxies rating patients as more limited than patients rate themselves (Introductory Chapter, Table 1.1). However, as reported in the Stroke Study, a trend towards stroke patients reporting greater limitations than proxies was seen. Proxies appeared to perceive patients' activity to be improving over time whereas patients did not. Exploratory analyses suggested it to be proxies who initially reported greater activity limitations than the patients who perceived this improvement. Epstein et al. (1989) found the time proxies spent helping patients to predict discrepancies in reported functional status such that the more time a proxy spent helping a patient, the more impaired the proxy rated the patient, relative to the patient's own report. It could be that proxies' ratings of patients' activity limitations improved with patients requiring less help over time.

However, it is also reported in the Stroke Study that, the less severe the limitation, the greater the discrepancy between patient and proxy rating. Sneeuw, Aaronson, deHaan et al. (1997) also found discrepancies greater where patients were more limited whereas Sneeuw et al. (1998) found poorer agreement at intermediate functioning levels. It is not clear how these samples compare in terms of the overall impairment of participants. It could be that Sneeuw et al.'s (1998) 'intermediate' functioning patients are at an equivalent level to Sneeuw, Aaronson, deHaan et al.'s (1997) and the Stroke Study's patients, such that agreement is high for participants whose limitations are either very high or very low, but concordance is lower for more ambiguous, 'intermediate' ability levels.

In summary, it would seem, therefore, that the closer the activities assessed are, the closer the correlation between measure types is. However, even when the associations between measures are at their highest, much variance remains unaccounted. Not only is the association between measures low, but biases have also been found on some measures, for example the finding that participants consistently under-reported distance but over-reported time spent sitting and standing. It would appear, therefore, that the different measure types are accessing different aspects of 'activity'. Whether biases of negative affectivity, social desirability or gender contribute to this effect is discussed in the next section.

Research Question 2: Are measures affected by biases of negative affectivity, social desirability and gender?

The biasing effects of negative affectivity and gender were investigated in the Pilot, Walking Difficulties, Stroke and Experimental Studies; the effects of social desirability were examined in the Pilot and Walking Difficulties Studies only.

Negative Affectivity

The negative affectivity hypothesis would predict negative affectivity to be associated with self-report activity measures such that people high in negative affectivity would report less activity. More 'objective', i.e. non self-report measures would be expected to not correlate, or to correlate less strongly with negative affectivity than would self-report measures. This hypothesis was investigated in all four studies (Pilot, Walking Difficulties, Stroke and Experimental) but was not supported.

It is reported in the Pilot and Walking Difficulties studies that associations (significant at least at trend level) between self-reported activity and Negative Affectivity were very similar to the associations between observed performance measures and Negative Affectivity, providing no evidence to support the hypothesis that Negative Affectivity's relationship with self-report activity would be the stronger. However, the correlations between Negative Affectivity and continuous monitoring of activity were not significant. This raises the question of whether negative affectivity was not only biasing self-report but observed performance measures also. It could be that, of two people who ordinarily do similar amounts of physical activity, the person higher in negative affectivity would feel less motivated to perform well in an observed performance test. Nevertheless, the Stroke Study provides further evidence against the negative affectivity hypothesis. Here, self-report, observer-rated and proxy-report measures of activity were compared. Negative affectivity could bias self-report and observer-rated measures, but it is not clear how a patient's negative affectivity level would bias a caregiver's perception of their physical activity, unless the caregiver's own negative affectivity was associated with that of the patient. However, whilst patient Negative Affectivity was associated with proxy-reported activity at all 3 time points, patient Negative Affectivity was significantly associated with self(patient)-report at just 2 time points and proxy Negative Affectivity was associated with proxy-reported activity at only one time point. As in the earlier chapters, observer-rated activity was associated with Negative Affectivity as strongly as was self-report activity. Further evidence is provided in the Experimental Study. Negative affectivity might be expected to bias self-reported walking speed but is unlikely to have biased the Baseline Observed measure where participants were asked to walk at their normal walking pace and even less likely to have biased the Unaware Observed walking speed measure where there was no notion of its being a test situation. However, Negative Affectivity correlated significantly with Baseline Observed and Unaware Observed walking speed but did not correlate with self-reported walking speed.

Hence it would appear that the association between negative affectivity and activity is real, and not an artifactual bias found only on 'objective' activity measures. These findings contradict Watson and Pennebaker's (1989) reports of symptom perception biases. This lack of agreement could be due the differences in outcome variables: Watson and Pennebaker (1989) concentrated on the perception of bodily symptoms whereas the present studies assessed recall and assessment of activity. Thus, whilst negative affectivity could influence how bodily sensations are perceived, it may not influence how activity is recalled and reported. Nevertheless, theories of mood-congruent learning and recall (e.g. Teasdale and Fogarty, 1979; Bower et al., 1981) would also predict that mood on performing and recalling activity would affect self-reports. However, the stimuli recalled in such experiments were somewhat different to the activity-related behaviours recalled here. Some of the studies assessed recall of material presented in the laboratory; for example, Bower and Mayer (1989) gave participants word lists to memorise. Bower et al. (1981) asked participants to recall real childhood incidents but such events may be more difficult to recall and people may be more dependent on cues such as mood to bring them to mind than activity events occurring more recently. Bower (1981) did find recollection of emotional events occurring 1 week earlier to be affected by mood at recall, but it seems plausible that the association of mood with emotional events might be stronger than the association with non-emotional events such as going for a walk.

Mora, Robitaille, Leventhal, Swigar and Leventhal (2002) use the 'commonsense framework' to suggest that the vigilance associated with trait negative affectivity would increase motivation for developing expertise in identifying when they are ill. Therefore an individual high in negative affectivity, in particular, an elderly person with much experience, would have detailed knowledge of ongoing chronic symptoms. Negative affectivity would be expected to be associated to such 'everyday' symptoms but not to the reporting of acute conditions such as colds. In a sample of 790 (at baseline, 719 at 12-month follow-up) older adults, Mora et al. (2002) found 'Negative Affect' (operationalised with a scale of anxiety and depression items; these are 2 components of Watson and Pennebaker's (1989) definition of 'Negative Affectivity') to be related to perceived severity of the prior week's chronic symptoms at both time points but unrelated to the perceived severity or duration of acute symptoms. The authors suggest that the findings replicate the positive relationship of negative affectivity with symptoms but do not support the hypothesis, derived from the symptom perception hypothesis, that negative affectivity biases all health reports.

Mora et al's framework provides an explanation for the lack of a negative affectivity bias in the present studies. Someone high in negative affectivity may be highly vigilant for signs of illness and yet be no more motivated than someone low in negative affectivity to be vigilant in remembering how active they have been.

Social Desirability

It was hypothesised that social desirability could bias self-report activity measures, and may also influence observed performance activity measures as participants high in social desirability might be motivated to put more effort into task performance. This hypothesis was investigated in the Pilot and Walking Difficulties Studies. However, Social Desirability was not found to be associated with any activity measure in either study. The social desirability hypothesis was not, therefore, supported. However, further work with larger samples should be done before a social desirability effect on activity reporting can be ruled out. The two studies in which this hypothesis were both very small; n = 14 in the Pilot Study 2, n = 20 in the Walking Difficulties Study. Lack of power to detect effects or sampling error could have affected results. It is unfortunate that a social desirability was not available in the Stroke Study and that a social desirability measure was not included in the Experimental Study.

Gender

The gender bias hypothesis, supported by the findings of, for example, Merrill et al. (1997) and Hakala et al. (1994) suggested that women would under-report and men would over-report activity in relation to observed performance measures. This was investigated in all four study chapters (Pilot, Walking Difficulties, Stroke and Experimental), but no support for the gender bias hypothesis was found. The present findings are not inconsistent with the literature: a number of other studies also found no biasing effect of gender. For example, Ferrer et al. (1999) found disagreement on self-report and performance scores to not be influenced by gender and Davis (1981) did not find women to be more likely than men to report chronic joint problems. It may be that such a gender effect is only detected in some studies because of its small size. Even Merrill et al. (1997) found self-reported and performed ability to match for most items; only where differences on items were seen were more men than women under-reporting disability and more women than men over-reporting disability, as compared with observed performance. Further studies could further illuminate this issue. As gender is routinely recorded, any study collecting both self-report and performance data could easily also investigate the gender bias hypothesis.

Further work could also investigate whether gender has a role in the proxy-report bias discussed in the Introductory Chapter (see Table 1.1). Although a similar effect was not found in the Stroke Study, in many studies, proxies were reported to under-report activity compared with patients' self-report. Caregivers are often female whilst patients are often male (for example, in the Stroke Study, most (68/101) patients were male, and most of the caregivers were the patient's spouse). It seems plausible, therefore, that gender could be a factor in proxy-report bias, although *why* gender should bias activity reports should also be investigated (e.g. is gender associated with emotion or control cognitions?).

Research Question 3: Do the abilities of emotion and control cognitions to predict activity depend on the measurement method used?

Associations of emotion and control cognitions with activity levels were investigated in all four studies (Pilot, Walking Difficulties, Stroke and Experimental).

Emotion

In the Pilot Study, Depression neared significance with activity assessed only on the usual self-reported activity measure.

In the Walking Difficulties study, one significant association was seen with a self-report measure: Anxiety was associated with Self-reported Minutes Active. Both Anxiety and Depression neared significance in predicting the observed performance measure. The continuous monitoring measure (Step Number) was predicted by Depression at trend level.

The Stroke Study found that, cross-sectionally, Depression was associated with self-, proxy-reported (time 1) and observer-rated activity (time 2); Anxiety was associated with only proxy- and observer-rated activity, not self-reported activity. Additionally, the correlation of Depression with activity was significantly stronger for proxy- than self-reported activity. In contrast, predictively, self-reported Recovery measures were associated more consistently with emotion variables than were proxy-reported Recovery measures. Emotion variables predicted self-reported Recovery at both times 2 and 3 whereas proxy-reported Recovery was predicted only at time 2. Time 3 observer-determined Recovery was predicted by Depression. Unfortunately, the number of analyses involving observer-rated Recovery was limited by the measure's only being presented at times 2 and 3.

Emotion was not found to predict walking speed on either self-report, Baseline Observed or Unaware Observed measures in the Experimental Study. As discussed in the Experimental Study, it could be that the behaviour, normal walking speed, was too familiar, automatic and easy in a healthy adult population to be affected by emotional state.

It would appear that, overall, emotion reasonably consistently predicted activity assessed on both self-report and observer-rated activity. In the Pilot Study, only self-report activity was associated with emotion, but this was a small-sample exploratory pilot study and potential relationships may not have been detected. The Stroke Study's results suggest self-report Recovery measures to be predicted more consistently than proxy-report measures, but proxy-reported activity limitations were more strongly associated with Anxiety and Depression than were self-report measures cross-sectionally. Proxy-rated measures were only included in the Stroke Study; further work could compare the extent to which proxy-reported measures are predicted by emotion, compared with other measures, in different participant samples. Continuous monitored activity was almost significantly predicted by Depression in the Walking Difficulties Study; the ability of emotion to predict continuously monitored activity could be investigated further with more reliable and effective monitoring devices (see the Calibration Studies) in a larger study.

The effectiveness of emotion predicting self-reported activity is supported in the literature (e.g. Bruce et al., 1994; Johnston, Earll et al.,1999). However, the present results contradict the findings of Johnston, Morrison et al. (1999) and Bonetti et al. (2001) where emotion variables were found to predict self-report but not observed recovery. It is particularly interesting that Bonetti et al. (2001) did not find an association because the data reported in the Stroke Study was a sub-sample of that used by Bonetti et al. (2001). An important difference, however, is that only those observed items relating to ambulation were analysed in the Stroke Study whereas the full Observer Assessed Disability scale was used by Bonetti et al. (2001). Emotion could be associated with observed ambulatory movement but not general movements such as turning the head and clasping hands. Kaivanto et al. (1995) found depression did not significantly predict the performance of patients with chronic back pain but whether depression would predict their self-reported activity in comparison was not assessed.

Control Cognitions

The association of Generalised Self-Efficacy with activity assessed on the usual self-report activity measure was found to near significance in the Pilot Study.

In the Walking Difficulties study, a number of self-report activity measures were found to be predicted by control cognitions. However, apart from the association between FLP and Generalised Self-Efficacy, these associations were not in the expected direction. Further investigations suggested the relationship between control cognitions and self-reported activity was confounded by reporting accuracy. In contrast, observed performance measures were found to be consistently predicted by perceived control in the expected direction. Control cognitions were not found to predict continuous monitoring scores.

The Stroke Study reports that, cross-sectionally, control cognitions were associated with activity on all three measures. Significant differences in correlations for self- and proxy-rated activity were not found. However, longitudinally, control cognitions predicted Recovery only at trend level and only on self-reported Recovery at time 3.

Control cognitions were not found to predict walking speed on either self-report, Baseline Observed or Unaware Observed measures in the Experimental Study. As discussed in the Experimental Study, the outcome activity, normal walking speed, may have been too automatic and not challenging enough to be affected by control cognitions. Ajzen (1991)'s perceived behavioural control is predicted to act, through intention as well as directly, on volitional behaviours. If a 'usual' walking pace is an automatic behaviour, Ajzen would not predict it to be affected by perceived behavioural control. Bandura (1986) proposes that, the greater the self-efficacy, the more effort put into a task. Walking down a corridor is not likely to require great effort in a healthy adult population and so self-efficacy is unlikely to be predictive.

Control cognitions were found to be associated with self-report activity measures in the Pilot, Walking Difficulties and Stroke Studies. However, the association of control cognitions with reporting accuracy was found to confound the Walking Difficulties Study's results. The predictive effects of control cognitions on observed performance measures varied between studies. In the Pilot Study, control cognitions did not predict observed activity. However, it is difficult to draw conclusions from this study due to its small sample size and, additionally, only one self-report measure was found to be associated with control cognitions. In the Walking Difficulties Study, control cognitions consistently predicted observed performance. It is reported in the Stroke study, however, that, whilst control cognitions are associated with observed performance cross-sectionally, they are not found to predict observer-derived Recovery. A factor in the differences between the findings reported in the Walking Difficulties and Stroke Studies is the level of specificity of control cognitions and observed activity measures. In the Walking Difficulties Study, perceived control over walking test performance predicted actual walking test performance. In the Stroke Study, control perceptions over recovery in general (RLOC) and whether or not participants would be able to perform a mobility task in the next month (PCI) did not predict participants' ability to perform a few short standing and walking tasks. Hence the control cognitions and activities performed are different. This hypothesis is supported by the finding that, for the self-report measure, the PCI control cognitions, which assessed control over an activity identified in the self-report measure, predicted self-reported Recovery more strongly in multiple regression than did the more general RLOC. This explanation is consistent with Bandura's (1986) proposal that discrepant results may be found where the level of generality differs between independent (control) and dependent (activity) variables. Seeman et al. (1999) found self-efficacy assessed at a general level (self-efficacy with respect to instrumental activities such as arranging transport) to predict self-reported physical limitations but not observed performance assessed using balance, gait, foot taps, leg strength and manual dexterity. Nevertheless, level of specificity does not seem to be the full explanation because Generalised Self-Efficacy also predicted observed walking task performance in the Walking Difficulties Study.

The finding of control cognitions predicting activity on both self-report and observed performance measures is supported in the literature (Johnston, Morrison et al., 1999; Rejeski et al., 1996; Rejeski et al., 2001).

Control cognitions were not found to predict activity as assessed by continuous monitoring or proxy-report (significant in cross-sectional analyses only). For each measure, further research would be advisable before concluding that control cognitions do not predict activity as assessed with these methods. The two studies involving the Numact monitor were small-scale and the Numact monitor was found to be unreliable (Calibration Studies) which limited the use of data collected with the device. Proxy data was collected in a large study, but the measure was only included in that one study and replication could confirm the findings.

Research Question 4: Does manipulating mood and perceived control affect activity levels (walking speed) when participants are unaware they are being observed?

Mood and perceived control manipulations were performed, and their effects on walking speed assessed, in the Experimental Study.

Neither mood nor perceived control manipulations were found to affect walking speed. A main effect of the mood manipulation on Mood was seen but manipulations were not found to have an effect on Perceived Control. In correlations, neither change in Mood nor Perceived Control was correlated with change in walking speed.

These results would suggest that mood does not independently influence activity, and that the findings of Fisher and Johnston (1996a), where emotion manipulations did influence task performance, may have been caused by cognitive changes caused by the emotion induction. However, as discussed in the Experimental Study, there are alternative explanations. First, manipulation effects may not have lasted long enough to have affected walking speed. Second, the task – walking down a corridor – may have been too easy for this healthy adult sample to have been affected by mood.

Unfortunately, the study could not properly address whether the effects found by Fisher and Johnston (1996a) could have been caused by changes in perceived control (found to successfully alter task performance by Fisher and Johnston (1996b) because the perceived control manipulation did not successfully alter perceived control levels, according to manipulation checks.

Nevertheless, the Stroke Study's results suggested that the effects of emotion on activity are not dependent on changes in control cognitions. Hierarchical multiple regression equations found that emotion (Depression) contributed to Recovery when entered after control cognitions (PCI) but on entering control cognitions after emotion, additional variance was not explained. This would suggest that the lack of effect of the mood induction may have been caused by the induction wearing off or the dependent variable (walking speed) being inappropriate. Additionally, Fisher and Johnston (1998) found distress to mediate the relationship between pain and disability but control cognitions, assessed using the Multidimensional Health Locus of Control did not. The authors suggested that the measure of control cognitions chosen may have limited the findings as other studies had found stronger associations between control and pain-relate disability using self-efficacy measures of control rather than Locus of Control. Such results support the findings of Härkäpää et al. (1991), where distress but not health locus of control was associated with the poorer performance of low back pain exercises. However, the literature more consistently finds control cognitions (other than locus of control measures) to predict activity independently of emotion (Holm et al., 1998; Johnston, Morrison et al., 1999; Lackner & Carosella, 1999; Sullivan et al., 1998). This would indicate a need to once more compare manipulations of perceived control and emotion on activity. However, further research should use well piloted, effective perceived control manipulations and an activity outcome measure that is challenging for participants. This could mean either maintaining the 'Unaware

Observed' walking speed outcome measure but working with a sample of people with walking difficulties or staying with a healthy adult sample but using a more challenging activity outcome measure. The latter is likely to pose difficulties if participants are to be unaware that their activity is being observed.

Measurement

Overall, self-report activity measures were found to be easy to administer and acceptable to participants. They were extremely flexible and allowed access to a greater range of activity than other measures: participants could be asked about activity over the past month, performance in laboratory tasks, and activity during the same time periods for which monitors were worn. It was not found to be vulnerable to biases of negative affectivity, social desirability or gender. However, it seems that people find it difficult to accurately report some activities; in particular, the time spent performing activities in the lab tended to be over-reported. Self-reported activity did seem, in general, to be predicted by emotion and control cognitions although researchers should be aware of the potential for accuracy to confound the association of control cognitions with self-reported activity.

In general, observed performance measures of activity were straightforward to administer and acceptable to participants; even when participants were not told in advance that behaviour would be observed (Experimental Study), none expressed dissatisfaction with the procedure. Participants did sometimes seem to be uncertain as to what exactly was expected of them in the Shuttle Walking Test (Walking Difficulties Study), but this reflects the task itself rather than the domain of observed performance measures. Although observed performance measures are often carried out in laboratories, they are not restricted to them; the Stroke Study used an observed measure in participants' homes. Whilst this restricted the range of activities that could be observed (e.g. walking was limited to a couple of steps), the measure was associated with other activity measures and was found to be predicted by some psychological variables. More generally, observed performance measures were predicted by psychological variables in both the Walking Difficulties Study and the Stroke Study.

The continuous monitor, the Numact monitor, was, although less convenient to use than other measures, not problematic to administer and was acceptable to participants, although some did express pleasure at being able to go home and bathe/shower after the study! However, as investigated in the Calibration Studies, the monitor was not found to be reliable; a number of monitors broke down and inter-monitor reliability was poor. It was only predicted by one psychological variable: depression, at trend level, in the Walking Difficulties Study. Small sample size caused by technical problems and the low reliability of the monitor itself make it difficult to determine from this whether psychological variables do not tend to predict activity assessed by continuous monitoring, or whether the absence of significant relationships is due to error associated with this particular monitoring device. Further research using more reliable monitors is needed to address these issues.

As for self-report measures, the proxy-report scale was easy to administer and acceptable to participants. Correlating highly with patient self-reported activity, it seemed to be assessing the same construct to a fair extent. However, patterns of association with psychological variables did seem to differ between self- and proxy-report measures, with proxy-reported activity being more strongly associated with psychological variables cross-sectionally, but self-reported Recovery being more consistently predicted longitudinally. It would appear, therefore, that proxy measures are a useful assessment of activity but researchers should not assume that they are assessing exactly the same aspect of activity as self-report measures. Hence it is not appropriate to simply replace self-report measures with proxy-report where patients are unable to respond for themselves and expect the proxy-report measures to behave in an identical way to the self-report assessments. However, proxy-reported activity was only included in one of the four studies so replications of the study would be beneficial.

Potential biasing factors (negative affectivity, social desirability and gender) were found to be less problematic than anticipated and researchers need not be unduly concerned about their effects. However, in studies where these factors are crucial to the research question, it may still be advisable to monitor them. In particular, whilst the present studies have found an absence of effect in the contexts of laboratory and home, no assessments were performed in clinical settings.

In summary, self-reported, proxy-reported and observed performance measures of activity were found to be practical and predictable; more evidence is required for continuous monitoring. It should be kept in mind that, at the highest, activity measures had little more than half the variance in common and differences were seen in the extent to which different measure types were predicted. It appears, therefore, that measures are not assessing identical phenomena. As there is, as yet, no 'gold standard' activity measure, it is not possible to determine which is the most 'accurate'. Where possible, using a combination of measurement methods should be considered.

General Limitations

One limitation potentially affecting 3 studies reported here is the calculation of PPAQ score. The PPAQ measure of self-reported activity was used in the Pilot and Experimental Studies, with their method of calculating MET counts for climbing stairs being used in the Walking Difficulties Study also. However, Bassett, Vachon, Kirkland, Howley, Duncan and Johnson (1997), following tests of oxygen uptake in participants performing stair-climbing tasks, argue that the algorithm for calculating scores should reflect an energy cost of .2 kcal for going up and down one step, rather than Paffenbarger et al.'s (1978) .4 kcal. It could be, therefore, that energy expenditure estimates for self-report measures are over-estimated on this measure as a result.

An assumption made in this study is that the measures used accurately reflected their activity domains, for example, that activity measures reflected activity measures in general. Although efforts were made to ensure the best possible measures for each context were used, the measures available were not perfect (see discussions of measures in study chapters). There is also no really good, agreed measure with which to compare different types of measures: 'The major problem in the development and evaluation of physical activity assessment methods is the absence of a feasible and accepted gold standard against which estimates of physical activity can be compared' (Blair, 1992).

A major limitation that plagued the Walking Difficulties Study and may also have affected the Pilot Study was the Numact monitor's unreliability, outlined in full and evaluated in the Calibration Studies. The monitors seemed to be consistent mainly in their ability to break down. The Numact monitor was chosen for the study because, by combining accelerometers with movement sensors it gave the potential to give more information than either a simple pedometer or other devices using accelerometers only. When working, the Numact monitor gave a minute-by-minute readout indicating posture, steps taken and the energy with which the participant was moving about. Whilst a pedometer would only have given an estimate of steps taken or distance walked, the Numact monitor, by multiplying step number by step amplitude, gave a score reflecting intensity of activity. However, as the step amplitude measure was found to be particularly unreliable between monitors, only step number was used in between-participant analyses (in the Walking Difficulties Study). Hence, a cheaper, more convenient to wear, less prone to breaking down, pedometer device would probably have been more efficient and enabled a larger study to be performed. As it is, the Walking Difficulties Study was ended with fewer participants than originally intended because it was not clear how the monitor problem would best be resolved and the time available remaining would be better spent on other studies. As a result, too little data was gathered in the Walking Difficulties Study for really strong conclusions to be drawn from it.

Across all assessment methods, measurement problems are of particular importance in this study because, if individual measures are not performing adequately, correlations between measures may reflect the validity and reliability of the specific measures rather than the validity of the methodology type.

Conclusion

In conclusion, it would appear that different types of activity measure assess different, albeit related, phenomena. Each measures a slightly different aspect of activity and it cannot be assumed that a finding with one measure will replicate with another. On choosing activity measures, researchers must be clear about exactly what aspect of activity they wish to assess and select specific measures accordingly. Until a 'gold standard' activity measure can be devised, gaining the most accurate possible picture of what an individual does, or is able to do, is likely to require combining scores from more than one type of activity measure.

Measurement Recommendations for Future Research

Self-report

Self-report measures are extremely flexible and can be used to access activity information from any chosen environment and any time period. Measures may address both what participants *can* do and what they *actually* do in day-to-day living. However, questions asking about what participants *can* do risk reflecting participant confidence rather than actual ability; such items should be phrased with care. More generally, self-report questionnaires should be suitable for the sample population with respect to types of activity included and time frame.

Proxy-report

Similarly to self-report measures, proxy-report items can reflect activity over a range of time frames and environments and can refer both to target participants' *ability* and *actual* behaviour in everyday life. Researchers should be aware that items addressing what participants *can* do may reflect where proxies help participants, which may not also indicate participants' real ability. Questionnaires should address types of activity and time frames suitable for the population sampled.

Observed Performance

Observed performance methods allow measures to be taken of participants' ability (or what participants *actually* do in the assessment situation) that do not depend on the perceptions of either participants or proxy reporters. Tasks performed should be appropriate for the sample population. Training effects may occur and so, to gain a more accurate picture of participants' ability it may be necessary to build practice trials into study designs.

Continuous Monitoring

Continuous monitoring allows a relatively objective measure of activity performed in day-to-day activity to be taken. Care should be taken in selecting devices such that they detect activity relevant to the study population and that they are reliable. Clear evidence that monitors are not likely to break down and that inter-monitor consistency is reasonable should be sought. It may be necessary to choose between a reliable, simple monitor and a relatively unproven, sophisticated device. Caution should be taken with the latter. Monitors should also be sufficiently comfortable and acceptable to participants.

Research question: What does the participant* actually *do?
a) in every day activity
Self-report
Proxy-report
Continuous Monitoring.
b) in the assessment situation
Observed Performance

Research question: Estimate what* can *the participant do?
Measurement methods:
Observed Performance
Self-report
Proxy-report

Box 7.1: Summary of use of measurement methods

References

Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology, 87*(1), 49-74.

Ainsworth, B. E., Haskell, W. L., Leon, A. S., Jacobs, D. R., Jr., Montoye, H. J., Sallis, J. F., & Paffenbarger, R. S., Jr. (1993). Compendium of physical activities: classification of energy costs of human physical activities. *Medicine and Science in Sports and Exercise, 25*, 71-80.

Ainsworth, B. E., Leon, A. S., Richardson, M. T., Jacobs, D. R., & Paffenbarger, R. S. (1993). Accuracy of the College Alumnus Physical-Activity Questionnaire. *Journal of Clinical Epidemiology, 46*(12), 1403-1411.

Ajzen, I. (1985). From intention to actions: a theory of planned behavior. In J. Kuhl & J. Beckman (Eds.), *Action Control: from Cognitions to Behavior* (pp. 11-39). New York: Springer-Verlag.

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes, 50*, 179-211.

Albersnagel, F. A. (1988). Velten and musical mood induction procedures: A comparison with accessibility of thought associations. *Behaviour Research and Therapy, 26*(1), 79-96.

Anastasiades, P., & Johnston, D. W. (1990). A simple activity measure for use with ambulatory subjects. *Psychophysiology, 27*(1), 87-93.

Andresen, E. M., Vahle, V. J., & Lollar, D. (2001). Proxy reliability: Health-related quality of life (HRQoL) measures for people with disability. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation., 10*(7), 609-619.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioural change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behaviour: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*(2), 230-244.

Barlow, J. H., Williams, B., & Wright, C. C. (1999). 'Instilling the strength to fight the pain and get on with life': learning to become an arthritis self-manager through an adult education programme. *Health Education Research, 14*(4), 533-544.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality & Social Psychology, 51*(6), 1173-1182.

Bassett, D. R., Cureton, A. L., & Ainsworth, B. E. (2000). Measurement of daily walking distance-questionnaire versus pedometer. *Medicine and Science in Sports and Exercise, 32*(5), 1018-1023.

Bassett, D. R., Vachon, J. A., Kirkland, E. T., Howley, G. E., Duncan, G. E., & Johnson, K. R. (1997). Energy cost of stair climbing and descending on the college alumnus questionnaire. *Medicine and Science in Sports and Exercise, 29*, 1250-1254.

Bendel, R. B., & Afifi, A. A. (1977). Comparison of stopping rules in forward 'stepwise' regression. *Journal of the American Statistical Association, 72*(357), 46-53.

Blair, S. N. (1992). Assessment of physical-activity in clinical and epidemiologic research. *Health Psychology, 11*, 48-48.

Blair, S. N., Dowda, M., Pate, R. R., Kronenfeld, J., Howe, H. G., Jr., Parker, G., Blair, A., & Fridinger, F. (1991). Reliability of long-term recall of participation in physical activity by middle aged men and women. *American Journal of Epidemiology, 133*(3), 266-275.

Blair, S. N., Haskell, W. L., Ho, P., Paffenbarger, R. S., Jr., Vranizan, K. M., Farquhar, J. W., & Wood, P.D. (1985). Assessment of habitual physical activity by a seven-day recall in a community survey and controlled experiments. *American Journal of Epidemiology, 122*(5), 794-804.

Bond, M. J., & Clark, M. S. (2000). The accuracy of proxy ratings of life-style activities using the Adelaide Activities Profile. *Aging-Clinical and Experimental Research, 12*(5), 388-392.

Bonetti, D. (2000). *Perceived control, desire for control and health outcomes.* Unpublished PhD Thesis, University of St Andrews.

Bonetti, D., Johnston, M., Morrison, V., MacWalter, R., & Pollard, B. (2001). *A randomised controlled trial of a workbook-based intervention for stroke patients: Effects on disability and distress in patients and carers.* (K/CR1/1/7): Report to the Chief Scientist Office.

Borson, S., McDonald, G. J., Gayle, T., Deffebach, M., Lakshminarayan, S., & VanTuinen, C. (1992). Improvement in mood, physical symptoms, and function with nortriptyline for depression in patients with chronic obstructive pulmonary disease. *Psychosomatics, 33*(2), 190-201.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36*(2), 129-148.

Bower, G. H., Gilligan, S. G., & Monteiro, K. P. (1981). Selectivity of learning caused by affective states. *Journal of Experimental Psychology: General, 110*(4), 451-473.

Bower, G. H., & Mayer, J. D. (1989). In search of mood-dependent retrieval. *Journal of Social Behavior and Personality, 4*(2), 121-156.

Brown, J. D., & Mankowski, T. A. (1993). Self-esteem, mood, and self-evaluation: Changes in mood and the way you see you. *Journal of Personality and Social Psychology, 64*(3), 421-430.

Bruce, M. L., Seeman, T. E., Merrill, S. S., & Blazer, D. G. (1994). The impact of depressive symptomatology on physical disability: MacArthur studies of successful aging. *American Journal of Public Health, 84*(11), 1796-1799.

Bryman, A., & Cramer, D. (1997). *Quantitative Data Analysis with SPSS for Windows*. London & New York: Routledge.

Bussmann, J. B. J., Tulen, J. H. M., van Heral, E. C. G., & Stam, H. J. (1998). Quantification of physical activities by means of ambulatory accelerometry: A validation study. *Psychophysiology, 35*, 488-496.

Carter, F. A., Wilson, J. S., Lawson, R. H., & Bulik, C. M. (1995). Mood induction procedure: Importance of individualising music. *Behaviour Change, 12*(3), 159-161.

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Reports, 100*, 126-130.

Charlton, J. R. H. (1989). Measuring disability in a longitudinal study. In D. L. Patrick & H. Peach (Eds.), *Disablement in the Community*. Oxford: Oxford University Press.

Chastain, G., Seibert, P. S., & Ferraro, F. R. (1995). Mood and lexical access of positive, negative, and neutral words. *Journal of General Psychology, 122*(2), 137-157.

Clark, D. M. (1983). On the induction of depressed mood in the laboratory: Evaluation and comparison of the Velten and musical procedures. *Advances in Behavior Research and Therapy, 5*, 27-49.

Clark, D. M., & Teasdale, J. D. (1985). Constraints on the effects of mood on memory. *Journal of Personality and Social Psychology, 48*(6), 1595-1608.

Conn, V. S., Minor, M. A., Mehr, D. R., & Burks, K. J. (2000). Recording activity in older women with TriTrac. *American Journal of Health Behavior, 24*(5), 370-378.

Conner, M., & Norman, P. (1996). *Predicting Health Behaviour*. Buckingham: Open University Press.

Cooper, K. H. (1968). A means of assessing maximal oxygen intake: Correlation between field testing and treadmill testing. *Journal of the American Medical Association, 203*(3), 201-204.

Courneya, K. S., & McAuley, E. (1994). Are there different determinants of the frequency, intensity, and duration of physical activity? *Behavioral Medicine, 20*(2), 84-90.

Craig, K. D., Hyde, S., & Patrick, C. J. (1991). Genuine, suppressed, and faked facial behavior during exacerbation of chronic low back pain. *Pain, 46*, 161-172.

Craig, K. D., & Patrick, C. J. (1985). Facial expression during induced pain. *Journal of Personality and Social Psychology, 48*(4), 1080-1091.

Craig, K. D., Prkachin, K. M., & Grunau, R. V. E. (1992). The facial expression of pain. In D. C. Turk & R. Melzack (Eds.), *Handbook of pain assessment*. London: Guildford Press.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 344-354.

Davis, M. A. (1981). Sex differences in reporting osteoarthritic symptoms: A sociomedical approach. *Journal of Health and Social Behavior, 22*, 298-310.

DiPietro, L., Caspersen, C. J., Ostfeld, A. M., & Nadel, E. R. (1993). A survey for assessing physical activity among older adults. *Medicine and Science in Sports and Exercise, 25*, 628-642.

Dishman, R. K., Darracott, C. R., & Lambert, L. T. (1992). Failure to generalize determinants of self-reported physical activity to a motion sensor. *Medicine and Science in Sports and Exercise, 24*(8), 904-910.

Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Physiology, 37*, 90-93.

Ekman, P., & Friesen, P. (1978). *Investigator's guide to the Facial Action Coding System.* Palo Alto, CA: Consulting Psychologists Press.

Epstein, A. M., Hael, J. A., Tognetti, J., Son, L. H., & Conant, L. (1989). Using proxies to evaluate quality of life. *Med Care, 27*, S91-S98.

Estlander, A. M., Vanharanta, H., Moneta, G. B., & Kaivanto, K. (1994). Anthropometric variables, self-efficacy beliefs, and pain and disability ratings on the isokinetic performance of low back pain patients. *Spine, 19*(8), 941-947.

Farmer, M. E., Locke, B. Z., Moscicki, E. K., Dannenberg, A. L., Larson, D. B., & Radloff, L. S. (1988). Physical activity and depressive symptoms: The NHANES I Epidemiologic Follow-up Study. *American Journal of Epidemiology, 128*(6), 1340-1351.

Faul, F., & Erdfelder, E. (1992). GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS [computer program]. Bonn, FRG: Bonn University, Department of Psychology.

Ferrer, M., Lamarca, R., Orfila, F., & Alonso, J. (1999). Comparison of performance-based and self-rated functional capacity in Spanish elderly. *American Journal of Epidemiology, 149*(3), 228-235.

Fisher, K., & Johnston, M. (1996a). Emotional distress as a mediator of the relationship between pain and disability: An experimental study. *British Journal of Health Psychology, 1*, 207-218.

Fisher, K., & Johnston, M. (1996b). Experimental manipulation of perceived control and its effects on disability. *Psychology and Health, 11*, 657 - 669.

Fisher, K., & Johnston, M. (1997). Validation of the Oswestry Low Back Pain Disability Questionnaire, its sensitivity as a measure of change following treatment and its relationship with other aspects of the chronic pain experience. *Physiotherapy Theory and Practice, 13*, 67-80.

Fisher, K., & Johnston, M. (1998). Emotional distress and control cognitions as mediators of the impact of chronic pain on disability. *British Journal of Health Psychology, 3*, 225-236.

Frank, G., Johnston, M., Morrison, V., Pollard, B., & MacWalter, R. (2000). Perceived control and recovery from functional limitations: preliminary evaluation of a workbook based intervention for discharged stroke patients. *British Journal of Health Psychology, 5*, 413-420.

Gill, T. M., Robison, J. T., & Tinetti, M. E. (1997). Predictors of recovery in activities of daily living among disabled older persons living in the community. *Journal of General Internal Medicine, 12*(12), 757-762.

Glicksohn, J. (2001). Temporal cognition and the phenomenology of time: A multiplicative function for apparent duration. *Consciousness and Cognition, 10*, 1-25.

Guralnik, J. M., Branch, L. G., Cummings, S. R., & Curb, J. D. (1989). Physical performance measures in aging research. *Journal of Gerontology: Medical Sciences, 44*(5), M141-146.

Guyatt, G. H., Pugsley, S. O., Sullivan, M. J., Thompson, P. J., Berman, L. B., Jones, N. L., Fallen, E. L., & Taylor, D. W. (1984). Effect of encouragement on walking test performance. *Thorax, 39*, 818-822.

Guyatt, G. H., Sullivan, M. J., Thompson, P. J., Fallen, E. L., Pugsley, S. O., Taylor, D. W., & Berman, L. B. (1985). The 6-minute walk: A new measure of exercise capacity in patients with chronic heart failure. *Canadian Medical Association Journal, 132*, 919-923.

Hakala, M., Nieminen, P., & Manelius, J. (1994). Joint impairment is strongly correlated with disability measured by self-report questionnaires.

Functional status assessment of individuals with rheumatoid arthritis in a population based series. *The Journal of Rheumatology, 21*(1), 64-69.

Harding, V. R., Williams, A. C. d. C., Richardson, P. H., Nicholas, M. K., Jackson, J. L., Richardson, I. H., & Pither, C. E. (1994). The development of a battery of measures for assessing physical functioning of chronic pain patients. *Pain, 58*, 367-375.

Härkäpää, K., Järvikoski, A., Mellin, G., Hurri, H., & Luomo, J. (1991). Health locus of control beliefs and psychological distress as predictors for treatment outcome in low-back pain patients: Results of a 3-month follow-up of a controlled intervention study. *Pain, 46*, 35-41.

Härkäpää, K., Järvikoski, A., & Vakkari, T. (1996). Associations of locus of control beliefs with pain coping strategies and other pain-related cognitions in back pain patients. *British Journal of Health Psychology, 1*, 51-63.

Haskell, W. L., Yee, M. C., Evans, A., & Irby, P. J. (1993). Simultaneous measurement of heart rate and body motion to quantitate physical activity. *Medicine and Science in Sports and Exercise, 25*(1), 109-115.

Heatherton, T. E., Striepe, M., & Wittenberg, L. (1998). Emotional distress and disinhibited eating: The role of self. *Personality and Social Psychology Bulletin, 24*(3), 301-313.

Hebert, J. R., Clemow, L., Pbert, L., Ockene, I. S., & Ockene, J. K. (1995). Social desirability bias in dietary self-report may compromise the validity of dietary-intake measures. *International Journal of Epidemiology, 24*(2), 389-398.

Hermans, D., De Houwer, J., & Eelen, P. (1996). Evaluative decision latencies mediated by induced affective states. *Behaviour Research and Therapy, 34*(5-6), 483-488.

Hibbard, J. H., & Pope, C. R. (1983). Gender roles, illness orientation and use of medical services. *Social Science and Medicine, 17*(3), 129-137.

Holm, M. B., Rogers, J. C., & Kwoh, C. K. (1998). Predictors of functional disability in patients with rheumatoid arthritis. *Arthritis Care and Research, 11*(5), 346-355.

Howell, D. C. (1992). *Statistical methods for psychology* (3rd ed.). Belmont, California: Duxbury Press.

Jacobs, D. R., Ainsworth, B. E., Hartman, T. J., & Leon, A. S. (1993). A simultaneous evaluation of 10 commonly used physical activity questionnaires. *Medicine and Science in Sports and Exercise, 25*, 81-91.

Jerusalem, M., & Schwarzer, R. (1992). Self-efficacy as a resource factor in stress appraisal process. In R. Schwarzer (Ed.), *Self-efficacy: Thought control of action*. Washington, D.C.: Hemisphere.

Jette, A. M., & Branch, L. G. (1985). Impairment and disability in the aged. *Journal of Chronic Diseases, 38*, 59-65.

Johnston, M. (1999). Mood in chronic disease: Questioning the answers. *Current Psychology: Developmental ·Learning ·Personality ·Social, 18*(1), 71-87.

Johnston, M., Earll, L., Giles, M., McClenahan, R., Stevens, D., & Morrison, V. (1999). Mood as a predictor of disability and survival in patients newly diagnosed with ALS/MND. *British Journal of Health Psychology, 4*, 127-136.

Johnston, M., Morrison, V., MacWalter, R., & Partridge, C. (1999). Perceived control, coping and recovery from disability following stroke. *Psychology and Health, 14*(2), 181-192.

Johnston, M., & Pollard, B. (1996). Measuring emotional, disability, handicap and quality of life outcomes of chronic disease. *Report to NHS Executive R&D Programme for People with Physical and Complex Disabilities.*

Johnston, M., Pollard, B., & Hennessey, P. (2000). Construct validation of the Hospital Anxiety and Depression Scale with clinical populations. *Journal of Psychosomatic Research, 48*(6), 579-584.

Kaivanto, K. K., Estlander, A. M., Moneta, G. B., & Vanharanta, H. (1995). Isokinetic performance in low back pain patients: The predictive power of the Self-Efficacy Scale. *Journal of Occupational Rehabilitation, 5*(2), 87-99.

Kamarck, T. W., Shiffman, S. M., Smithline, L., Goodie, J. L., Thompson, H. S., Ituarte, P. H. G., Jong, J. Y.-K., Pro, V., Paty, J. A., Kassel, J. D., Gnys, M., & Perz, W. (1998). The diary of ambulatory behavioral states: A new approach to the assessment of psychosocial influences on ambulatory cardiovascular activity. In D. Krantz & H. Baum (Eds.), *Perspectives in Behavioral Medicine: Technology and Methodology in Behavioral Medicine*. Hillsdale, N.J.: Lawrence Erlbaum.

Kelly-Hayes, M., Jette, A. M., Wolf, P. A., D'Agostino, R. B., & Odell, P. M., 82(6), 841-845. (1992). Functional limitations and disability among elders in the Framingham Study. *American Journal of Public Health, 82*(6), 841-845.

Kelvin, R. G., Goodyer, I. M., Teasdale, J. D., & Brechin, D. (1999). Latent negative self-schema and high emotionality in well adolescents at risk for psychopathology. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 40*(6), 959-968.

Kempen, G. I. J. M., van Sonderen, E., & Ormel, J. (1999). The impact of psychological attributes on changes in disability among low-functioning older persons. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 54*(1), P23-P29.

Kenealy, P. (1988). Validation of a music mood induction procedure: Some preliminary findings. *Cognition and Emotion, 2*(1), 41-48.

Kenealy, P. M. (1997). Mood-state-dependent retrieval: The effects of induced mood on memory reconsidered. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 2*, 290-317.

Klassen, D., Hornstra, R. K., & Anderson, P. B. (1976). Influence of social desirability on symptom and mood reporting in a community survey. *Journal of Consulting and Clinical Psychology, 43*(4), 448-452.

Klesges, R. C., Hanson, C. L., Eck, L. H., Mellon, M., Fulliton, W., & Somes, G. (1990). The accuracy of self-reports of physical activity. *Medicine and Science in Sports and Exercise, 22*(5), 690-697.

Klesges, R. C., Klesges, L. M., Swenson, A. M., & Pheley, A. M. (1985). A validation of two motion sensors in the prediction of child and adult physical activity levels. *American Journal of Epidemiology, 122*, 400-410.

Koivula, N. (1996). Estimation of time: Effects of locus of control, mental arithmetic and length of target interval. *Personality and Individual Differences, 20*(1), 25-32.

Kovar, M. G., & Lawton, M. P. (1994). Functional disability: Activities and instrumental activities of daily living. In M. P. Lawton & J. A. Teresi (Eds.), *Annual review of gerontology and geriatrics: Focus on assessment techniques.* (Vol. 14, pp. 57-75). New York: Springer Publishing Company.

Kriska, A. M., & Caspersen, C. J. (1997). A collection of physical activity questionnaires for health-related research. *Medicine and Science in Sports and Exercise, 29(suppl)*, S1-S205.

Kwiatkowski, S. J., & Parkinson, S. R. (1994). Depression, elaboration, and mood congruence - differences between natural and induced mood. *Memory & Cognition, 22*(2), 225-233.

Lackner, J. M., & Carosella, A. M. (1999). The relative influence of perceived pain control, anxiety, and functional self efficacy on spinal function among patients with chronic low back pain. *Spine, 24*(21), 2254-2260.

Lackner, J. M., Carosella, A. M., & Feuerstein, M. (1996). Pain expectancies, pain and functional self-efficacy expectancies as determinants of disability in patients with chronic low back disorders. *J Consult Clin Psychol, 64*(1), 212-220.

Langer, E. J. (1983). *Psychology of Control*. Beverly Hills, CA: Sage.

LaPorte, R. E., Black-Sandler, R., Cauley, J. A., Link, M., Bayles, C., & Marks, B. (1983). The assessment of physical activity in older women; Analysis of the interrelationship and reliability of activity monitoring, activity surveys, and caloric intake. *Journal of Gerontology, 38*, 394-397.

Léger, L. A., & Lambert, J. (1982). A maximal 20-m shuttle run test to predict VO2 max. *European Journal of Applied Physiology and Occupational Physiology, 49*, 1-12.

Lenton, S. R., & Martin, P. R. (1991). Case histories and shorter communications: The contribution of music vs instructions in the Musical Mood Induction Procedure. *Behaviour Research & Therapy, 29*(6), 623-625.

Levin, S., Jacobs, D. R., Ainsworth, B. E., Richardson, M. T., & Leon, A. S. (1999). Intra-individual variation and estimates of usual physical activity. *Annals of Epidemiology, 9*(8), 481-488.

Lewis, L. M., Dember, W. N., Schefft, B. K., & Radenhausen, R. A. (1995). Can experimentally induced mood affect optimism and pessimism scores? *Current Psychology: Developmental, Learning, Personality, Social, 14*(1), 29-41.

Lorish, C. D., Abraham, N., Austin, J., Bradley, L. A., & Alarcón, G. S. (1991). Disease and psychosocial factors related to physical functioning in rheumatoid arthritis. *Journal of Rheumatology, 18*, 1150-1157.

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

Magaziner, J., Zimmerman, S. I., Gruber-Baldini, A. L., Hebel, J. R., & Fox, K. M. (1997). Proxy reporting in five areas of functional status - Comparison with self-reports and observations of performance. *American Journal of Epidemiology, 146*(5), 418-428.

Marteau, T., & Johnston, M. (1986). Determinants of beliefs about illness: A study of parents of children with diabetes, asthma, epilepsy and no chronic illness. *Journal of Psychosomatic Research, 30*(6), 673-683.

Martin, M. A., & Metha, A. (1997). Recall of early childhood memories through musical mood induction. *Arts in Psychotherapy, 24*(5), 447-454.

Matthews, C. E., & Freedson, P. S. (1995). Field trial of a three-dimensional activity monitor: comparison with self-report. *Medicine and Science in Sports and Exercise, 27*, 1071-1100.

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

McCrae, R. R., Bartone, P. T., & Costa, P. T., Jr. (1976). Age, anxiety and self-reported health. *Aging and Human Development, 7*(1), 49-58.

McDermott, M. M., Liu, K., O'Brien, E., Guralnik, J. M., Criqui, M. H., Martin, G. J., & Greenland, P. (2000). Measuring physical activity in peripheral arterial disease: A comparison of two physical activity questionnaires with an accelerometer. *Angiology, 51*(2), 91-100.

McGavin, C. R., Gupta, S. P., & McHardy, G. J. R. (1976). Twelve-minute walking test for assessing disability in chronic bronchitis. *British Medical Journal, 1*, 822-823.

Mechanic, D. (1976). Sex, illness behavior, and the use of health services. *Journal of Human Stress, 2*, 29-40.

Mendes de Leon, C. F., Seeman, T. E., Baker, D. I., Richardson, E. D., & Tinetti, M. E. (1996). Self-efficacy, physical decline, and change in functioning in community-living elders: A prospective study. *Journal of Gerontology: Social Sciences, 51B*(4), S183-S190.

Merrill, S. S., Seeman, T. E., Kasl, S. V., & Berkman, L. F. (1997). Gender differences in the comparison of self-reported disability and performance measures. *Journal of Gerontology: Medical Sciences, 52A*(1), M19-M26.

Miller, D. J., Freedson, P. S., & Kline, G. M. (1994). Comparison of activity levels using the Caltrac accelerometer and five questionnaires. *Medicine and Science in Sports and Exercise, 26*, 376-382.

Moinpour, C. M., Lyons, B., Schmid, S. P., Chansky, K., & Patchell, R. A. (2000). Substituting proxy ratings for patient ratings in cancer clinical trials: An analysis based on a Southwest Oncology Group trial in patients with brain metastases. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation., 9*(2), 219-231.

Moorey, S., Greer, S., Watson, M., Gorman, C., Rowden, L., Tunmore, R., Robertson, B., & Bliss, J. (1991). The factor structure and factor stability of the Hospital Anxiety and Depression Scale in patients with cancer. *British Journal of Psychiatry, 158*, 255-259.

Mora, P. A., Robitaille, C., Leventhal, H., Swigar, M., & Leventhal, E. A. (2002). Trait negative affect relates to prior-week symptoms, but not to reports of illnss episodes, illness symptoms, and care seeking among older persons. *Psychosomatic Medicine, 64*(3), 436-449.

Mungall, I. P. F., & Hainsworth, R. (1979). Assessment of respiratory function in patients with chronic obstructive airways disease. *Thorax, 34*, 254-258.

Myers, A. M., Holliday, P. J., Harvey, K. A., & Hutchinson, K. S. (1993). Functional performance measures: Are they superior to self-assessments? *Journal of Gerontology: Medical Sciences, 48*(5), M196-M206.

Neumann, P. J., Araki, S. S., & Gutterman, E. M. (2000). The use of proxy respondents in studies of older adults: Lessons, challenges, and opportunities. *Journal of the American Geriatrics Society, 48*(12), 1646-1654.

Nisbett, R., & Ross, L. (1980). *Human Inference: Strategies and shortcomings of social judgement.* Englewood Cliffs, NJ: Prentice-Hall.

Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression: Evidence and theory. *Psychological Bulletin, 101*(2), 259-282.

O'Leary, A., Shoor, S., Lorig, K., & Holman, H. R. (1988). A cognitive-behavioural treatment for rheumatoid arthritis. *Health Psychology, 7*(6), 527-544.

Orbell, S., Johnston, M., Rowley, D., Davey, P., & Espley, A. (2001). Self-efficacy and goal importance in the prediction of physical disability in people following hospitalization: A prospective study. *British Journal of Health Psychology, 6*, 25-40.

Ostir, G. V., Carlson, J. E., Black, S. A., Rudkin, L., Goodwin, J. S., & Markides, K. S. (1999). Disability in older adults 1: Prevalence, causes and consequences. *Behavioral Medicine, 24*(4), 147-156.

Paffenbarger, R. S., Jr., Wing, A. L., & Hyde, R. T. (1978). Physical activity as an index of heart attack risk in college alumni. *American Journal of Epidemiology, 108*(3), 161-175.

Parrott, W. G., & Sabini, J. (1990). Mood and memory under natural conditions: Evidence for mood incongruent recall. *Journal of Personality and Social Psychology, 59*(2), 321-336.

Partridge, C., & Johnston, M. (1989). Perceived control of recovery from physical disability: Measurement and prediction. *British Journal of Clinical Psychology, 28*, 53-59.

Partridge, C. J., Johnston, M., & Edwards, S. (1987). Recovery from physical disability after stroke: normal patterns as a basis for evaluation. *The Lancet*(February 14), 373-375.

Patrick, C. J., Craig, K. D., & Prkachin, K. M. (1986). Observer judgments of acute pain: Facial action determinants. *Journal of Personality and Social Psychology, 50*(6), 1291-1298.

Patrick, D. L., & Peach, H. (1989). *Disablement in the Community*. Oxford: Oxford University Press.

Patterson, S. M., Krantz, D. S., Montgomery, L. C., Deuster, P. S., Hedgcs, S. M., & Nebel, L. E. (1993). Automated physical activity monitoring: Validation and comparison with physiological and self-report measures. *Psychophysiology, 30*, 296-305.

Peck, J. R., Smith, T. W., Ward, J. R., & Milano, R. (1989). Disability and depression in rheumatoid arthritis: A multi-trait, multi-method investigation. *Arthritis and Rheumatism, 32*(9), 1100-1106.

Phillips, D. L., & Clancy, K. J. (1970). Response biases in field studies of mental illness. *American Sociological Review, 35*, 503-515.

Pignatiello, M., Camp, C. J., Elder, S. T., & Rasar, L. A. (1989). A psychophysiological comparison of the Velten and musical mood induction techniques. *Journal of Music Therapy, 26*(3), 140-154.

Pignatiello, M. F., Camp, C. J., & Rasar, L. A. (1986). Musical Mood Induction - an Alternative to the Velten Technique. *Journal of Abnormal Psychology, 95*(3), 295-297.

Podsiadlo, D., & Richardson, S. (1991). The timed 'up & go': A test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society, 39*, 142-148.

Pollard, B., & Johnston, M. (2001). Problems with the Sickness Illness Profile: A theoretically based analysis and a proposal for a new method of implementation and scoring. *Social Science and Medicine, 52*(6), 921-934.

Pols, M. A., Peeters, P. H. M., Kemper, H. C. G., & Collette, H. J. A. (1996). Repeatability and relative validity of two physical activity questionnaires in elderly women. *Medicine and Science in Sports and Exercise, 28*(8), 1020-1025.

Prkachin, K. M. (1992). Dissociating spontaneous and deliberate expressions of pain: Signal detection analyses. *Pain, 51*, 57-65.

Prkachin, K. M., Berzins, S., & Mercer, S. (1994). Encoding and decoding of pain expressions: A judgement study. *Pain, 58*, 253-259.

Rejeski, W. J., Craven, T., Ettinger, W. H., McFarlane, M., & Shumaker, S. (1996). Self-efficacy and pain in disability with osteoarthritis of the knee. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 51*(1), P24-P29.

Rejeski, W. J., Miller, M. E., Foy, C., Messier, S., & Rapp, S. (2001). Self-efficacy and the progression of functional limitations and self-reported disability in older adults with knee pain. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 56*(5), S261-S265.

Rennie, K. L., & Wareham, N. (1998). The validation of physical activity instruments for measuring energy expenditure: problems and pitfalls. *Public Health Nutrition, 1*(4), 265-271.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 38*(1), 119-125.

Rholes, W. S., Riskind, J. H., & Lane, J. W. (1987). Emotional states and memory biases: Effects of cognitive priming and mood. *Journal of Personality & Social Psychology, 52*(1), 91-99.

Riskind, J. H., & Rholes, W. S. (1985). The cognitive model of depression and mood induction procedures: A reply to Clark. *Behavior Research and Therapy, 23*(6), 663-666.

Riskind, J. H., Rholes, W. S., & Eggers, J. (1982). The Velten Mood Induction Procedure - Effects on Mood and Memory. *Journal of Consulting and Clinical Psychology, 50*(1), 146-147.

Rothman, M. L., Hedrick, S. C., Bulcroft, K. A., Hickam, D. H., & Rubenstein, L. Z. (1991). The validity of proxy-generated scores as measures of patient health status. *Med Care, 29*, 115-124.

Rubenstein, L. Z., Schairer, C., Wieland, G. D., & Kane, R. (1984). Systematic biases in functional status assessments of elderly adults: Effects of different data sources. *Journal of Gerontology, 39*(6), 686-691.

Sallis, J. F., Haskell, W. L., Wood, P. D., Fortmann, S. P., Rogers, T., Blair, S. N., & Paffenbarger, R. S. (1985). Physical activity assessment methodology in the five-city project. *American Journal of Epidemiology, 121*(1), 91-106.

Salonen, J. T., & Lakka, T. (1987). Assessment of physical activity in population studies - validity and consistency of the methods in teh Kuopio ischemic heart disease risk factor study. *Scand. J. Sports Sci., 9*, 89-95.

Samsom, D., & Rachman, S. (1989). The effect of induced mood on fear reduction. *British Journal of Clinical Psychology, 28*(3), 227-238.

Sanders, S. H. (1983). Automated versus self-monitoring of 'up-time' in chronic low-back pain patients: a comparative study. *Pain, 15*, 399-405.

Schiaffino, K. M., & Revenson, T. A. (1992). The role of perceived self-efficacy, perceived control, and causal attributions in adaptation to rheumatoid-arthritis: Distinguishing mediator from moderator effects. *Personality and Social Psychology Bulletin, 18*(6), 709-718.

Schiaffino, K. M., Revenson, T. A., & Gibofsky, A. (1991). Assessing the impact of self-efficacy beliefs on adaptation to rheumatoid arthritis. *Arthritis Care and Research, 4*(4), 150-157.

Schwarzer, R., & Schröder, K. (1997). Effects of self-efficacy and social support on postsurgical recovery of heart patients. *Irish Journal of Psychology, 18*(1), 88-103.

Seeman, T. E., Unger, J. B., McAvay, G., & Mendes de Leon, C. F. (1999). Self-efficacy beliefs and perceived declines in functional ability: MacArthur studies of successful aging. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 54*(4), P214-P222.

Singh, S. J., Morgan, M. D. L., & Hardman, A. E. (1992). The Shuttle Walking Test. © The Glenfield Hospital, Leicester and Loughborough University.

Singh, S. J., Morgan, M. D. L., Scott, S., Walters, D., & Hardman, A. E. (1992). Development of a shuttle walking test of disability in patients with chronic airways obstruction. *Thorax, 47*, 1019-1024.

Smarr, K. L., Parker, J. C., Wright, G. E., StuckyRopp, R. C., Buckelew, S. P., Hoffman, R. W., O'Sullivan, F. X., & Hewett, J. E. (1997). The importance of enhancing self-efficacy in rheumatoid arthritis. *Arthritis Care and Research, 10*(1), 18-26.

Sneeuw, K. C., Aaronson, N. K., Osoba, D., Muller, M. J., Hsu, M. A., Yung, W. K. A., Brada, M., & Newlands, E. S. (1997). The use of significant others as proxy raters of the quality of life of patients with brain cancer. *Medical Care, 35*, 490-506.

Sneeuw, K. C., Aaronson, N. K., Sprangers, M. A., Detmar, S. B., Wever Lidwina, D., & Schornagel, J. H. (1998). Comparison of patient and proxy EORTC QLQ-C30 ratings in assessing the quality of life of cancer patients. *Journal of Clinical Epidemiology, 51*, 617-631.

Sneeuw, K. C. A., Aaronson, N. K., deHaan, R. J., & Limburg, M. (1997). Assessing quality of life after stroke: The value and limitations of proxy ratings. *Stroke, 28*(8), 1541-1549.

Sneeuw, K. C. A., Aaronson, N. K., Sprangers, M. A. G., Detmar, S. B., Wever, L. D. V., & Schornagel, J. H. (1997). Value of caregiver ratings in evaluating the quality of life of patients with cancer. *Journal of Clinical Oncology, 15*(3), 1206-1217.

Sneeuw, K. C. A., Aaronson, N. K., Sprangers, M. A. G., Detmar, S. B., Wever, L. D. V., & Schornagel, J. H. (1999). Evaluating the quality of life of cancer patients: assessments by patients, significant others, physicians and nurses. *British Journal of Cancer, 81*(1), 87-94.

Sneeuw, K. C. A., Albertsen, P. C., & Aaronson, N. K. (2001). Comparison of patient and spouse assessments of health related quality of life in men with metastatic prostate cancer. *Journal of Urology, 165*(2), 478-482.

SOLO. (1992). SOLO Statistical System - Power Analysis [computer program]. LA, California: BMDP Statistical Software Inc.

Spiegel, J. S., Leake, B., Spiegel, T. M., Paulus, H. E., Kane, R. L., Ward, N. B., & Ware, J. E., Jr. (1988). What are we measuring? An examination of self-reported functional status measures. *Arthritis and Rheumatism, 31*(6), 721-728.

SPSS. (1999). *SPSS Advanced Models 9.0.* Chicago: SPSS Inc.

Steele, B. G., Holt, L., Belza, B., Ferris, S., Lakshminaryan, S., & Buchner, D. M. (2000). Quantitating physical activity in COPD using a triaxial accelerometer. *Chest, 117*(5), 1359-1367.

Stein, K. D., Goldman, M. S., & Del Boca, F. K. (2000). The influence of alcohol expectancy priming and mood manipulation on subsequent alcohol consumption. *Journal of Abnormal Psychology, 109*(1), 106-115.

Straseske, C. A. (1989). Musically induced moods: Effects on judgments of self-efficacy. *Dissertation Abstracts International, 50*(1-B), 355.

Sullivan, M. D., LaCroix, A. Z., Russo, J., & Katon, W. J. (1998). Self-efficacy and self-reported functional status in coronary heart disease: A six-month prospective study. *Psychosomatic Medicine, 60*(4), 473-478.

Sutherland, G., Newman, B., & Rachman, S. (1982). Experimental investigations of the relations between mood and intrusive, unwanted cognitions. *British Journal of Medical Psychology, 55*, 127-138.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using Multivariate Statistics* (3rd ed.). New York: HarperCollins College Publishers.

Taylor, C. B., Coffey, T., Berra, K., Iaffaldano, R., Casey, K., & Haskell, W. L. (1984). Seven-day physical activity and self-report compared to a direct measure of physical activity. *American Journal of Epidemiology, 120*, 818-824.

Teasdale, J. D., & Fogarty, S. J. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology, 88*(3), 248-257.

Tessler, R., & Mechanic, D. (1978). Psychological distress and perceived health status. *Journal of Health and Social Behavior, 19*, 254-262.

Tran, P.-V., Schwarz, J., Gorman, M., & Helme, R. D. (1997). Validation of an automated up-timer for measurement of mobility in older adults. *Medical Journal of Australia, 167*(8), 434-436.

Tuomisto, M. T., Johnston, D. W., & Schmidt, T. F. H. (1996). The ambulatory measurement of posture, thigh acceleration and muscle tension and their relationship to heart rate. *Psychophysiology, 33*, 409-415.

Velten, E. (1968). A laboratory task for the induction of mood states. *Behavior Research and Therapy, 6*, 473-482.

Verbrugge, L. M., & Wingard, D. L. (1987). Sex differentials in health and mortality. *Women and Health, 12*(2), 103-145.

Voorrips, L. E., Ravelli, A. C. J., Dongelmans, P. C. A., Deurenberg, P., & Van Staveren, W. A. (1991). A physical activity questionnaire for the elderly. *Medicine and Science in Sports and Exercise, 23*, 974-979.

Walker, D. J., Heslop, P. S., Kay, L. J., & Chandler, C. (1998). Spontaneous activity as a quantifiable outcome measure for osteoarthritis of the knee. *British Journal of Rheumatology, 37*, 969-971.

Walker, D. J., Heslop, P. S., Plummer, C. J., Essex, T., & Chandler, C. (1997). A continuous patient activity monitor: Validation and relation to disability. *Physiological Measurement, 18*, 49-59.

Wallston, K. A., Wallston, B. S., Smith, S., & Dobbins, C. J. (1989). Perceived Control and Health. In M. Johnston & T. Marteau (Eds.), *Applications in Health Psychology*. New Brunswick, NJ: Transaction Publishers.

Wareham, N. J., & Rennie, K. L. (1998). The assessment of physical activity: Why try to be more precise about how physical activity is assessed? *International Journal of Obesity, 22*(suppl 2), S30-S38.

Washburn, R. A., Jette, A. M., & Janney, C. A. (1990). Using age-neutral physical activity questionnaires in research with the elderly. *Journal of Aging and Health, 2*(3), 341-356.

Watson, D., & Clark, L. A. (1984). Negative affectivity : The disposition to experience aversive emotional states. *Psychological Bulletin, 96*(3), 465-490.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Watson, D., & Pennebaker, J. W. (1989). Health complaints, stress and distress: Exploring the central role of negative affectivity. *Psychological Review, 96*(2), 234-254.

Wenzlaff, R. M., Wegner, D. M., & Klein, S. B. (1991). The role of thought suppression in the bonding of thought and mood. *Journal of Personality & Social Psychology, 60*(4), 500-508.

Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of Social Psychology, 26*(4), 557-580.

WHO. (1999). *ICIDH-2: International Classification of Functioning and Disability.* Beta-2 draft, Full Version. Geneva: World Health Organization.

Williams, E., Klesges, R. C., Hanson, C. L., & Eck, L. H. (1989). A prospective study of the reliability and convergent validity of three physical activity measures. *Journal of Clinical Epidemiology, 42*(12), 1161-1170.

Williams, J. M. G. (1992). *The Psychological Treatment of Depression* (2nd ed.). London: Routledge.

Williams, R. G. A., Johnston, M., Willis, L. A., & Bennett, A. E. (1976). Disability: A model and measurement technique. *Brit. J. prev. soc. Med., 30*, 71-78.

Willner, P., Benton, D., Brown, E., Cheeta, S., Davies, G., Morgan, J., & Morgan, M. (1998). "Depression" increases "craving" for sweet rewards in animal and human models of depression and craving. *Psychopharmacology, 136*(3), 272-283.

Willner, P., Field, M., Pitts, K., & Reeve, G. (1998). Mood, cue and gender influences on motivation, craving and liking for alcohol in recreational drinkers. *Behavioural Pharmacology, 9*(7), 631-642.

Willner, P., & Jones, C. (1996). Effects of mood manipulation on subjective and behavioural measures of cigarette craving. *Behavioural Pharmacology, 7*(4), 355-363.

Wingard, D. L. (1984). The sex differential in morbidity, mortality, and life-style. *Annual Review of Public Health, 5*, 443-458.

Wood, J., Saltzberg, J. A., & Goldsamt, L. A. (1990). Does affect induce self-focused attention? *Journal of Personality and Social Psychology, 58*(5), 899-908.

Zelman, D. C., Howland, E. W., Nichols, S. N., & Cleeland, C. S. (1991). The effects of induced mood on laboratory pain. *Pain, 46*(1), 105-111.

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavia, 67*, 361-370.

APPENDICES

Appendix A: Activity Measures

A.i: Adapted Paffenbarger Physical Activity Questionnaire: Past Week.

1. How far do you normally walk each day?_________miles/km (circle)

2. What is your usual pace of walking? (please check one)
   a. Casual or strolling (less than 2mph) ☐ b. Average or normal (2-3mph) ☐
   c. Fairly brisk (3-4mph) ☐ d. Brisk or striding (4mph or faster) ☐

3. How many flights of stairs do you climb up each day?________ *(let 1 flight = 10 steps).*

4. List any sports or recreation you have actively participated in during the past week.

Sport, recreation or other physical activity.	Number of times/week.	Average time/episode. (hours & minutes)	Intensity (light, moderate, vigorous, or very vigorous).
……………………..…………..	…………………..	…………………..	………………………
……………………..…………..	…………………..	…………………..	………………………
……………………..…………..	…………………..	…………………..	………………………
……………………..…………..	…………………..	…………………..	………………………
……………………..…………..	…………………..	…………………..	………………………

5. Which of these statements best expresses your view? (Please check one).
   a. I take enough exercise to keep healthy. ☐ b. I ought to take more exercise. ☐
   c. Don't know. ☐

6. At least once a week, do you engage in regular activity akin to brisk walking, jogging, bicycling, swimming, etc. long enough to work up a sweat, get your heart thumping, or get out of breath?
   No ☐ Why not? ____________________________ Yes ☐ How many times per week? __
   Activity:____________________

7. When you are exercising in your usual fashion, how would you rate your level of exertion (degree of effort)? Please circle one number.

0	0.5	1	2	3	4	5	6	7	8	9	10	*
Normal	very very weak (just noticeable)	Very weak	Weak	Moderate	Somewhat strong	Strong (heavy)		Very strong			Very very strong (almost maximal)	Maximal

8. On a usual day, how much time do you spend on the following activities? Total should add up to 24 hours.

Activity	Time
a) **Vigorous activity** (digging in the garden, strenuous sports, jogging, aerobic dancing, sustained swimming, brisk walking, bicycling on hills, etc.)	
b) **Moderate activity** (housework, light sports, regular walking, golf, painting, repairing, ballroom dancing, bicycling on level ground, etc.)	
c) **Light activity** (office work, driving car, strolling, personal care, standing with little motion, etc.)	
d) **Sitting activity** (eating, reading, desk work, watching TV, listening to radio, etc.)	
e) **Sleeping or reclining.**	

A.ii: Adapted Paffenbarger Physical Activity Questionnaire: Previous Day.

1. How far have you walked during the past day? ___________ miles/km (circle)

2. What was your usual pace of walking? (please check one)
   a. Casual or strolling (less than 2mph) ☐ b. Average or normal (2-3mph) ☐
   c. Fairly brisk (3-4mph) ☐ d. Brisk or striding (4mph or faster) ☐

3. How many flights of stairs have you climbed during the past day? _____*(let 1 flight = 10 steps).*

4. List any sports or recreation you have actively participated in during the past day.

Sport, recreation or other physical activity.	Number of times/day.	Average time/episode. (hours & minutes)	Intensity (light, moderate, vigorous, or very vigorous).
…………………………..…………..	…………………..	…………………..	………………………
…………………………..…………..	…………………..	…………………..	………………………
…………………………..…………..	…………………..	…………………..	………………………
…………………………..…………..	…………………..	…………………..	………………………
…………………………..…………..	…………………..	…………………..	………………………

5. When you were exercising, how would you rate your level of exertion (degree of effort)? Please circle one number. (Scale from 0 – 10).

0	0.5	1	2	3	4	5	6	7	8	9	10	*
Normal	very very weak (just noticeable)	Very weak	Weak	Moderate	Somewhat strong	Strong (heavy)		Very strong			Very very strong (almost maximal)	Maximal

6. During the past day, how much time did you spend on the following activities? Total should add up to 24 hours.

Activity	Time
b) Vigorous activity (digging in the garden, strenuous sports, jogging, aerobic dancing, sustained swimming, brisk walking, bicycling on hills, etc.)	
b) Moderate activity (housework, light sports, regular walking, golf, painting, repairing, ballroom dancing, bicycling on level ground, etc.)	
c) Light activity (office work, driving car, strolling, personal care, standing with little motion, etc.)	
d) Sitting activity (eating, reading, desk work, watching TV, listening to radio, etc.)	
e) Sleeping or reclining.	

A.iii: Self-reports of standardised laboratory tasks.

1.How far do you think you walked in doing the walking task? _____metres/feet/yards (circle)

2. For how long do you think you were walking in the task? _____ minutes _____ seconds

3. When you were asked to sit, stand and lie down for specific time periods, for how long do you think you were:

a. Lying? _____ minutes _____ seconds
b. Sitting? _____minutes _____ seconds
c. Standing?_____ minutes _____seconds

4. Over the whole laboratory session, how long do you think you have spent completing questionnaires? _____ minutes _____ seconds

A.iv: Activity Checklist.

Please tick the boxes corresponding to activities undertaken each hour during the time you were wearing the monitor. Please also indicate time spent lying down. Tick as many boxes as necessary for each time period

Please circle the time at which this 24-hour record begins (this will be the time at which you left the laboratory yesterday).

Time	**Lying**	**Sitting**	**Standing**	**Walking**	**Running**	**Cycling**
00.00 - 01.00						
01.00 - 02.00						
02.00 - 03.00						
03.00 - 04.00						
04.00 - 05.00						
05.00 - 06.00						
06.00 - 07.00						
07.00 - 08.00						
08.00 - 09.00						
09.00 - 10.00						
10.00 - 11.00						
11.00 - 12.00						
12.00 - 13.00						
13.00 - 14.00						
14.00 - 15.00						
15.00 - 16.00						
16.00 - 17.00						
17.00 - 18.00						
18.00 - 19.00						
19.00 - 20.00						
20.00 - 21.00						
21.00 - 22.00						
22.00 - 23.00						
23.00 - 00.00						

A.v: Modified Functional Limitations Profile.

The following statements describe walking and use of stairs. Please think of yourself *today*, and tell me if you agree or disagree with the statement, *and if this is due to the state of your health.*

1. I do not walk at all.	126
2. I get about in a wheel chair.	121
3. I do not use stairs at all.	106
4. I only walk with help from somebody else.	98
5. I get about only by using a walking frame, crutches, stick, walls, or hold on to furniture.	96
6. I only go up and down stairs with assistance from somebody else.	87
7. I only use stairs with a physical aid; for example, special rail, stick or crutches.	82
8. I walk by myself but with some difficulty; for example, I limp, wobble, stumble or I have a stiff leg.	71
9. I do not walk up or down hills.	64
10. I go up and down stairs more slowly; for example, one step at a time or I often have to stop.	62
11. I walk shorter distances or often stop for a rest.	54
12. I walk more slowly [than I would without my walking difficulty].	39

A.vi: Yale Physical Activity Survey, Section 2.

These next questions are about how active you have been over the past month. They can be quite difficult to answer, so take your time.

1. About how many times during the month did you participate in vigorous activities that lasted at least 10 minutes and caused large increases in breathing, heart rate, or leg fatigue or caused you to perspire? ....

0 = Not at all (go to 3)
1 = 1-3 times per month
2 = 1-2 times per week
3 = 3-4 times per week
4 = 5+ times per week

7 = Refused
8 = Don't know. **Frequency score** = ………………

2. About how long did you do this vigorous activity(ies) each time?……………………..

0 = Not applicable
1 = 10-30 minutes
2 = 31-60 minutes
3 = 60+ minutes
7 = Refused
8 = Don't know. **Duration score** = …………………
Weight = **5**

VIGOROUS ACTIVITY INDEX SCORE:
Freq score………x dur score………x weight …………. = …………...
[responses of 7 or 8 are scored as missing]

1. Think about the walks you have taken during the past month. About how many times per month did you walk for at least 10 minutes or more without stopping which was not strenuous enough to cause large increases in breathing, heart rate, or leg fatigue or cause you to perspire?…………………………………………………………

0 = Not at all
1 = 1-3 times per month
2 = 1-2 times per week
3 = 3-4 times per week
4 = 5+ times per week
7 = Refused
8 = Don't know. **Frequency score** = ………………

2. When you did this walking, for how many minutes did you do it?……………………..

0 = Not applicable
1 = 10-30 minutes
2 = 31-60 minutes
3 = 60+ minutes
7 = Refused
8 = Don't know. **Duration score** = …………………
Weight = **4**

LEISURELY WALKING INDEX SCORE:
Freq score………x dur score………x weight …………. = …………...

3. About how many hours a day do you spend moving around on your feet while doing things? Please report only the time that you are actually moving……………………..

0 = Not at all
1 = Less than 1 hr per day
2 = 1 to less than 3 hrs per day
3 = 3 to less than 5 hrs per day
4 = 5 to less than 7 hrs per day
5 = 7+ hrs per day
7 = Refused
8 = Don't know.

Moving Score = ……………………
Weight = 3

MOVING INDEX SCORE:
Moving score ………… x weight …………….. = ……………..

4. Think about how much time you spend standing or moving around on your feet on an average day during the past month. About how many hours per day do you stand? ……………………………………………………………………………………

0 = Not at all
1 = Less than 1 hr per day
2 = 1 to less than 3 hrs per day
3 = 3 to less than 5 hrs per day
4 = 5 to less than 7 hrs per day
5 = 7+ hrs per day
7 = Refused
8 = Don't know.

Standing score = …………………
Weight = 2

STANDING INDEX SCORE:
Standing score ………… x weight …………….. = ……………..

5. About how many hours did you spend sitting on an average day during the past month?……………………………………………………………………………………..

0 = Not at all
1 = Less than 3 hrs
2 = 3 to less than 6 hrs
3 = 6 to less than 8 hrs
4 = 8+ hrs per day
7 = Refused
8 = Don't know.

Sitting score = ……………………
Weight = 1

SITTING INDEX SCORE:
Sitting score ………… x weight …………….. = ……………..

6. About how many flights of stairs do you climb up each day? (let 10 steps = 1 flight)………………………………………………………………………………………

7. Please compare the amount of physical activity that you do during other seasons of the year with the amount of activity you just reported for a typical week in the past month. For example, in the summer, do you do more or less activity than what you reported doing in the past month?

	Lot more	Little more	Same	Little less	Lot less	Don't know
Spring	1.30	1.15	1.00	0.85	0.70	.
Summer	1.30	1.15	1.00	0.85	0.70	.
Fall	1.30	1.15	1.00	0.85	0.70	.
Winter	1.30	1.15	1.00	0.85	0.70	.

Seasonal adjustment score = sum over all seasons / 4 = …………………………………

A.vii: Modified Yale Physical Activity Survey: 24-hours.

1. About how many times during the past 24 hours did you participate in vigorous activities that lasted at least 10 minutes and caused large increases in breathing, heart rate, or leg fatigue or caused you to perspire?…………………………………………..

0 = Not at all (go to 3)
7 = Refused
8 = Don't know. **Frequency score =** ………………

2. About how long did you do this vigorous activity(ies) each time?……………………

0 = Not applicable
1 = 10-30 minutes
2 = 31-60 minutes
3 = 60+ minutes
7 = Refused
8 = Don't know. **Duration score =** …………………
Weight = 5

3. Think about the walks you have taken during the past 24 hours. About how many times did you walk for at least 10 minutes or more without stopping which was not strenuous enough to cause large increases in breathing, heart rate, or leg fatigue or cause you to perspire?………………………………………………………………...

0 = Not at all
7 = Refused
8 = Don't know. **Frequency score** = ………………

4. When you did this walking, for how many minutes did you do it?………………

0 = Not applicable
1 = 10-30 minutes
2 = 31-60 minutes
3 = 60+ minutes
7 = Refused
8 = Don't know. **Duration score** = …………………
Weight = 4

5. About how many hours during the past 24 hours did you spend moving around on your feet while doing things? Please report only the time that you are <u>actually moving</u>………………………………………………………………………………

0 = Not at all
1 = Less than 1 hr per day
2 = 1 to less than 3 hrs per day
3 = 3 to less than 5 hrs per day
4 = 5 to less than 7 hrs per day
5 = 7+ hrs per day
7 = Refused
8 = Don't know. **Moving Score** = ……………………
Weight = 3

MOVING INDEX SCORE:
Moving score ………… x weight …………….. = ……………..

6. Think about how much time you spent standing or moving around on your feet during the past 24 hours. About how many hours per day do you <u>stand?</u>……………………

0 = Not at all
1 = Less than 1 hr per day
2 = 1 to less than 3 hrs per day
3 = 3 to less than 5 hrs per day
4 = 5 to less than 7 hrs per day
5 = 7+ hrs per day
7 = Refused
8 = Don't know. **Standing score** = …………………
Weight = 2

STANDING INDEX SCORE:
Standing score ………… x weight …………….. = ……………..

7. About how many hours did you spend sitting during the past 24 hours?

0 = Not at all
1 = Less than 3 hrs
2 = 3 to less than 6 hrs
3 = 6 to less than 8 hrs
4 = 8+ hrs per day
7 = Refused
8 = Don't know.

Sitting score = ……………………
Weight = 1

SITTING INDEX SCORE:
Sitting score ………… x weight ……………... = ……………..

8. About how many flights of stairs did you climb <u>up</u> during the past 24 hours? (let 10 steps = 1 flight)……………………………………………………………………………

9. Please compare the amount of physical activity that you did during the past 24 hours with the amount of activity you do during a typical day. Did you do more or less activity than you would usually do?

Lot more	Little more	Same	Little less	Lot less	Don't know
1.30	1.15	1.00	0.85	0.70	.

If more/less:
Why did you do more/less activity than during a typical day?
…………………………………………………………………………………………

A.viii: Observer Assessed Disability (standing and walking).

<u>STANDING AND WALKING</u>

10. Without using your hands to pull yourself up Can you now stand up so that you are standing free
11. Can you stay there standing in that position for 2 minutes
12. I would like you now to take 2 steps forward
13. and...2 steps backwards
14. Can you walk for me now...back [to the living room]
15. You have shown me your walking in the house, rather than go outside can you just tell me if you have been outside in the garden or street and have walked alone (no other persons help)

Appendix B: Measures of Bias

B.i: Positive and Negative Affectivity Schedule.

This scale consists of a number of words that describe different feelings and emotions. Read each item and then circle the number corresponding to the appropriate answer. Indicate to what extent you feel this way right now, that is, at the present moment. Or: For each item, would you tell me to what extent you feel this way right now, that is, at the present moment.

	very slightly or not at all	a little	Moderately	quite a bit	extremely
interested	1	2	3	4	5
distressed	1	2	3	4	5
excited	1	2	3	4	5
upset	1	2	3	4	5
strong	1	2	3	4	5
guilty	1	2	3	4	5
scared	1	2	3	4	5
hostile	1	2	3	4	5
enthusiastic	1	2	3	4	5
proud	1	2	3	4	5
irritable	1	2	3	4	5
alert	1	2	3	4	5
ashamed	1	2	3	4	5
inspired	1	2	3	4	5
nervous	1	2	3	4	5
determined	1	2	3	4	5
attentive	1	2	3	4	5
jittery	1	2	3	4	5
active	1	2	3	4	5
afraid.	1	2	3	4	5

B.ii: Marlowe-Crowne Social Desirability Scale.

Listed below are a number of statements concerning personal attitudes and traits. Read each item and decide whether the statement is **true** or **false** as it pertains to you personally.

	True	False
1. Before voting I thoroughly investigate the qualifications of all the candidates.	☐	☐
2. I never hesitate to go out of my way to help someone in trouble.	☐	☐
3. It is sometimes hard for me to go on with my work if I am not encouraged.	☐	☐
4. I have never intensely disliked someone.	☐	☐
5. On occasion I have had doubts about my ability to succeed in life.	☐	☐
6. I sometimes feel resentful when I don't get my way.	☐	☐
7. I am always careful about my manner of dress.	☐	☐
8. My table manners at home are as good as when I eat out in a restaurant.	☐	☐
9. If I could get into a movie without paying and be sure I was not seen I would probably do it.	☐	☐
10. On a few occasions, I have given up doing something because I thought too little of my ability.	☐	☐
11. I like to gossip at times.	☐	☐
12. There have been times when I felt like rebelling against people in authority even though I knew they were right.	☐	☐
13. No matter who I'm talking to, I'm always a good listener.	☐	☐
14. I can remember 'playing sick' to get out of something.	☐	☐
15. There have been occasions when I took advantage of someone.	☐	☐
16. I'm always willing to admit it when I make a mistake.	☐	☐
17. I always try to practise what I preach.	☐	☐
18. I don't find it particularly difficult to get along with loud-mouthed, obnoxious people.	☐	☐
19. I sometimes try to get even rather than forgive and forget.	☐	☐
20. When I don't know something I don't at all mind admitting it.	☐	☐
21. I am always courteous, even to people who are disagreeable.	☐	☐
22. At times I have really insisted on having things my own way.	☐	☐
23. There have been occasions when I felt like smashing things.	☐	☐
24. I would never think of letting someone else be punished for my wrong-doing.	☐	☐
25. I never resent being asked to return a favour.	☐	☐
26. I have never been irked when people expressed ideas very different from my own.	☐	☐
27. I never make a long trip without checking the safety of my car.	☐	☐
28. There have been times when I was quite jealous of the good fortune of others.	☐	☐
29. I have almost never felt the urge to tell someone off.	☐	☐
30. I am sometimes irritated by people who ask favours of me.	☐	☐
31. I have never felt that I was punished without cause.	☐	☐
32. I sometimes think when people have a misfortune they only get what they deserve.	☐	☐
33. I have never deliberately said something that hurt someone's feelings.	☐	☐

B.iii: Marlowe-Crowne Social Desirability Scale: Short Form.

I will read a number of statements concerning personal attitudes and traits. For each item, decide whether the statement is **true** or **false** as it pertains to you personally.

	True	False
1. It is sometimes hard for me to go on with my work if I am not encouraged.	☐	☐
2. I sometimes feel resentful when I don't get my way.	☐	☐
3. On a few occasions, I have given up doing something because I thought too little of my ability.	☐	☐
4. There have been times when I felt like rebelling against people in authority even though I knew they were right.	☐	☐
5. No matter who I'm talking to, I'm always a good listener.	☐	☐
6. There have been occasions when I took advantage of someone.	☐	☐
7. I'm always willing to admit it when I make a mistake.	☐	☐
8. I sometimes try to get even rather than forgive and forget.	☐	☐
9. I am always courteous, even to people who are disagreeable.	☐	☐
10. I have never been irked when people expressed ideas very different from my own.	☐	☐
11. There have been times when I was quite jealous of the good fortune of others.	☐	☐
12. I am sometimes irritated by people who ask favours of me.	☐	☐
13. I have never deliberately said something that hurt someone's feelings.	☐	☐

Appendix C: Emotion and Control Cognitions Measures

C.i: Hospital Anxiety and Depression Scale.

Please indicate which comes closest to how you have been feeling in the past 7 days (tick the appropriate box) Or: Please tell me which comes closest to how you have been feeling in the past 7 days:

1. I feel tense or 'wound-up':
    1. Most of the time ☐
    2. A lot of the time ☐
    3. Time to time, occasionally ☐
    4. Not at all. ☐

2. I still enjoy the things I used to enjoy:
    1. Definitely as much ☐
    2. Not quite so much ☐
    3. Only a little ☐
    4. Hardly at all. ☐

3. I get a sort of frightened feeling as if something awful is about to happen:
    1. Very definitely and quite badly ☐
    2. Yes, but not too badly ☐
    3. A little but it doesn't worry me ☐
    4. Not at all. ☐

4. I can laugh and see the funny side of things:
    1. As much as I always could ☐
    2. Not quite so much now ☐
    3. Definitely not so much now ☐
    4. Not at all. ☐

5. Worrying thoughts go through my head:
    1. A great deal of the time ☐
    2. A lot of the time ☐
    3. From time to time but not too often ☐
    4. Only occasionally ☐

6. I feel cheerful:
    1. Not at all ☐
    2. Not often ☐
    3. Sometimes ☐
    4. Most of the time ☐

7. I can sit at ease and feel relaxed:
    1. Definitely ☐
    2. Usually ☐
    3. Not often ☐
    4. Not at all. ☐

8. I feel as if I am slowed down:
    1. Nearly all the time ☐
    2. Very often ☐
    3. Sometimes ☐
    4. Not at all. ☐

9. I get a sort of frightened feeling like 'butterflies' in the stomach:
    1. Not at all ☐
    2. Occasionally ☐
    3. Quite often ☐
    4. Very often ☐

10. I have lost interest in my appearance:
    1. Definitely ☐
    2. I don't take so much care as I should ☐
    3. I may not take quite as much care ☐
    4. I take just as much care as ever. ☐

11. I feel restless as if I have to be on the move:
    1. Very much indeed ☐
    2. Quite a lot ☐
    3. Not very much ☐
    4. Hardly at all. ☐

12. I look forward with enjoyment to things:
    1. As much as I ever did ☐
    2. Rather less than I used to ☐
    3. Definitely less than I used to ☐
    4. Hardly at all ☐

13. I get sudden feelings of panic:
    1. Very often indeed ☐
    2. Quite often ☐
    3. Not very often ☐
    4. Not at all. ☐

14. I can enjoy a good book or radio or TV programme:
    1. Often ☐
    2. Sometimes ☐
    3. Not often ☐
    4. Very seldom. ☐

C.ii: Generalised Self-Efficacy Scale.

Please circle the appropriate number in response to the following statements. Or: I will read a series of statements. For each one, please tell me the extent to which each statement applies to you; how true each statement is, from not at all true to exactly true.

	Not at all true	*Barely true*	*Moderately true*	*Exactly true*
1. I can always manage to solve difficult problems if I try hard enough.	1	2	3	4
2. If someone opposes me, I can find means and ways to get what I want.	1	2	3	4
3. It is easy for me to stick to my aims and accomplish my goals.	1	2	3	4
4. I am confident that I could deal efficiently with unexpected events.	1	2	3	4
5. Thanks to my resourcefulness, I know how to handle unforeseen situations.	1	2	3	4
6. I can solve most problems if I invest the necessary effort.	1	2	3	4
7. I can remain calm when facing difficulties because I can rely on my coping abilities.	1	2	3	4
8. When I am confronted with a problem, I can usually find several solutions.	1	2	3	4
9. If I am in a bind, I can usually think of something to do.	1	2	3	4
10. No matter what comes my way, I'm usually able to handle it.	1	2	3	4

C.iii: Perceived Behavioural Control over Walking Task.

1. Whether I do or do not perform well on the walking task is entirely up to me.

Strongly disagree *Strongly agree*

1 2 3 4 5 6 7

2. How much control do you feel you have over walking task performance?

No control *Complete control*

1 2 3 4 5 6 7

3. I would like to perform well on the walking task but I don't really know if I can.

Strongly disagree *Strongly agree*

1 2 3 4 5 6 7

4. I am confident that I could perform well on the walking task if I wanted to.

Strongly disagree *Strongly agree*

1 2 3 4 5 6 7

5. For me to perform well on the walking task is:

Difficult *Easy*

1 2 3 4 5 6 7

C.iv: Perceived Behavioural Control over 24-hour Activity.

1. Whether I am or am not very active in the next 24 hours is entirely up to me.

Strongly disagree *Strongly agree*

1 2 3 4 5 6 7

2. How much control do you feel you have over being very active in the next 24 hours?

No control *Complete control*

1 2 3 4 5 6 7

3. I would like to be very active in the next 24 hours but I don't really know if I can.

Strongly disagree *Strongly agree*

1 2 3 4 5 6 7

4. I am confident that I could be very active in the next 24 hours if I wanted to.

Strongly disagree *Strongly agree*

1 2 3 4 5 6 7

5. For me to be very active in the next 24 hours is:

Difficult *Easy*

1 2 3 4 5 6 7

C.v: Recovery Locus of Control.

If 1 is Strongly Agree and 5 is Strongly Disagree, tell me the extent to which you agree or disagree with the following statements.

1. How I manage in the future depends on me, not on what other people can do for me.
2. It's often better just to wait and see what happens.
3. It's what I do to help myself that's really going to make all the difference.
4. My own efforts are not very important, my recovery really depends on others.
5. Its up to me to make sure that I make the best recovery possible under the circumstances.

6. I have little or no control over my progress from now on.
7. My own contribution to my recovery doesn't amount to much.
8. Getting better now is a matter of my own determination rather than anything else.
9. It doesn't matter how much help you get - in the end it's your own efforts that count.

C.vi: Specific Perceived Control Items (PCI).

(PBC). If 1 is No Control at all and 5 is Complete Control, how much control do you feel you have over whether you...... over the next month?

(PBC). If 1 is Not at all Difficult and 5 is Extremely Difficult, how difficult will it be
for you to...... over the next month

(SE). If 1 is Not at all confident and 5 is Extremely Confident, how confident are you that you will ...... over the next month?

(LOC). If 1 is Strongly Agree and 5 is Strongly Disagree, do you think that it is entirely up
to you whether you......over the next month?

C.vii: Perceived Control over Activity Visual Analogue Scales.

Indicate, by marking the horizontal lines, the extent to which you agree with the statements below.

	Agree	*Neither*	***Disagree***
1. I am in control of being as active as I would wish to be.			

	Agree	*Neither*	***Disagree***
2. It is difficult to be as active as I would wish			

	Agree	*Neither*	***Disagree***
3. I am confident in my ability to be as active as I would wish			

C.viii: Mood Visual Analogue Scales.

Indicate, by placing a line on the scales, your answers to the questions below.
At this moment in time, to what extent are you feeling:

	Not at all	***Extremely***
1. happy	______________________	
2. energetic	______________________	
3. distressed	______________________	
4. content	______________________	
5. sluggish	______________________	
6. sad	______________________	
7. sleepy	______________________	

Appendix D: Other Measures

D.i: Intention: Walking Task.

1. How much effort do you intend to put into the walking task?

No effort at all … *As much effort as possible*

1 2 3 4 5 6 7

2. I intend to perform well on the walking task

Definitely do not … *Definitely do*

1 2 3 4 5 6 7

3. I plan to perform well on the walking task

Definitely do not … *Definitely do*

1 2 3 4 5 6 7

4. I would like to perform well on the walking task

Definitely yes … *Definitely no*

1 2 3 4 5 6 7

5. I want to perform well on the walking task

Strongly disagree … *Strongly agree*

1 2 3 4 5 6 7

6. I expect to perform well on the walking task

Unlikely *Likely*

1 2 3 4 5 6 7

7. How likely is it that you will perform well on the walking task?

Unlikely *Likely*

1 2 3 4 5 6 7

D.ii: Intention: 24-hour Activity:

1. How active do you intend to be over the next 24 hours?

Not at all active *As active as possible*

1 2 3 4 5 6 7

2. I intend to be very active over the next 24 hours.

Definitely do not *Definitely do*

1 2 3 4 5 6 7

3. I plan to be very active over the next 24 hours.

Definitely do not *Definitely do*

1 2 3 4 5 6 7

4. I would like to be very active over the next 24 hours.

Definitely yes *Definitely no*

1 2 3 4 5 6 7

5. I want to be very active over the next 24 hours.

Strongly disagree *Strongly agree*

1 2 3 4 5 6 7

6. I expect to be very active over the next 24 hours.

Unlikely *Likely*

1 2 3 4 5 6 7

7. How likely is it that you will be very active over the next 24 hours?

Unlikely *Likely*

1 2 3 4 5 6 7

D.iii: Whether could have walked further, if so how far.

After performing the walking test, some people think that they could actually have walked faster and therefore further than they did walk in the time allowed. Do you think that you could have walked further? Yes ☐ No ☐
If 'yes', how many more laps do you think you could have walked? (where one lap is a 'single', not 'return', journey between the walls) ….. laps.

Appendix E: Data

E.i: Pilot Study Frequency Tables: Self-reports.

Table i: Report of walking task distance (N = 13)

Distance (m)	Frequency
60	1
200	1
300	2
500	3
700	1
800	1
900	1
1000	2
1300	1

Table ii: Report of time completing questionnaires (N = 12) (2 participants without actual time records excluded)

Time (s)	Frequency
240	1
300	1
600	1
720	1
900	4
1200	3
1500	1

Table iii: Report of time lying (N = 13)

Time (s)	Frequency
60	1
120	11
180	1

Table iv: Report of time sitting (N = 12)

Time (s)	Frequency
60	1
120	9
180	1
270	1

Table v: Report of time standing (N = 13)

Time (s)	Frequency
60	2
120	9
180	2

Table vi: Report of walking task time (N = 1

Time (s)	Frequency
600	8
720	2
900	2

E.ii: Walking Difficulties Study Frequency Tables.

Tables show number of half hour periods for which activity is reported by the monitor as steps taken, or activity is reported by the participant as having spent time moving about, walking or performing vigorous activity. Number of half-hour blocks spent in the laboratory is also reported. Thus, total number of blocks should sum to 48.

S2P1

	Self-Report Yes	Self-Report No
Monitor Yes	9	14
Monitor No	3	19

Lab blocks: 3

Monitor Yes, Self-report No – Self-report Yes, Monitor No = 11

S2P2

	Self-Report Yes	Self-Report No
Monitor Yes	4	26
Monitor No	0	16

Lab blocks: 2

Monitor Yes, Self-report No – Self-report Yes, Monitor No = 26

S2P3

	Self-Report Yes	Self-Report No
Monitor Yes	25	4
Monitor No	1	15

Lab blocks: 3

Monitor Yes, Self-report No – Self-report Yes, Monitor No = 3

S2P5

	Self-Report Yes	Self-Report No
Monitor Yes	13	18
Monitor No	1	15

Lab blocks: 1
Monitor Yes, Self-report No – Self-report Yes, Monitor No = 17

S2P6

	Self-Report Yes	Self-Report No
Monitor Yes	2	12
Monitor No	4	27

Lab blocks: 3
Monitor Yes, Self-report No – Self-report Yes, Monitor No = 8

S2P11

	Self-Report Yes	Self-Report No
Monitor Yes	14	13
Monitor No	0	14

Lab blocks: 2
Monitor Yes, Self-report No – Self-report Yes, Monitor No = 13

S2P12

	Self-Report Yes	Self-Report No
Monitor Yes	10	24
Monitor No	0	12

Lab blocks: 2
Monitor Yes, Self-report No – Self-report Yes, Monitor No = 24

S2P14

	Self-Report Yes	Self-Report No
Monitor Yes	9	15
Monitor No	0	22

Lab blocks: 2
Monitor Yes, Self-report No – Self-report Yes, Monitor No = 15

S2P15

	Self-Report Yes	Self-Report No
Monitor Yes	3	15
Monitor No	0	28

Lab blocks: 2
Monitor Yes, Self-report No – Self-report Yes, Monitor No = 15

S2P16

	Self-Report Yes	Self-Report No
Monitor Yes	7	13
Monitor No	2	24

Lab blocks: 2
Monitor Yes, Self-report No – Self-report Yes, Monitor No = 11

S2P17

	Self-Report Yes	Self-Report No
Monitor Yes	2	14
Monitor No	0	29

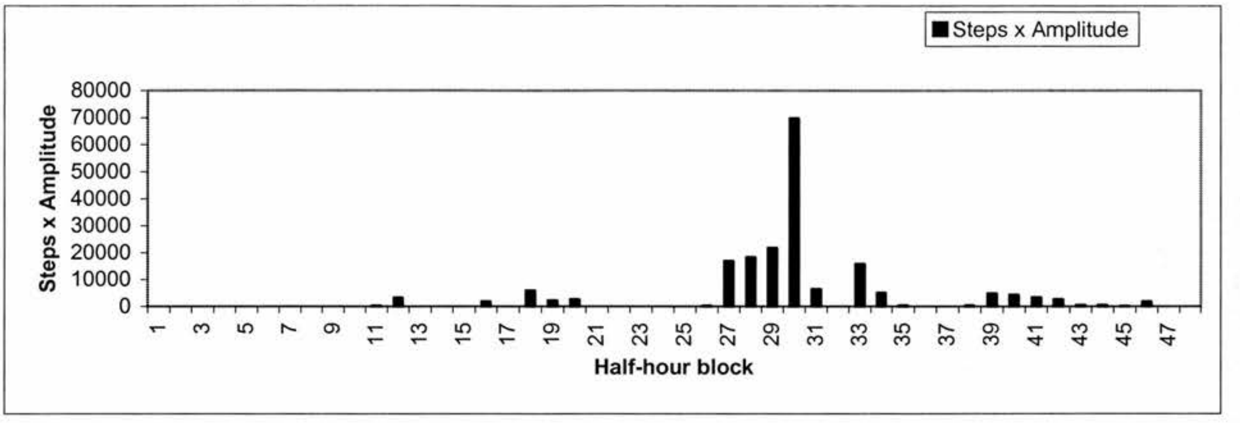
Lab blocks: 3
Monitor Yes, Self-report No – Self-report Yes, Monitor No = 14

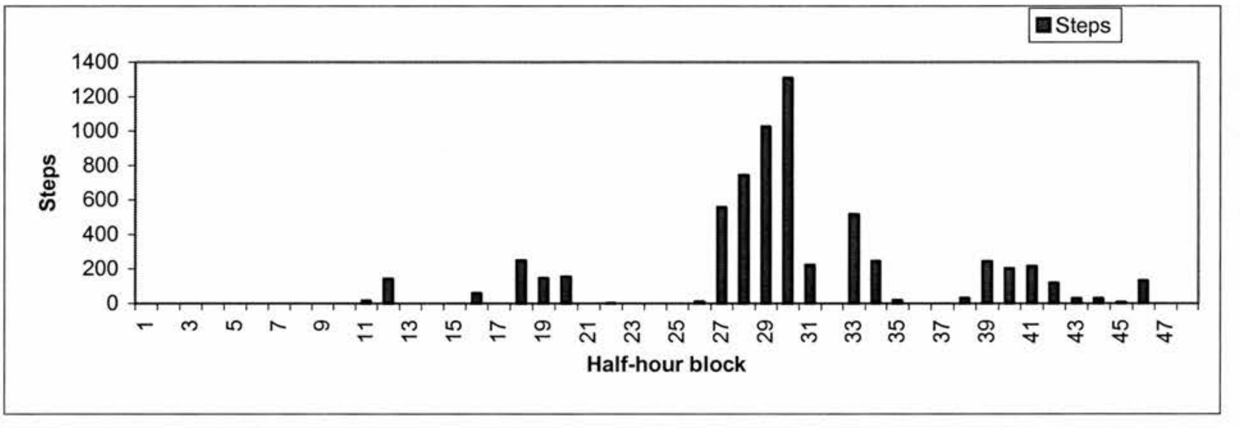
S2P19

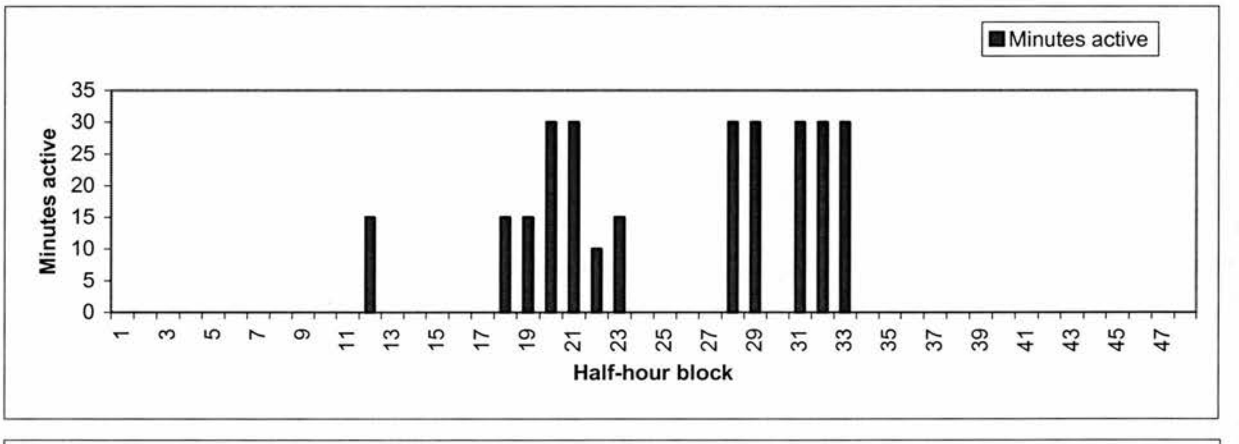
	Self-Report Yes	Self-Report No
Monitor Yes	7	23
Monitor No	0	15

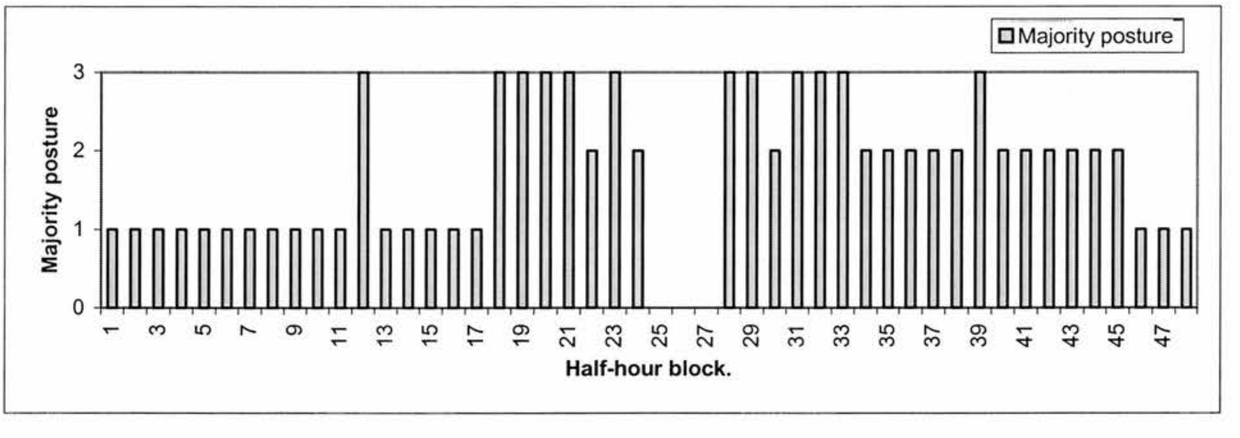
Lab blocks: 3
Monitor Yes, Self-report No – Self-report Yes, Monitor No = 23

E.iii: Walking Difficulties Study: 24-hour Data Charts. Block 1 = 00.00 (midnight) – 00.30.
S2P1.
Blocks in laboratory: 25, 26 and 27.



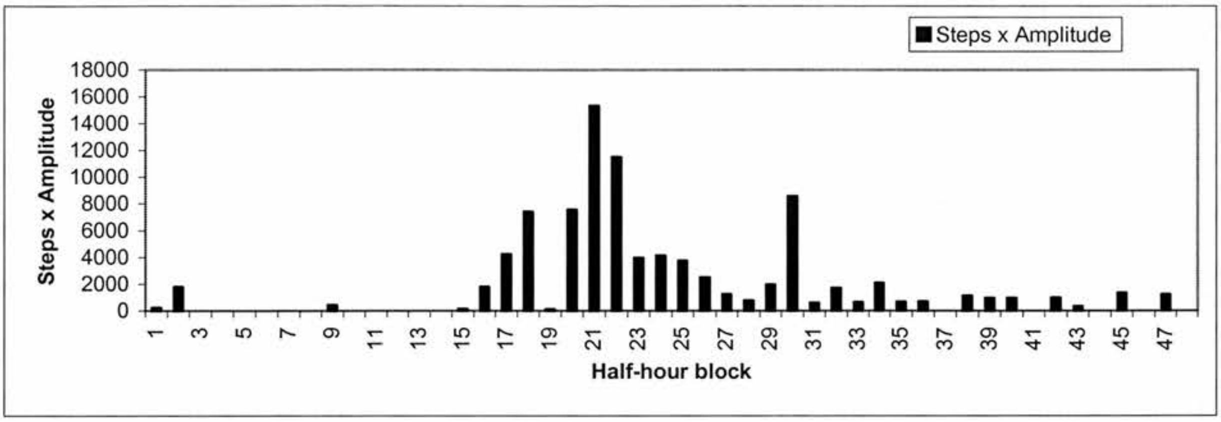


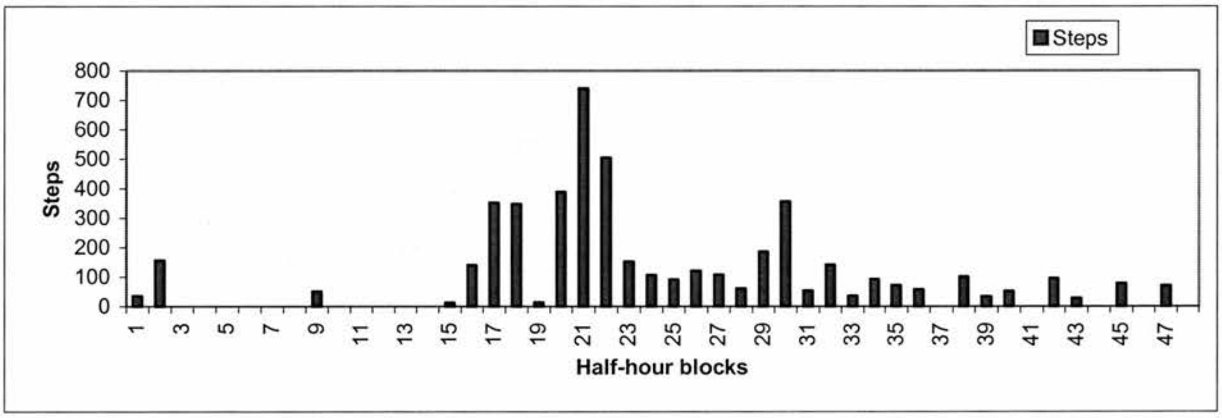


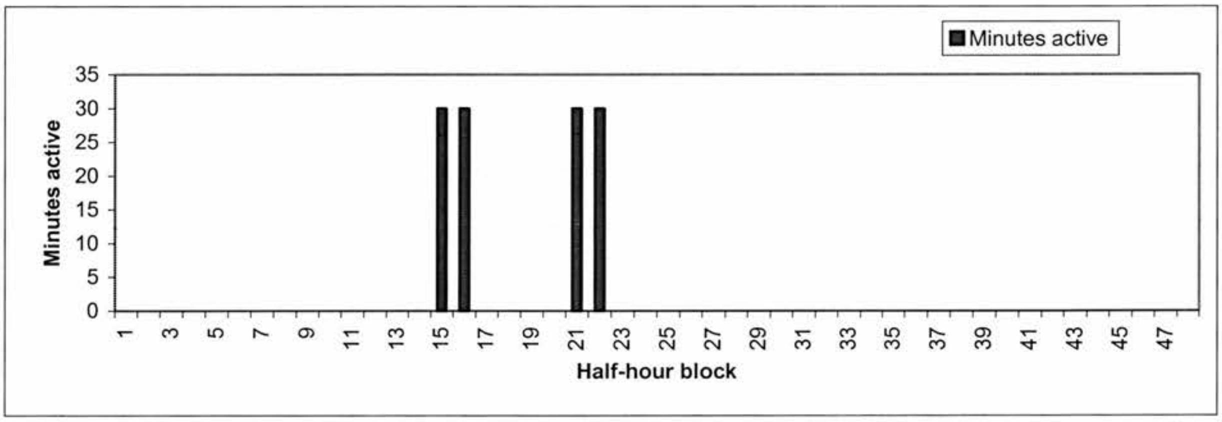


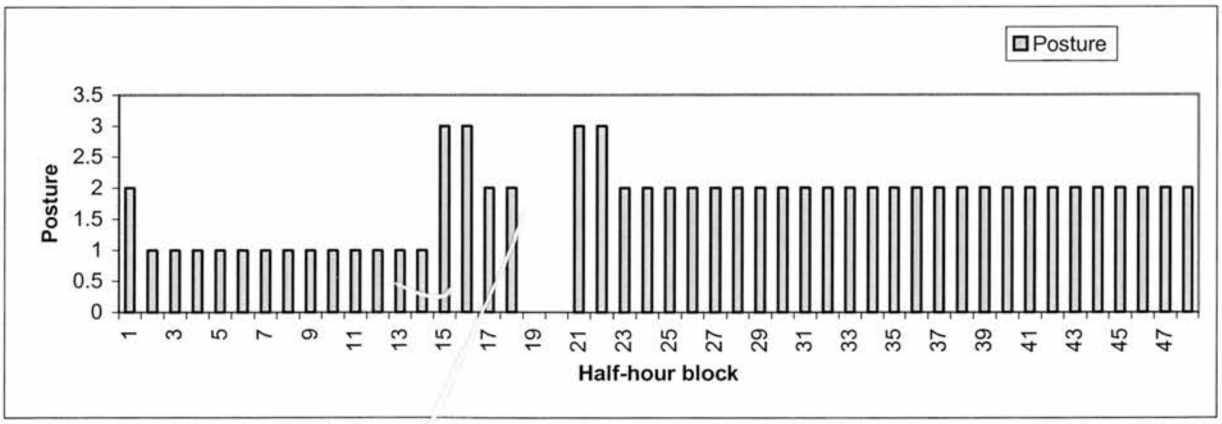
S2P2

Blocks in lab: 19 and 20. Monitor detached c. 4am – 7.20am (blocks 9 – 15).



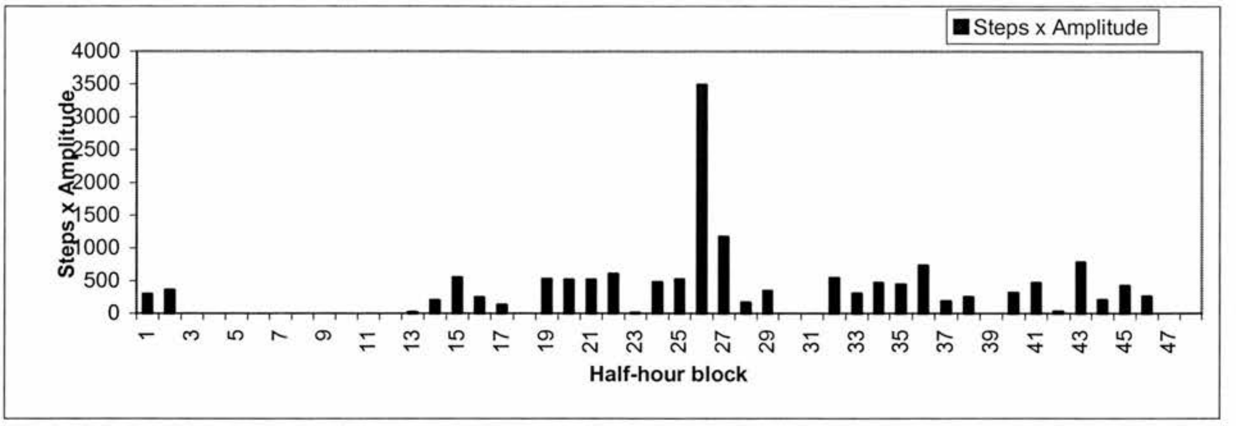


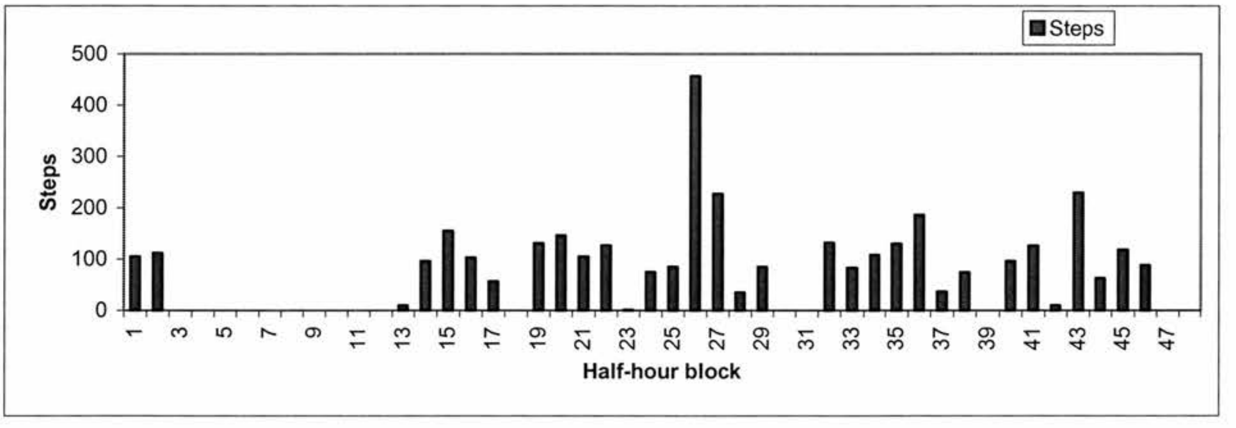


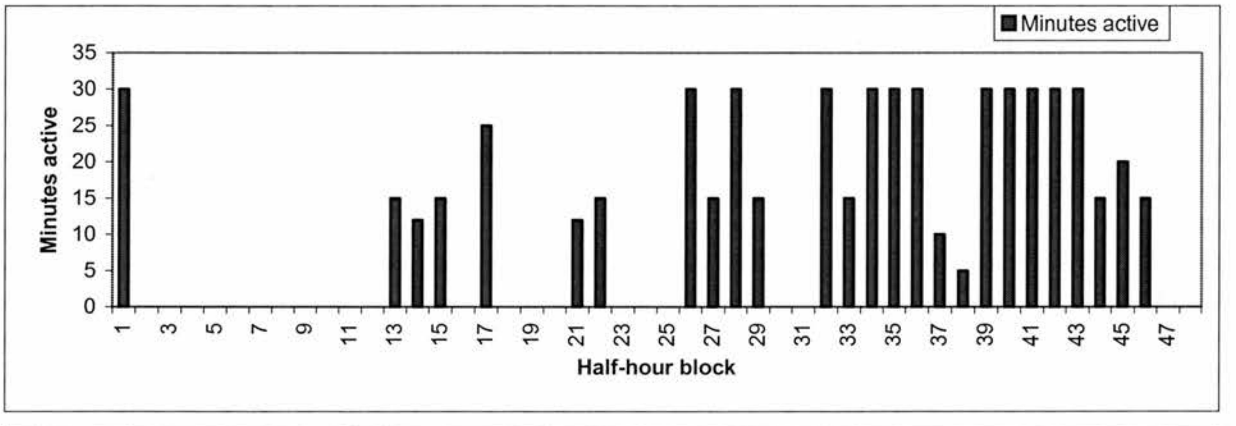


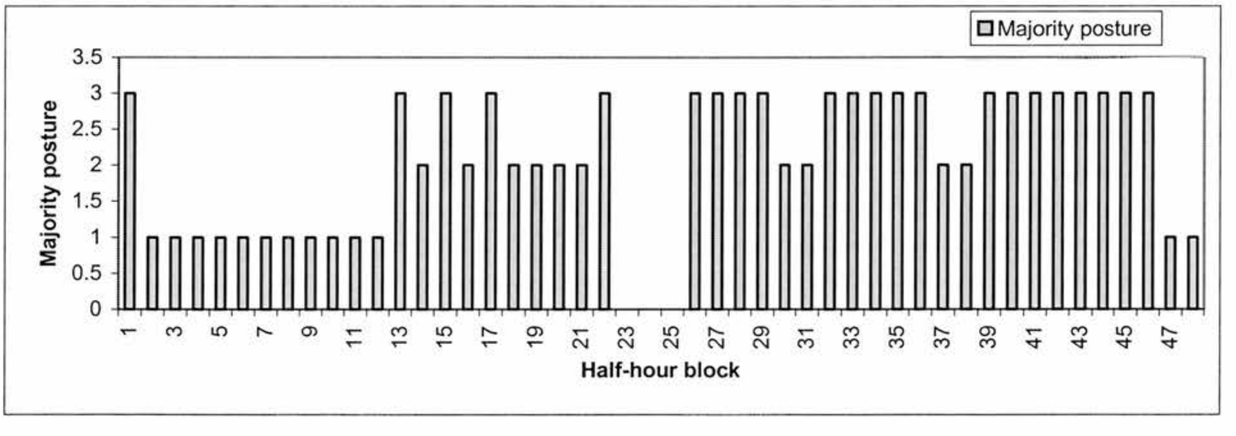
S2P3

Blocks in lab: 23, 24, 25.



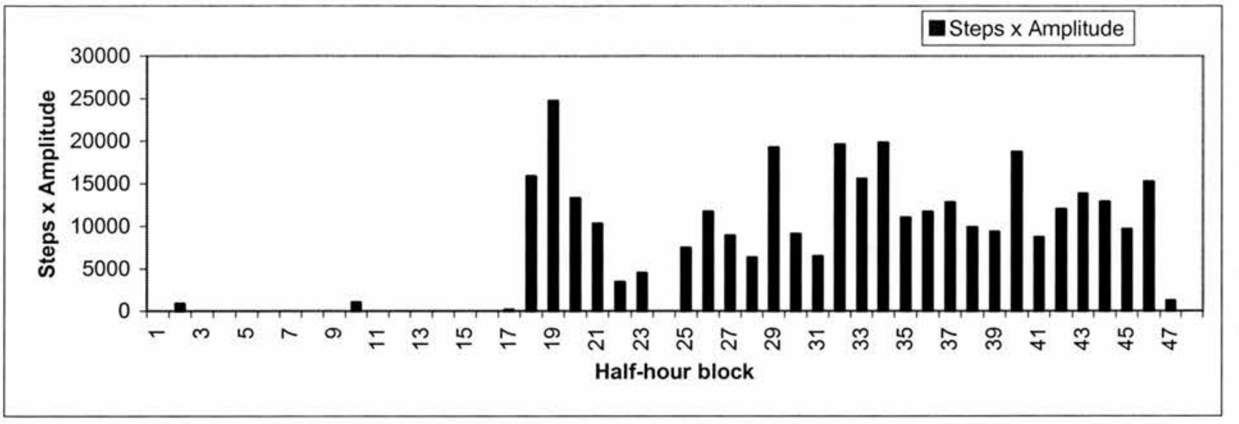


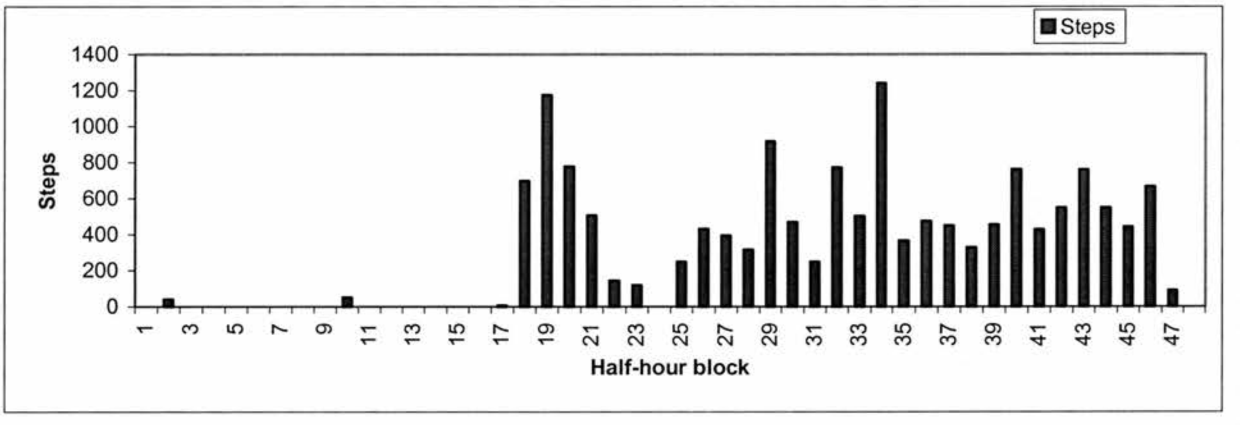


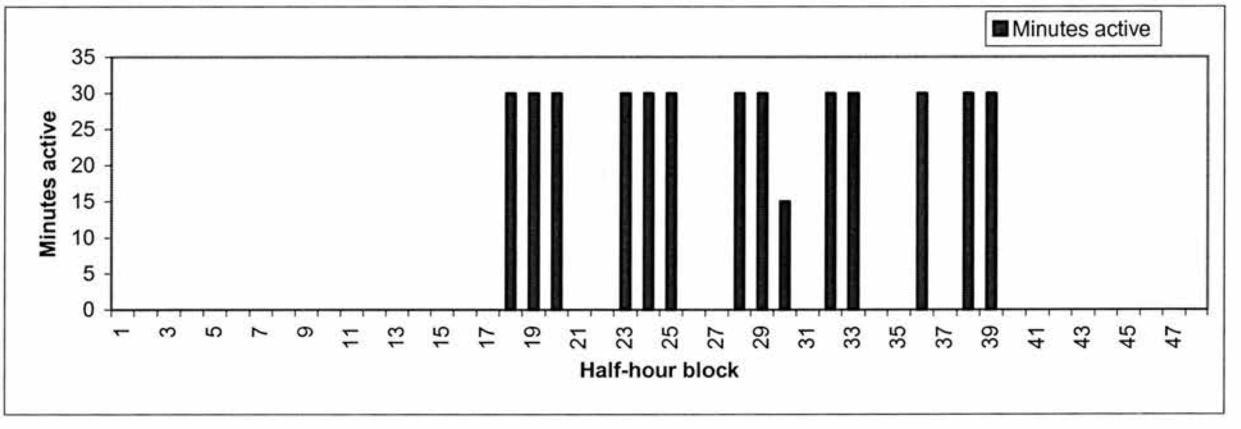


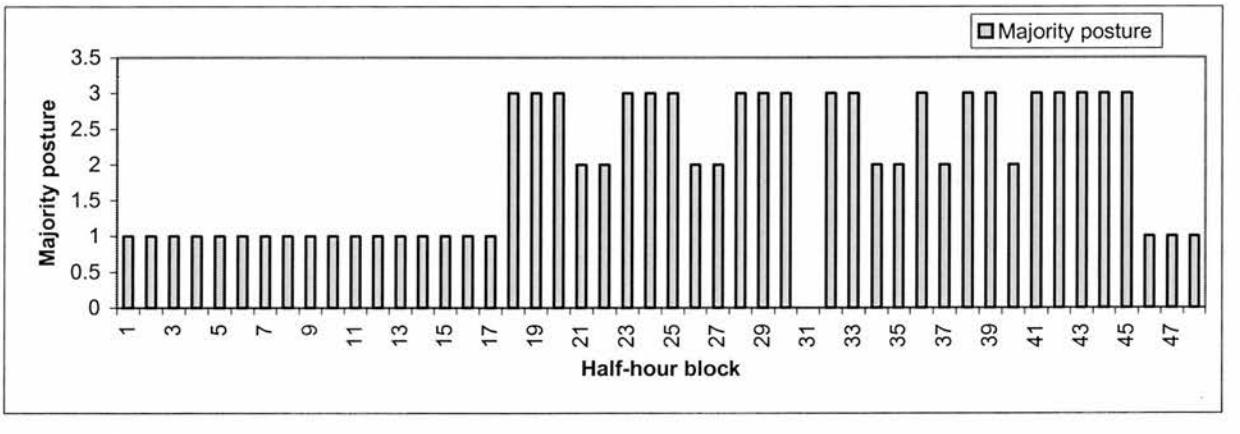
S2P5

Blocks in lab: half of 30 (start time: 2.45pm) and 31



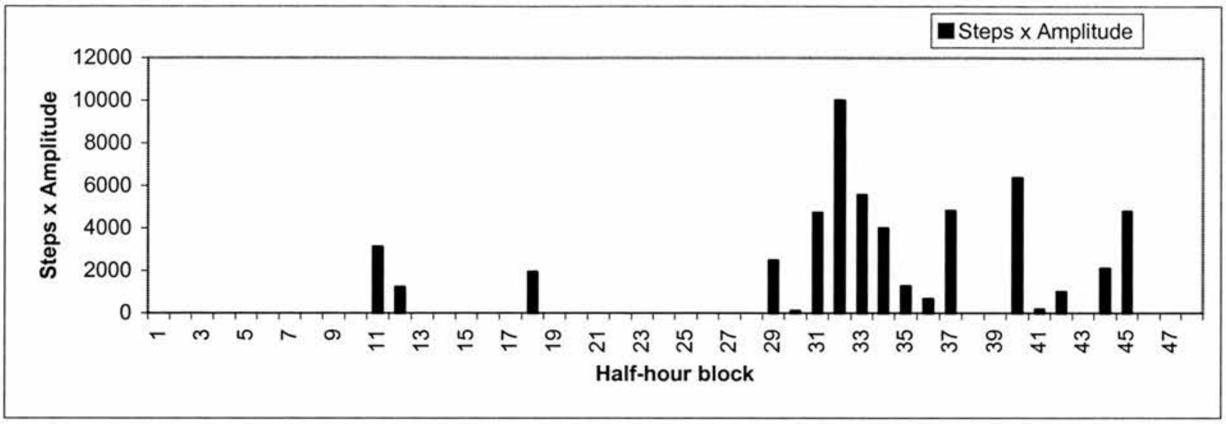


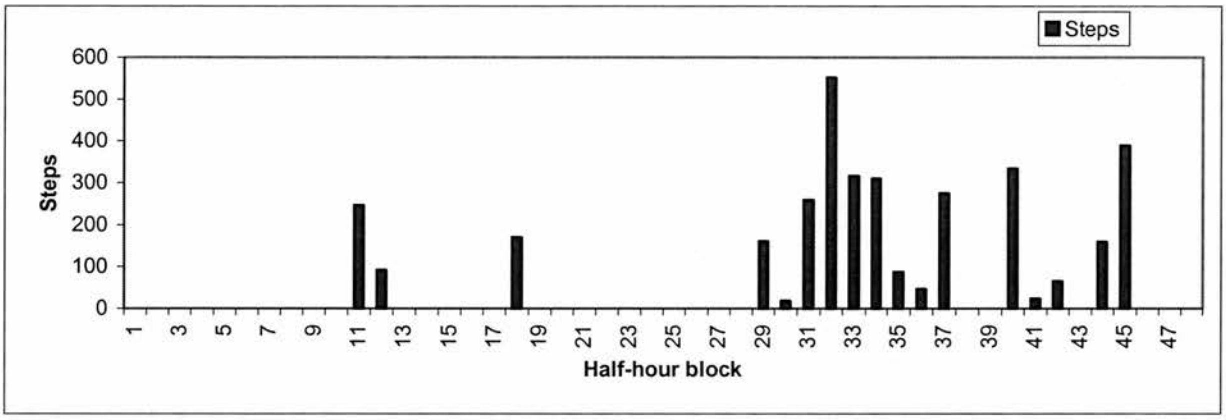


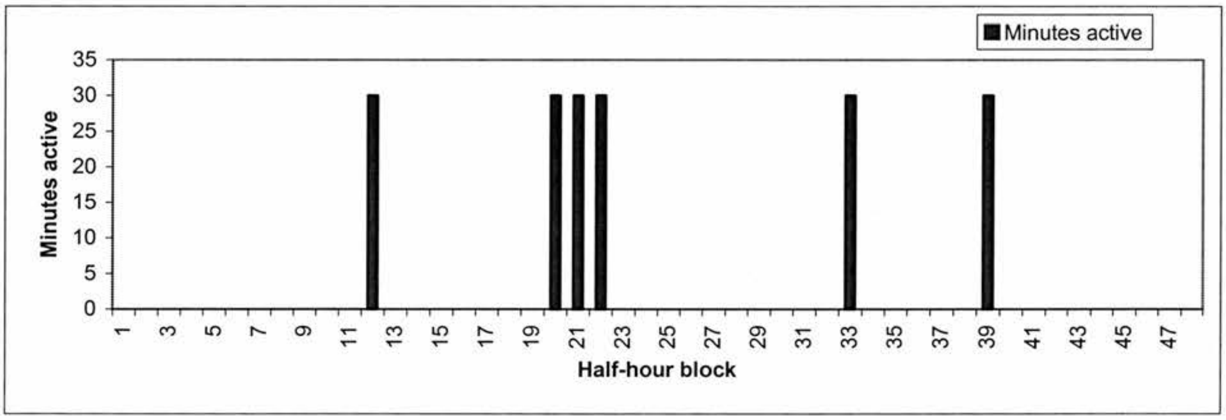


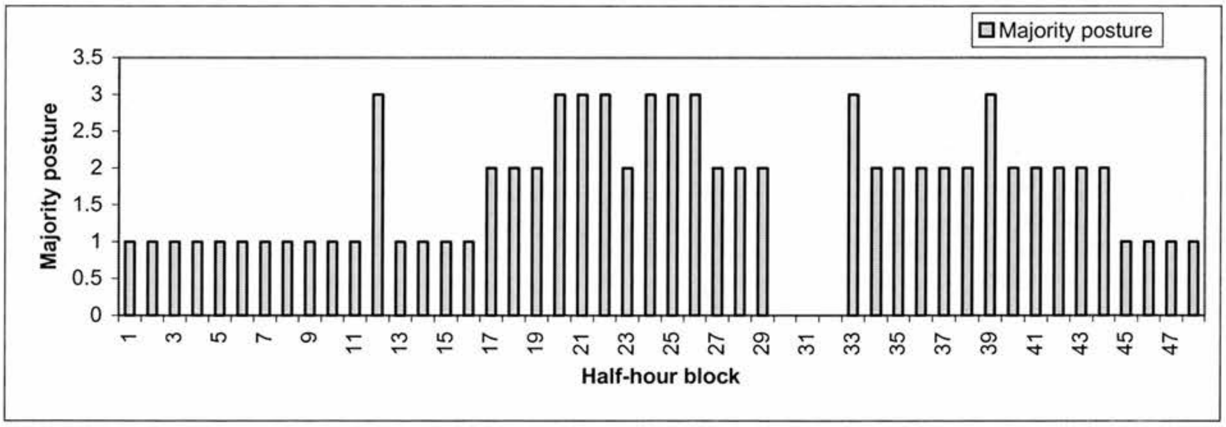
S2P6

Blocks in lab: 30, 31, 32.



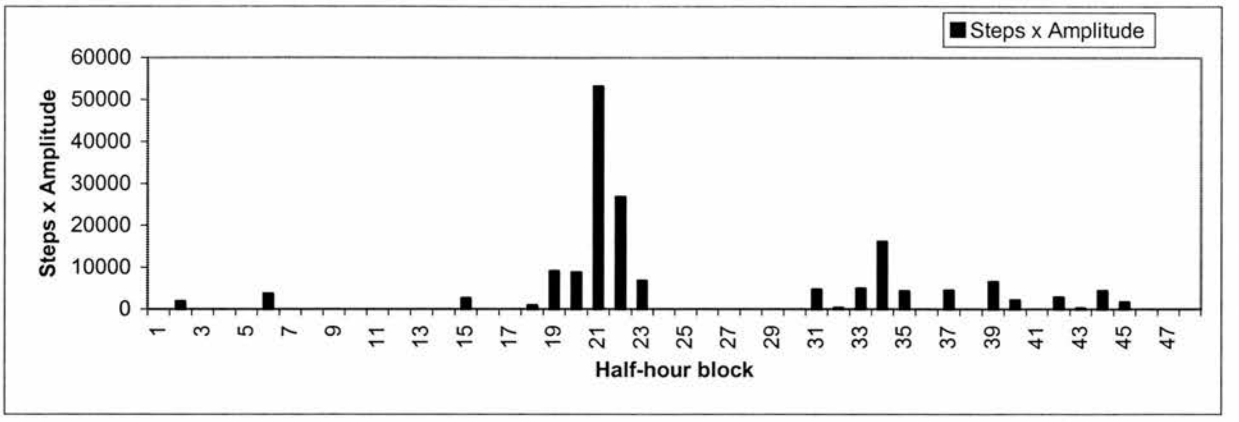


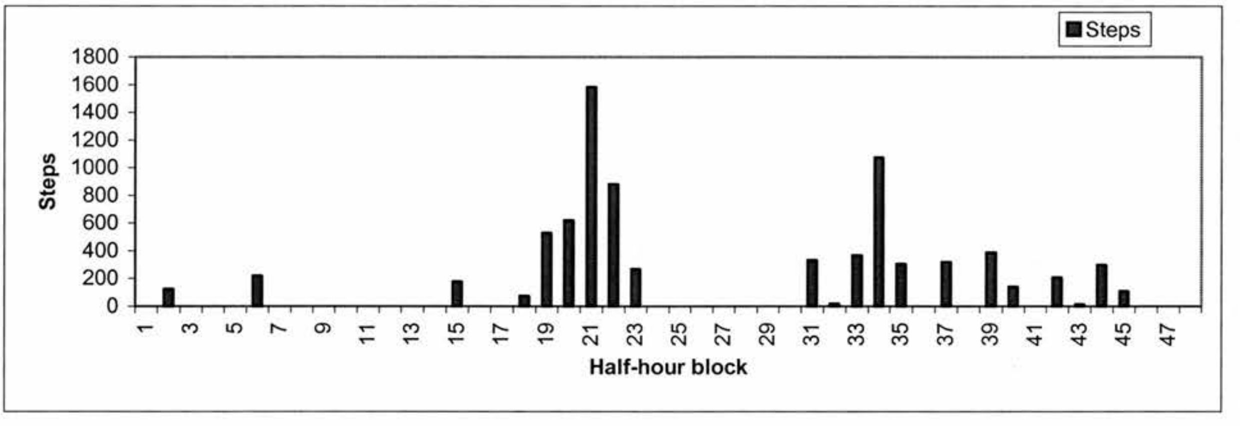


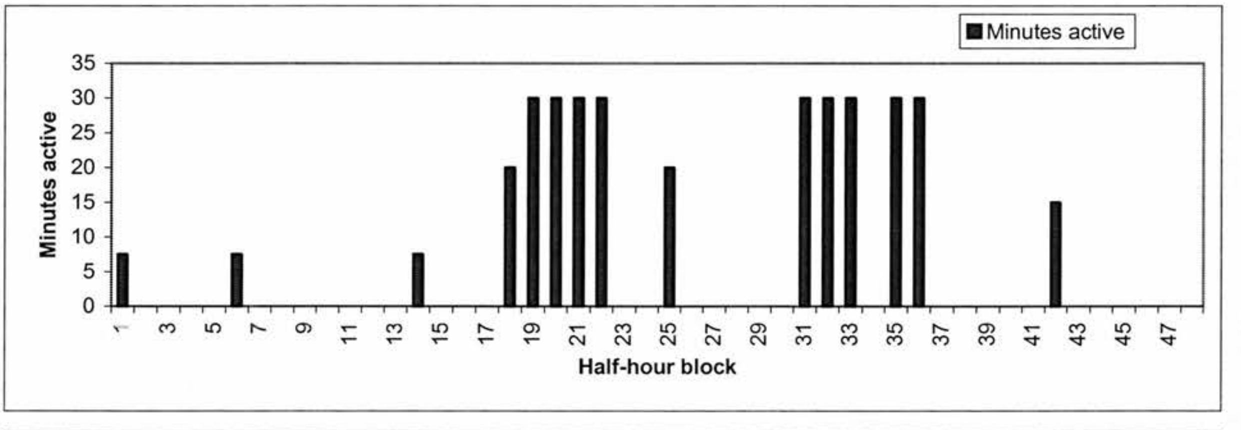


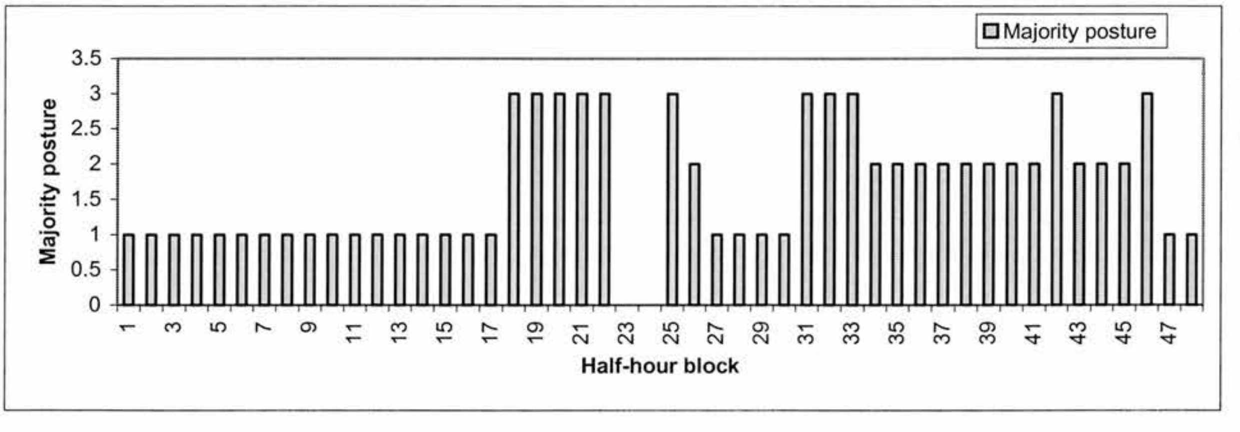
S2P8

Blocks in lab: 23 and 24. No monitor data for blocks 24-30.



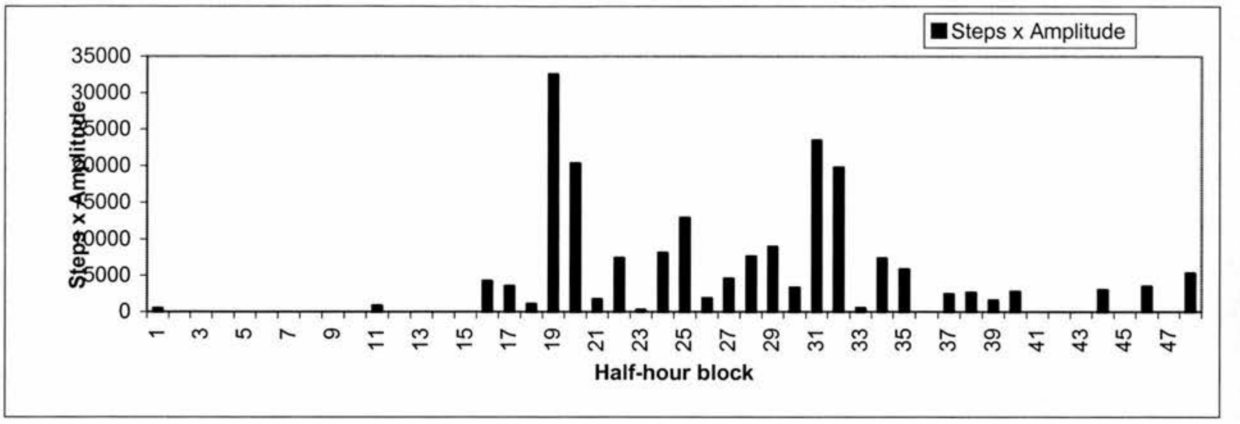


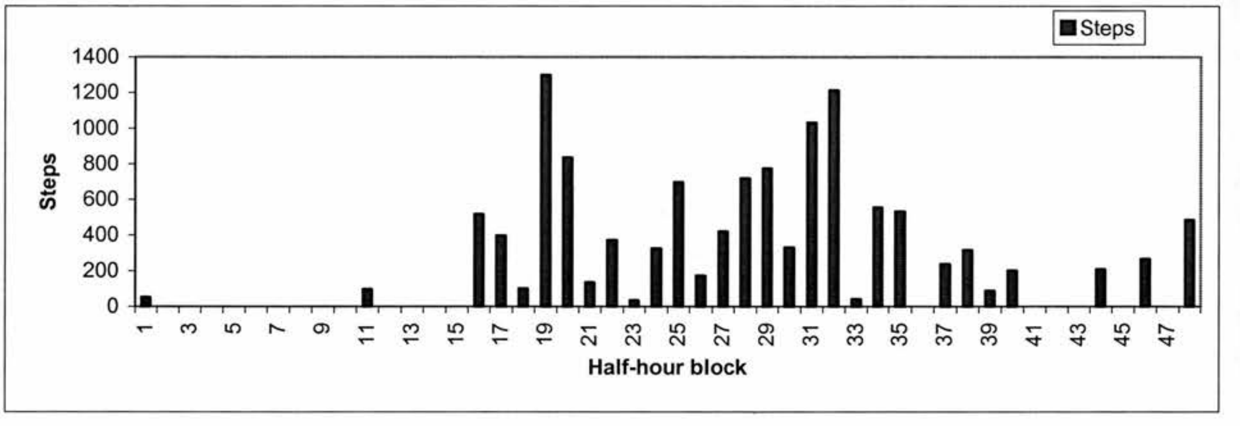


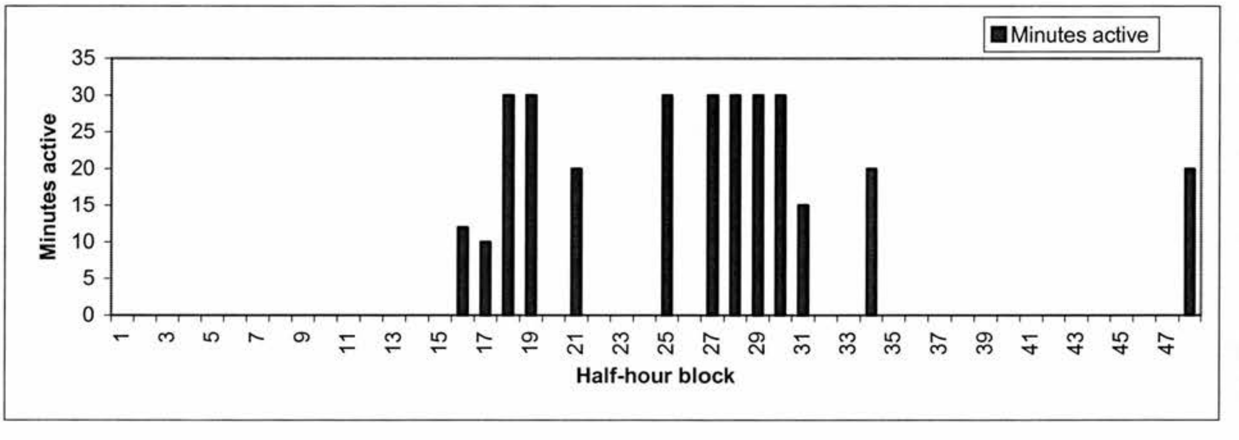


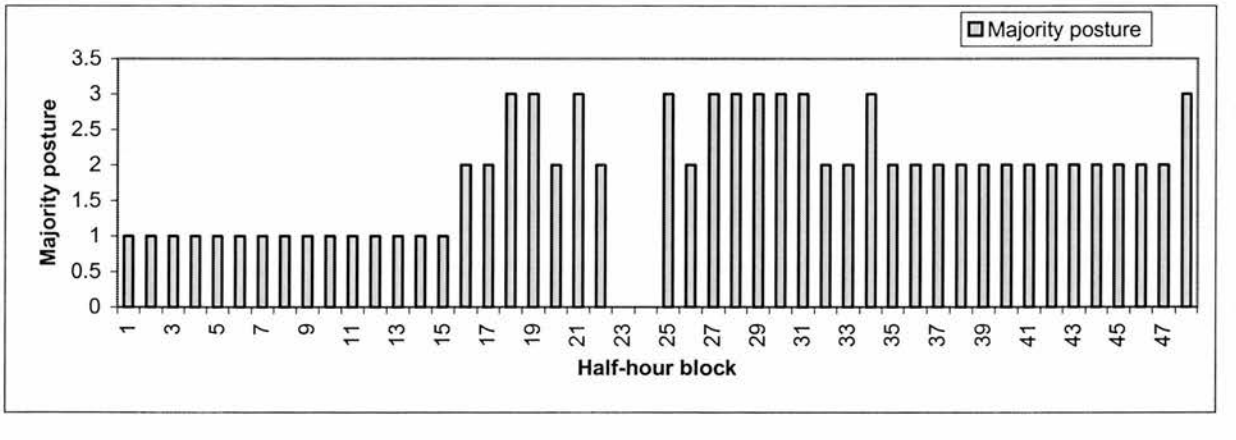
S2P11

Blocks in lab: 23 and 24.



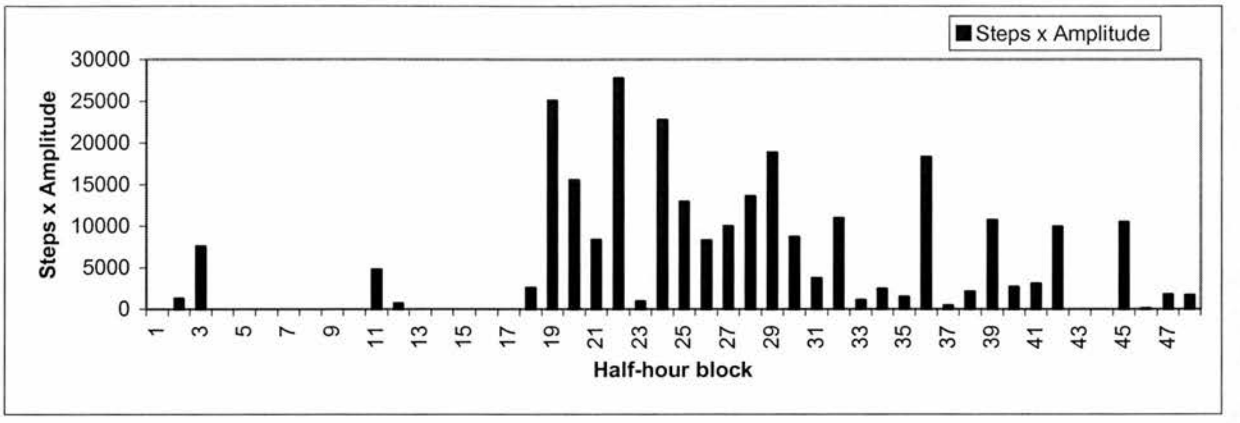


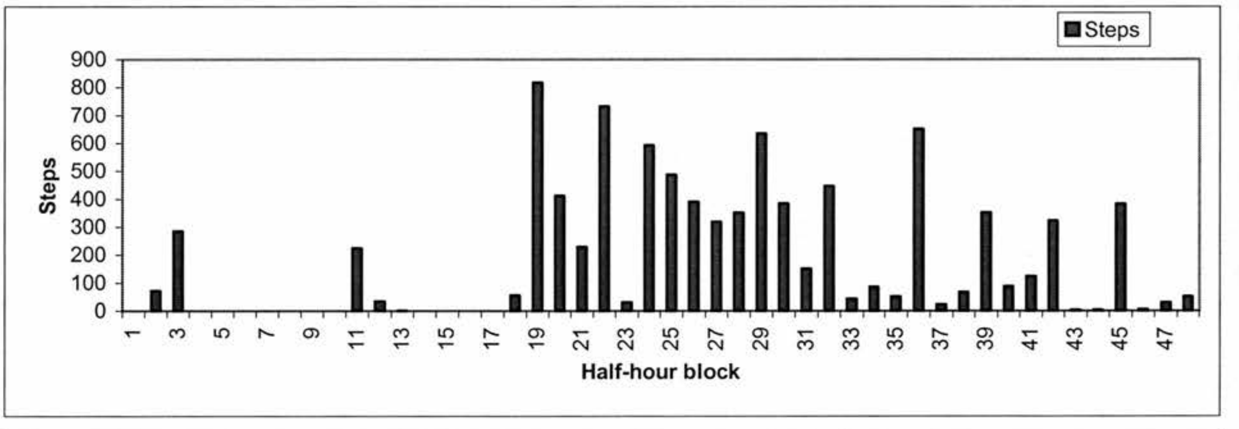


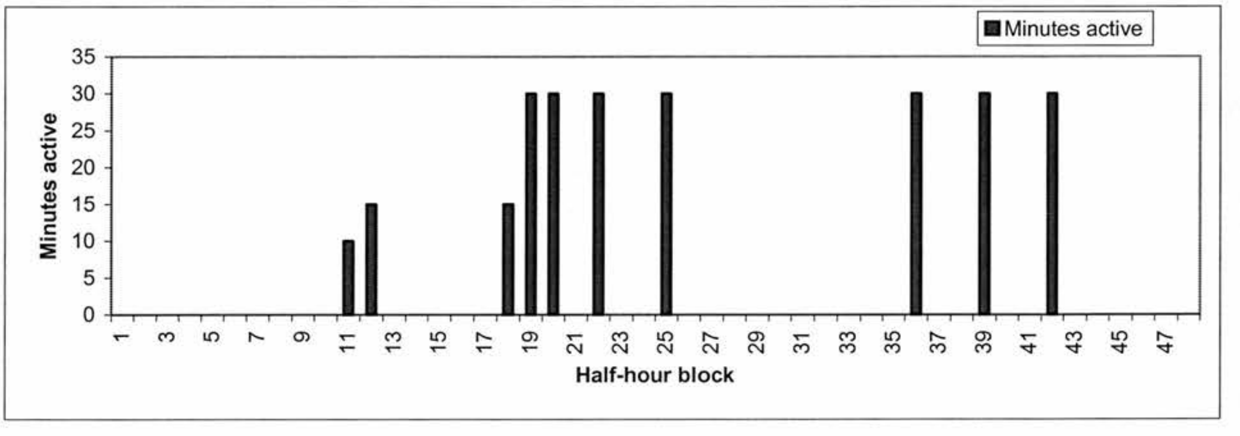


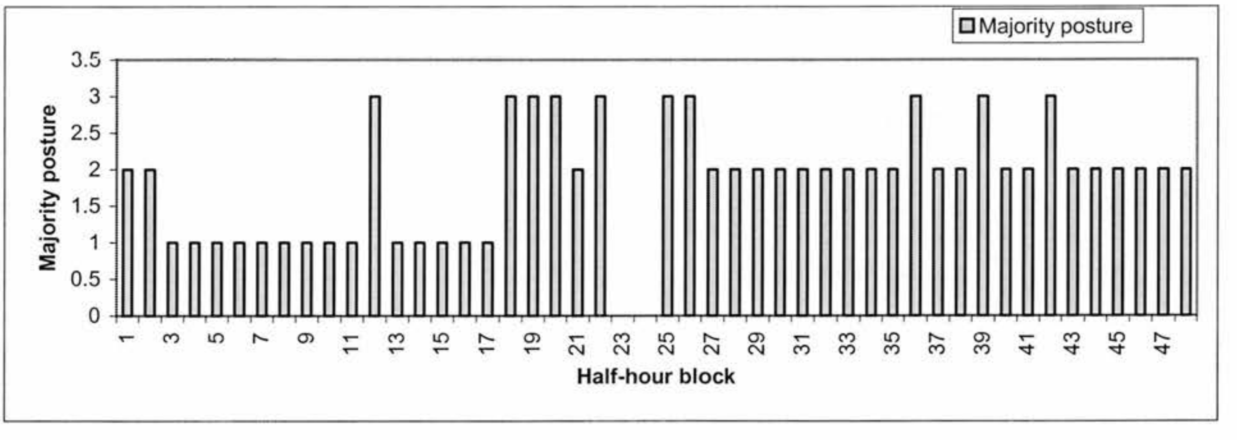
S2P12

Blocks in lab: 23 and 24.



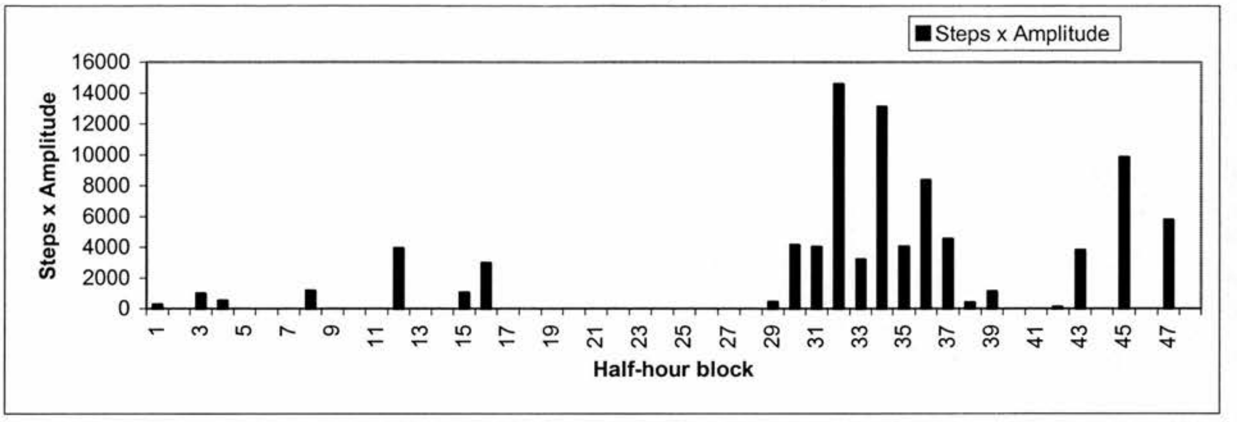


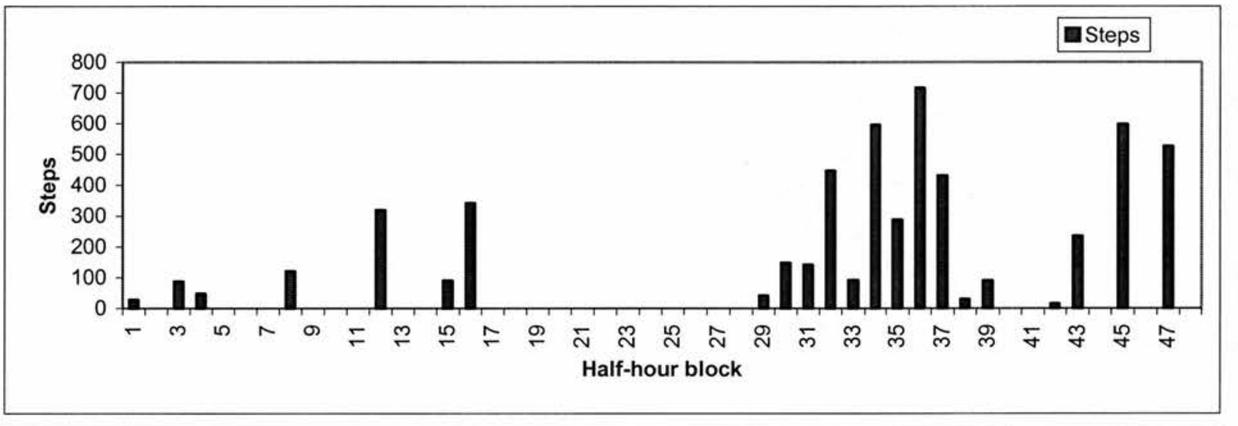


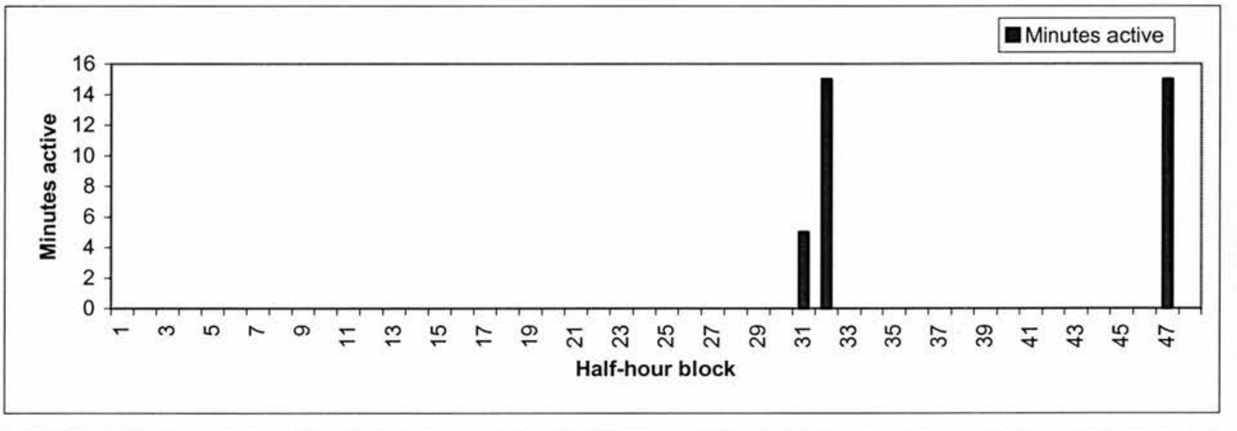


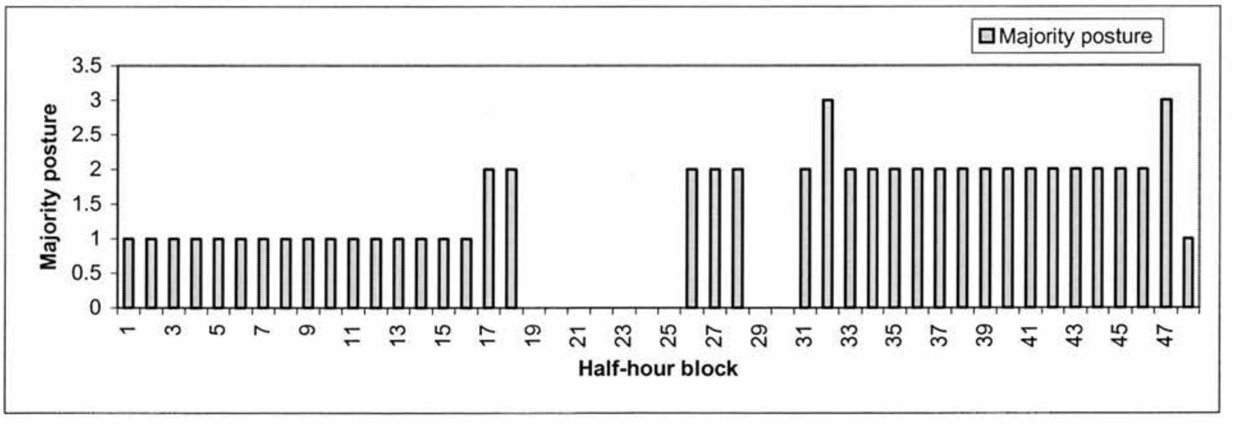
S2P13

Blocks in lab: 29 and 30. Monitor detached in blocks 33, 47 and 17. Not reattached after 17 → no data blocks 17-28. Self-report uninterpretable 19-25.

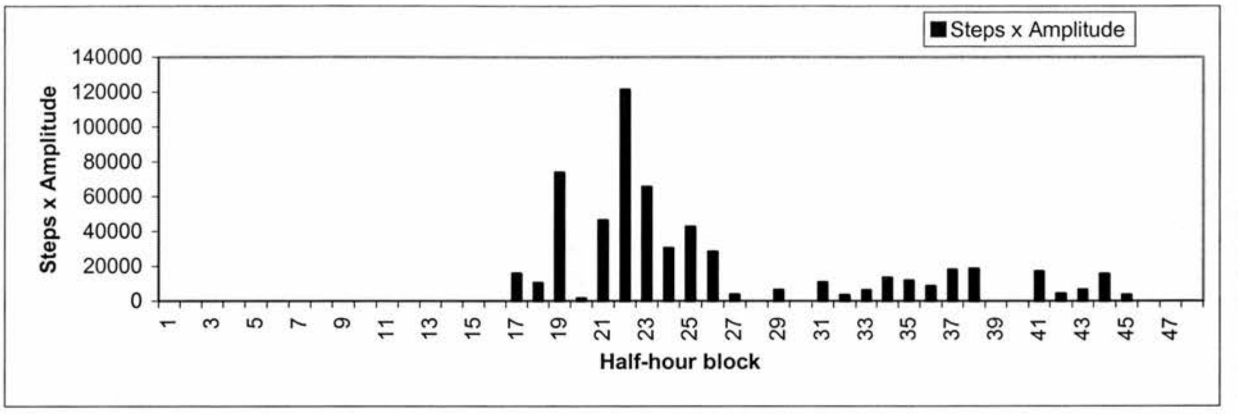


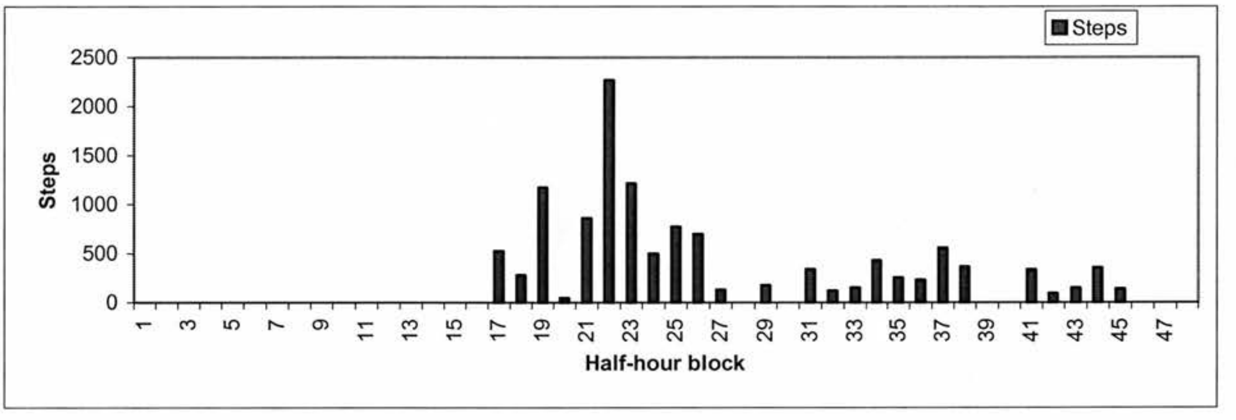


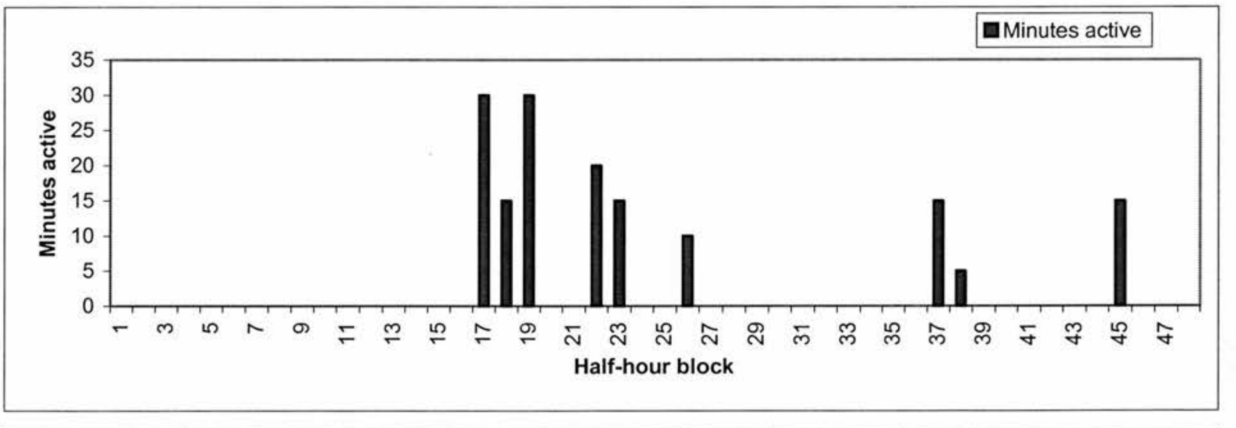


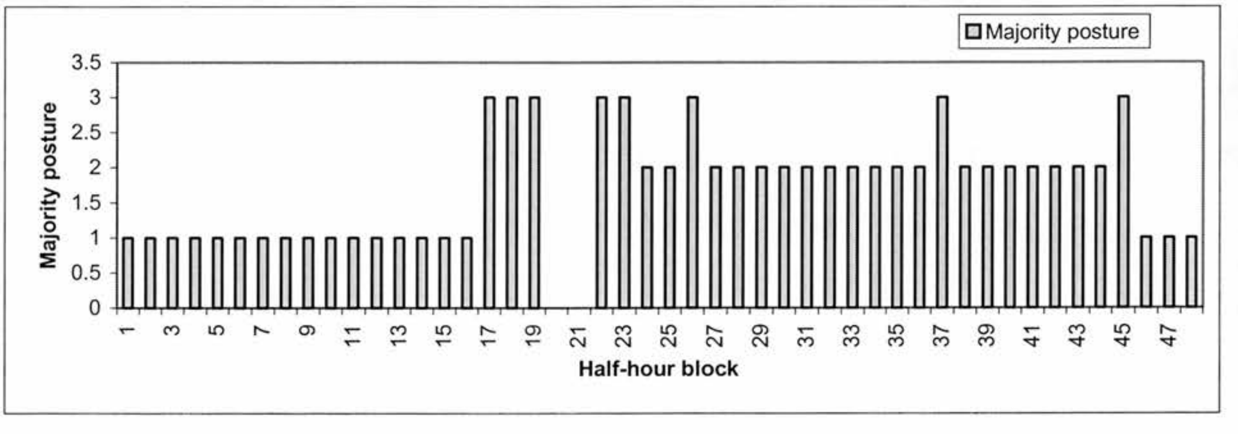


S2P14
Blocks in lab: 20 and 21



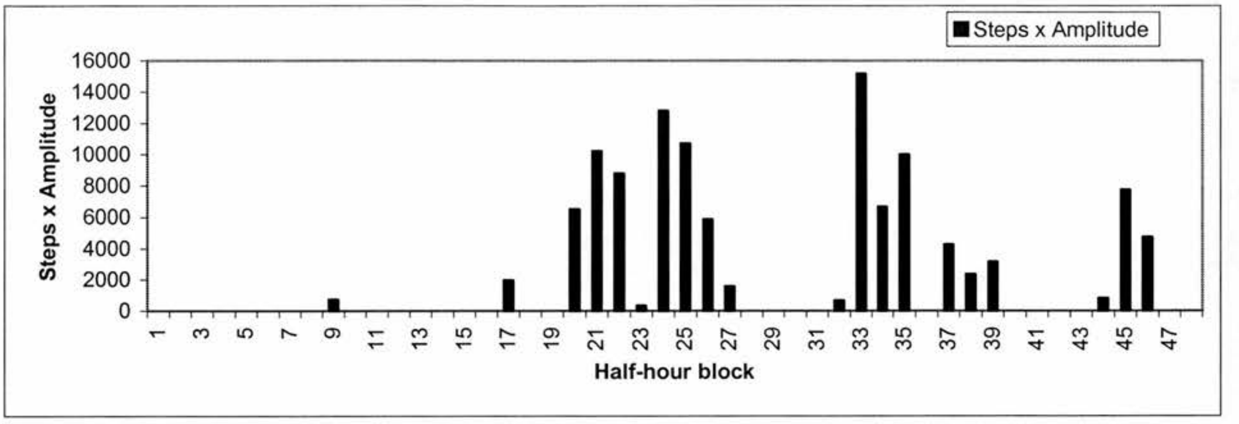


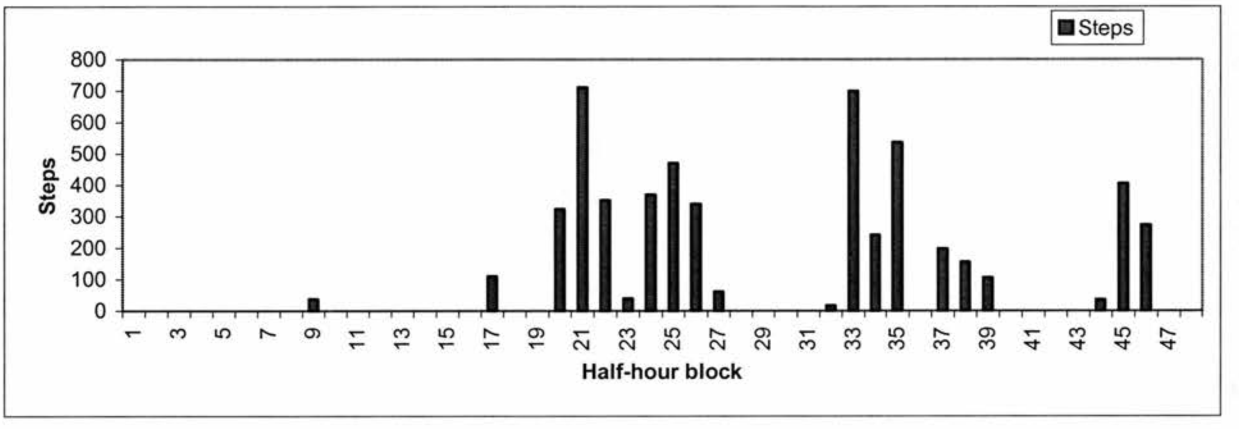


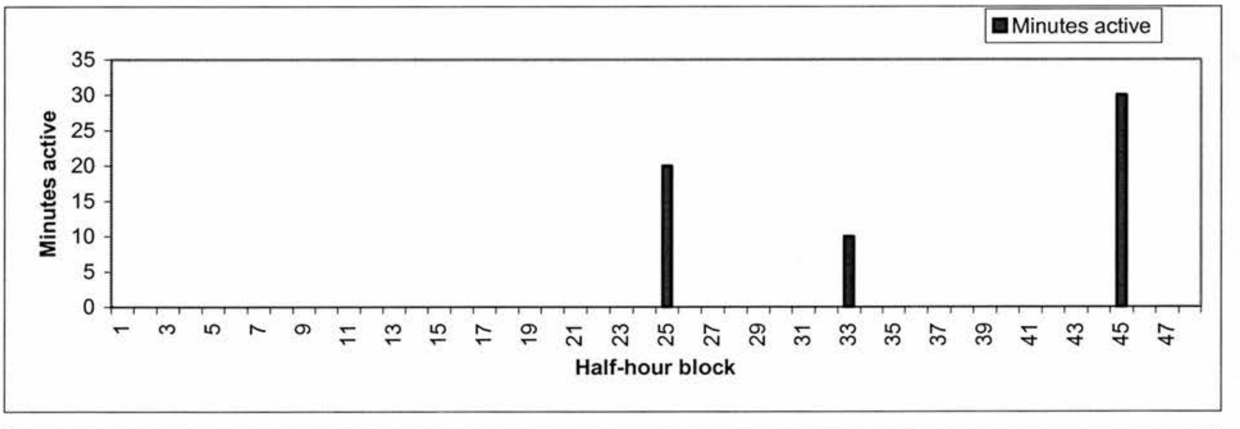


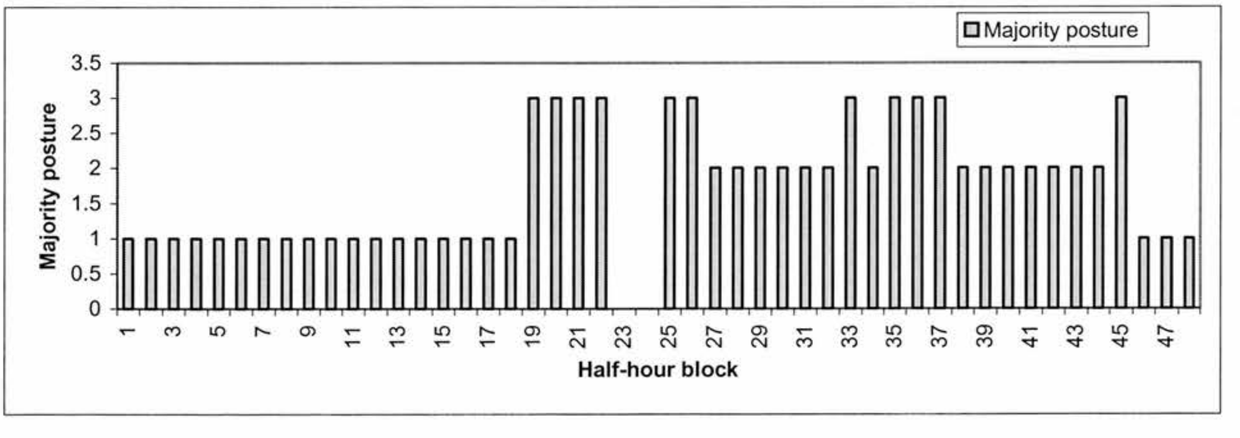
S2P15

Blocks in lab: 23 and 24.

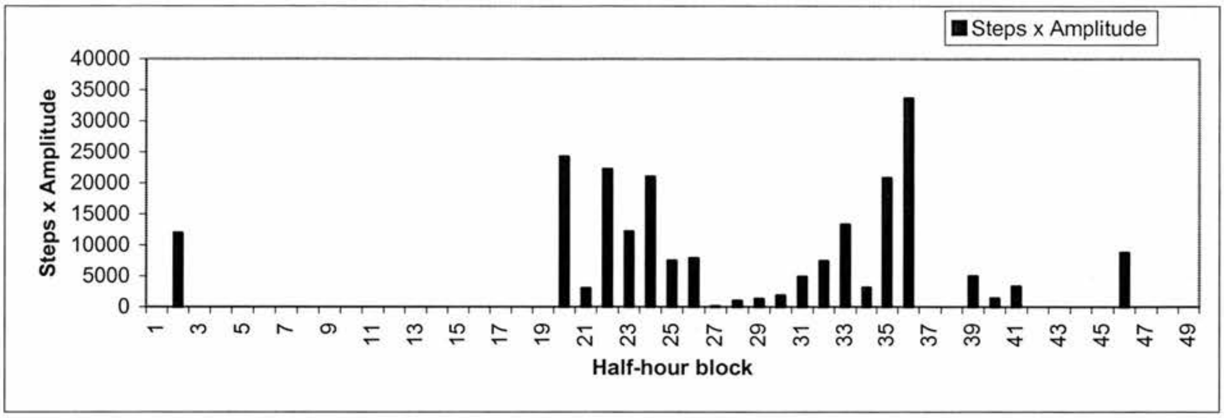


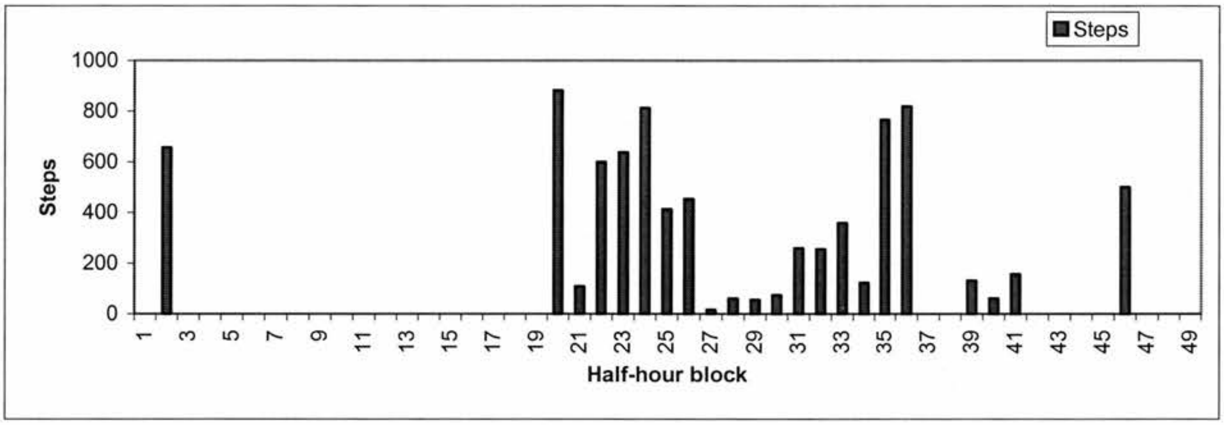


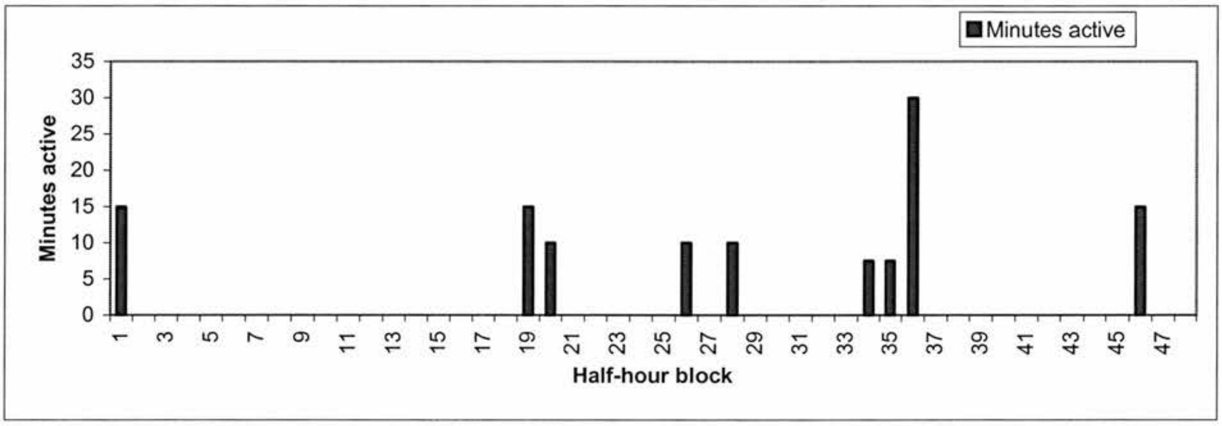


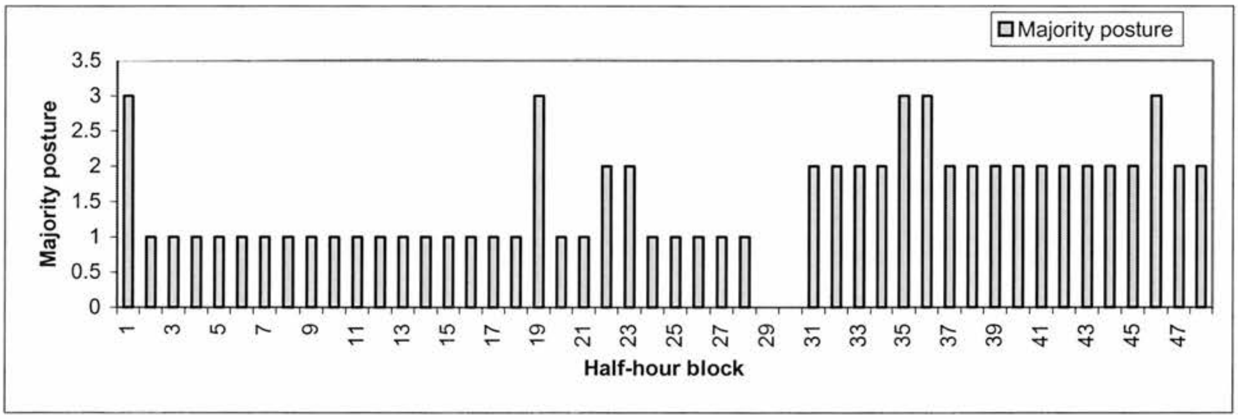


S2P16
Blocks in lab:29 and 30.



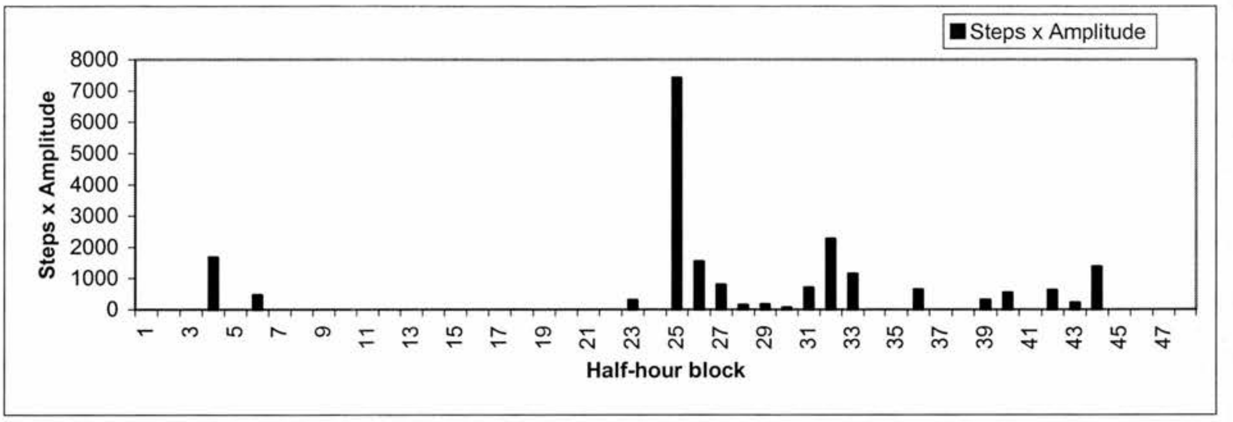


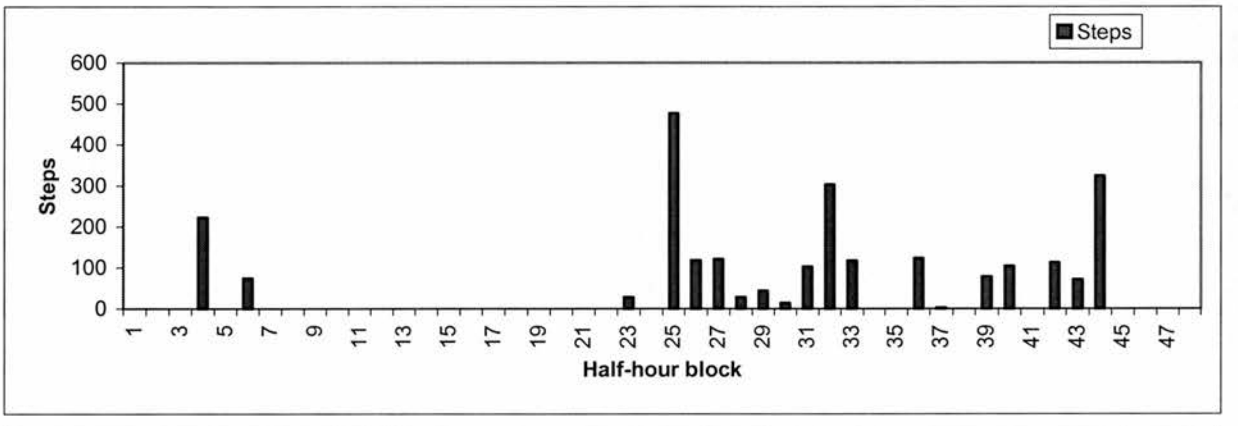


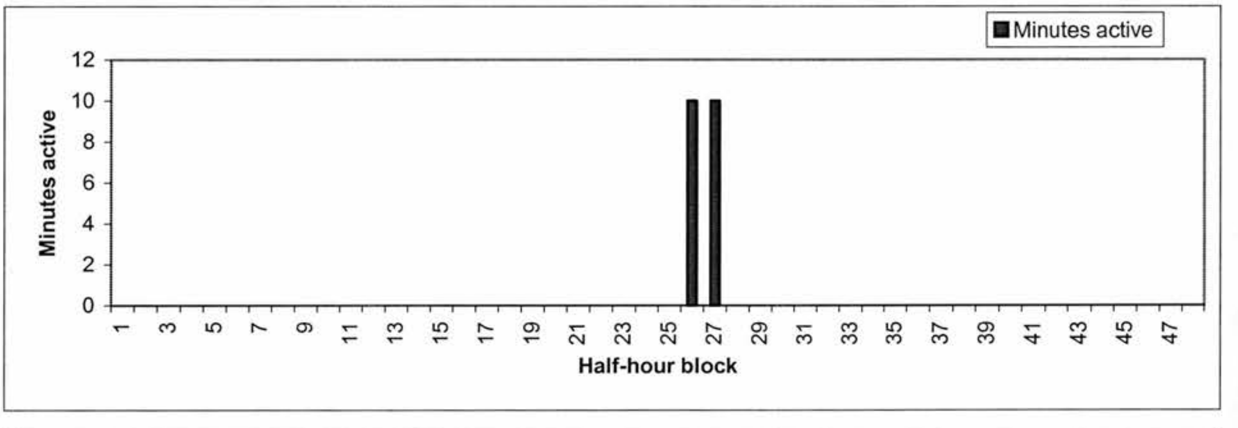


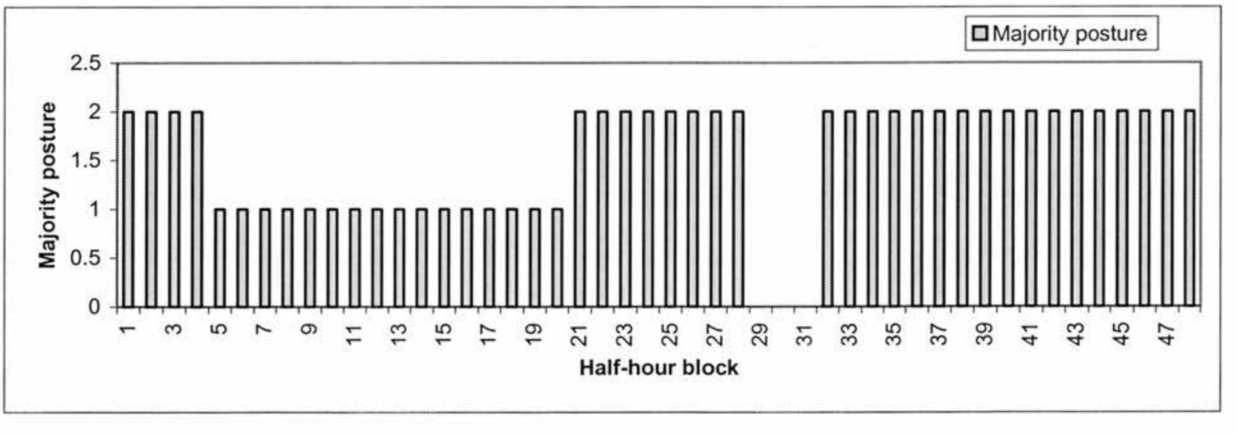
S2P17

Blocks in lab: 29, 30, 31.

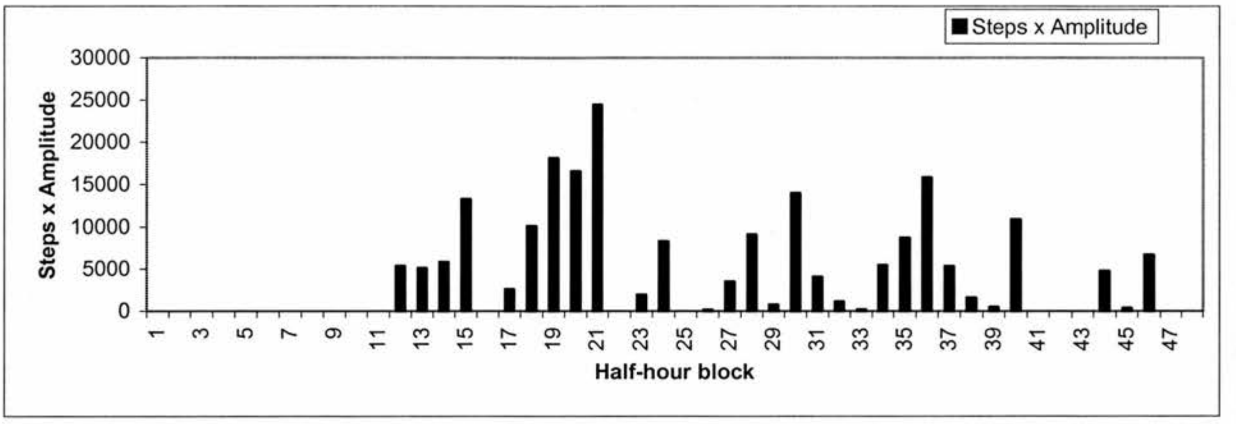


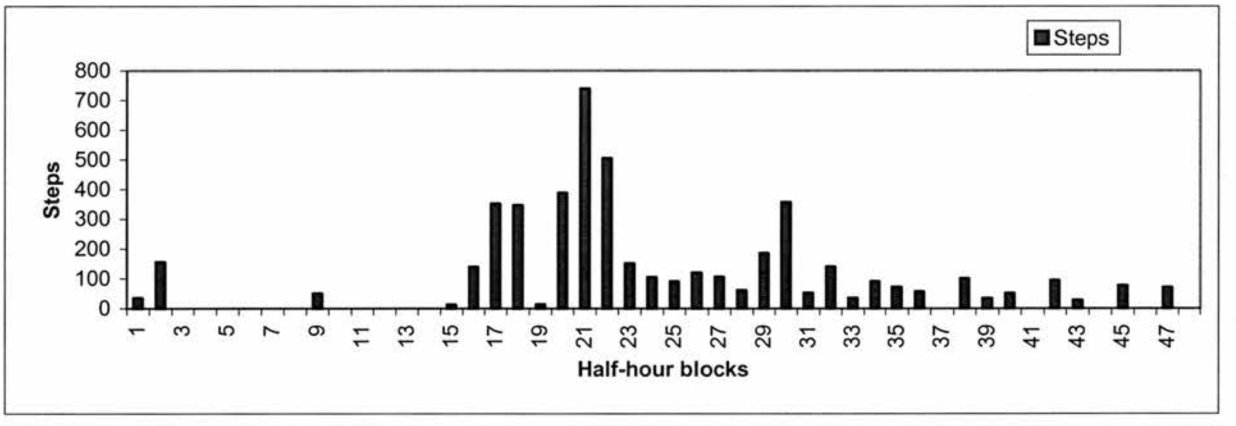


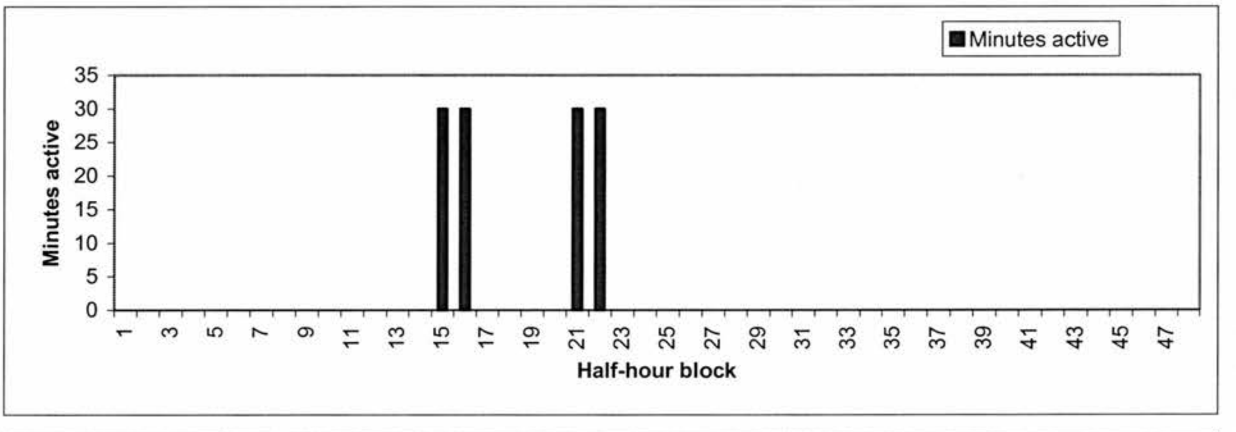


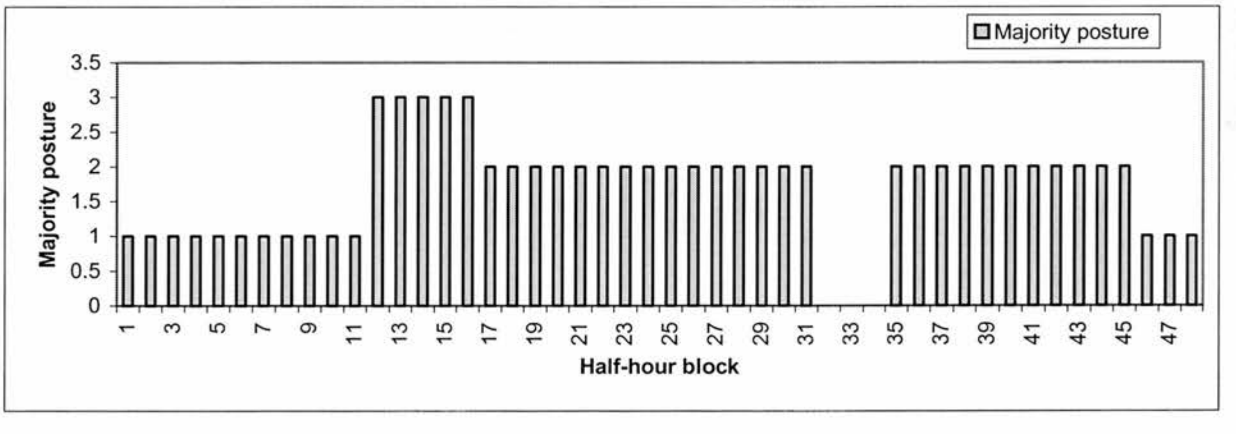


S2P19
Blocks in lab: 32, 33 and 34.









E.iv: Stroke Study: Baseline Comparisons.

Measure	t	p	N (included)	N (excluded)
P age	**-.45**	**.66**	**101**	**51**
C age	-.44	.66	98	46
P1flp	**-.24**	**.81**	**101**	**51**
C1flp	1.48	.14	101	50
P1dep	**.35**	**.73**	**100**	**51**
P1anx	-.44	.66	100	51
P1hads	-.07	.94	100	51
C1dep	-.86	.39	100	51
C1anx	-.57	.57	100	51
C1hads	-.77	.44	100	51
P1RLOC	**-1.14**	**.26**	**101**	**46**
P1PBC	-.35	.73	89	41
P1PCI	-.89	.38	89	40
TP1I	-.19	.85	90	41

Table shows results of independent t-tests comparing included and excluded participants. None of the differences were significant.

Appendix F: Adverts, Information Sheets and Consent Forms

F.i: Pilot Study: Advert

W A N T E D:

Healthy volunteers to take part in research looking at physical activity measurement.

You will be required to wear an activity monitor for 24 hours and to attend a laboratory session before and after this 24-hour period. These sessions will involve completing questionnaires and performing a fast walking task. In total, laboratory time will be about 1 hour.

£25 will be paid on completion of all tasks.

Please sign up below if you are interested in taking part.
For more information contact:
Rachael Powell, e-mail rkep@st-and.ac.uk, tel. 462077.

Name	Telephone number or e-mail address

F.ii: Walking Difficulties Study: Advert

Do you have difficulty walking?

We're looking into mobility problems at St Andrews University Psychology Department. If you have difficulty walking but can walk 100 yards we would like to ask you to take part in a study.

Please fill out and return one of the postcards below and you will be invited to meet the researcher to discuss what the study is about and whether you would like to participate. Or, for more information, contact:

Rachael Powell
School of Psychology
University of St Andrews
St Andrews
Fife
KY16 9JU

Phone: 01334 462052
e-mail: rkep@st-andrews.ac.uk

F.iii: Experimental Study: Advert

WANTED

Participants for a study investigating the relationship between beliefs and activity.

The study mainly involves the completion of questionnaires.

No longer than 1 hour should be required.
Participants are each paid £4.

Please e-mail rkep@st-andrews.ac.uk or phone 462052 (ext.2052) with times when you would be available, should you wish to take part.

Thank you!

Any questions? Contact Rachael Powell, e-mail rkep@st-and.ac.uk or phone (46)2052.

F.iv: Pilot Study: Information Sheet and Consent Form

Measuring Activity Levels as a Health Outcome:
Participant Information Sheet.

Thank you for volunteering to take part in this study.

This research aims to improve activity measurement methodology and to explore ways in which people's thoughts and feelings influence activity levels.

An activity monitor will be attached to you. It has 2 sensors, one of which is placed on the thigh, the other on the chest (it may be necessary for some participants to shave an area of the chest to prevent the sensor from moving and giving false readings). The sensors detect posture (whether you are lying, sitting or standing) and the number and intensity of steps taken.

There are 3 parts to the study:

1. You will be asked to complete some tasks in the laboratory, including 15-30 minutes of questionnaires and a walking task. The walking task consists of walking briskly, to cover as much distance as you can in 10-15 minutes.

2. You will be asked to wear the activity monitor for the next 24 hours and to return to the laboratory tomorrow. You should go about your ordinary activities as much as possible, but please keep the monitor away from water (ie do not swim, shower or bathe) and do not carry out contact sports.

It is unlikely that the tape used to attach the monitor will cause irritation, but if this does occur, you should take off the monitor.

3. When you return to the laboratory, the monitor will be removed and you will be given a short questionnaire to complete.

Please remember that you are free to withdraw from the experiment at any time and you do not need to give a reason for wishing to do so.

You will be paid £25 on completion of all tasks.

If you have any questions or if any problems arise during the course of the experiment, please do not hesitate to contact me:
Telephone: 01334 462077 (internal: 2077) (psychology department)
e-mail: rkep@st-and.ac.uk.

Measuring Activity Levels as a Health Outcome:
Volunteer Consent Form.

The purpose of this form is to ensure that you are willing to take part in this study and that you understand what it entails. Signing it does not commit you to anything you do not wish to do.

Have you read the information sheet? Yes / No

Have you had the opportunity to ask questions and to discuss the study? Yes / No

Have you received satisfactory answers to your questions? Yes / No

Do you understand that you are free to withdraw from the study
- at any time
- without having to give a reason for withdrawing? Yes / No

Have you any medical conditions that could be problematic in the walking task as described on the information sheet? Yes / No
If 'yes', please outline below.

Signed…………………………………..
Date……………………………………

Name (block letters)……………………………………………………………….

F.v: Walking Difficulties Study: Information Sheet and Consent Forms

Assessing Walking Limitations caused by Illness

Participant Information Sheet

Introduction

Thank you for showing interest in this research study. Before you decide whether or not to take part it is important that you understand why the research is being done and what it will involve. Please take your time to read the following information carefully and discuss it with friends, relatives and your GP if you wish. Please ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of the study?

When people have a health condition that affects their walking, it is important to be able to properly assess how well they can walk. A variety of different methods are used. Sometimes patients are asked to say what they can do, sometimes they are asked to actually show what they can do and sometimes they are asked to go about their ordinary business wearing a small instrument that measures what they do. This study aims to compare these different ways of assessing how well people can walk.

When we feel good we do more; when we feel bad we do less. A second aim of this study is to look at how the way people think and feel can affect how active they are.

Who is being asked to take part?

People who have difficulty walking due to a health condition but who are able to walk 100 yards are being asked to take part.

What will happen to me if I take part and what will I have to do?

If you agree to take part, you will come, on 2 consecutive days, to the Wolfson Health Psychology Laboratory at St. Andrews University. This is situated just off South Street in St. Andrews and disabled car parking spaces are available. Various measures of your walking will be taken, both in and outside the laboratory.

On your first visit to the laboratory, you will be asked to put on a gauge which tells us about your walking: when you have been walking and how energetically you have walked. We will show you an example of a record taken by the gauge before you are asked to wear it. It has 2 sensors, one of which will be taped to your thigh and the other to your chest. The sensors are connected, by wires, to a small, light box the size of a personal stereo. This is worn in a pouch at the waist.

There are 3 main parts to the study:

1. *In the laboratory:* you will be asked to do some tasks which will take about 1 hour to complete. This includes filling out some questionnaires and you will be asked to do a walking test to indicate how well you can walk. You will not be asked to do more walking than you feel comfortable with. Please wear comfortable clothes and shoes.

2. *Outside the laboratory:* you will be asked to wear the walking gauge for the next 24 hours and to return to the laboratory at the same time the following day. We would like you to go about your ordinary business as much as possible whilst wearing the gauge, but we ask that you keep the monitor dry (ie **do not shower, swim or bathe**).

It is unlikely that the tape used to attach the gauge will cause irritation, but if this does happen, you should take off the gauge.

3. *In the laboratory:* when you return to the laboratory, you will be asked to fill out a short questionnaire and the gauge will be taken off.

You will be paid £15 expenses when all the tasks are completed.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be asked to sign a consent form. You will still be free to withdraw at any time and without giving a reason. Your medical care will not be affected in any way.

Will my taking part be kept confidential?

All information collected about you during the course of the research will be kept strictly confidential. Any information which leaves the study centre will be anonymous; no-one will be able to recognise it as being about you.

If you are happy for me to contact your GP, your GP will be told that you are participating in this trial and it would be useful for me to ask your GP to confirm what the health condition is that causes your walking difficulty.

What will happen to the results of the study?

The results of the study will form part of Rachael Powell's PhD and may also be published in a scientific journal. The PhD thesis will be lodged in St Andrews University Library on completion.

Who is organising and funding the research?

The research is organised by the University of St Andrews, School of Psychology and is funded by the Medical Research Council Health Services Research Collaboration through a PhD studentship to Rachael Powell.

Who has reviewed the study?

The study was reviewed by the University of St Andrews School of Psychology Ethical Committee.

Contact for Further Information

If you have any questions before, during or after you take part in the study, please do not hesitate to make contact.

Rachael Powell
School of Psychology,
University of St. Andrews,
St. Andrews,
Fife.
KY16 9JU

Phone: 01334 462052
e-mail: rkep@st-andrews.ac.uk

Assessing Walking Limitations caused by Illness

Participant Consent Form

The purpose of this form is to ensure that you are willing to take part in this study and to let you understand what it entails. Signing this form does not commit you to anything you do not wish to do.

Have you read the Participant Information Sheet?	Yes / No*
Have you had the opportunity to ask questions and discuss the study?	Yes / No*
Have you received satisfactory answers to your questions?	Yes / No*
Do you understand that you are free to withdraw from the study: - at any time - without having to give a reason for withdrawing?	Yes / No*
Do you agree to take part in the study?	Yes / No*

*Delete as applicable

Signed……………………………………………………Date……………

Name in block letters………………………………………………………..
Address……………………………………………………………………..
……………………………………………………………………………….

Postcode…………………………………………………………………….

Assessing Walking Limitations caused by Illness

Participant Consent Form 2

We did not tell you in advance that we would record your walking speed as you left the laboratory yesterday. Do we have your permission to use this data? Yes / No*

*Delete as applicable

Signed……………………………………………………Date……………

Name in block letters……………………………………………………

Assessing Walking Limitations caused by Illness

Consent to contact GP

I (Rachael Powell) would like to contact your GP to a
~~the cause of your walking difficulty to be confirmed.~~
have your permission to do this? Yes / No*

*Delete as applicable

GP's name: ……………………………………………………………

GP's address: ……………………………………………………………
………………………………………………………………………
………………………………………………………………………

Signed……………………………………………………Date……………..

Name in block letters…………………………………………………..

Date of birth………………………………………………………………

F.vi: Experimental Study: Information Sheet and Consent Forms.

Beliefs and Activity

Participant Information Sheet

Thank you for taking part in this research. The study is looking at the relationships between how people think and feel and how active they are.

The study mainly involves completing some questionnaires. These will be given in 3 parts. After the first questionnaire section, you will be asked to walk down a corridor and back, counting your steps in order to assess step length. This will be followed by the other parts of the study. All answers will be kept strictly confidential.

You will be paid £4 in return for your participation.

Please remember that you are free to withdraw at any time.

Do you have any questions?

If you wish to ask any questions or make any comments after completing the study, please contact:

Rachael Powell
School of Psychology
University of St Andrews

e-mail: rkep@st-and.ac.uk
Telephone: 01334 462052 (internal extension: 2052)

Beliefs and Activity

Participant Consent Form

The purpose of this form is to ensure that you are willing to take part in this study and to let you understand what it entails. Signing this form does not commit you to anything you do not wish to do.

Have you read the Participant Information Sheet?	Yes / No*
Have you had the opportunity to ask questions and discuss the study?	Yes / No*
Have you received satisfactory answers to your questions?	Yes / No*
Do you understand that you are free to withdraw from the study:	
- at any time	Yes / No*
- without having to give a reason for withdrawing?	
	Yes / No*
Do you agree to take part in the study?	

*Circle appropriate answer

Signed………………………………………………Date………………

Name in block letters…………………………………………………….

Beliefs and Activity

Participant Consent Form 2

We did not tell you in advance that we would record your walking speed as you left the laboratory. Do we have your permission to use this data? Yes / No*

*Circle appropriate answer

Signed……………………………………………………Date………………

Name in block letters……………………………………………………

Appendix G: Ethics Committee Approval Letters: Pilot, Walking Difficulties and Experimental Studies.

University of St Andrews

School of Psychology

UNIVERSITY OF ST ANDREWS
SCHOOL OF PSYCHOLOGY ETHICS COMMITTEE

12 April, 2000

Rachel Powell
School of Psychology
University of St Andrews

Dear Rachel

Re: Measuring activity levels as a health outcome

The above-named project has been read and approved by the School of Psychology Ethics Committee.

However the committee have some concern that the payment of £25 made to participants on completion of all tasks may serve to exhibit the withdrawal from the study by those who would otherwise wish to do so. We wondered whether a mechanism could be arrived at for paying those who withdraw. In particular, if a subject withdraws because of an allergy to the tape attaching the monitor, shortly before conclusion of the 24 hours then it would seem unfair not to pay anything to that person. Perhaps you could let us know where you have a proposal which meets this concern.

If, during the course of the proposed research, any important condition were to alter, then the Committee would wish to be informed.

Yours sincerely

Dr Hugh Morris
Convener

Dictated but not read

St Andrews, Fife KY16 9JU, Scotland
Switchboard: (01334) 476161 Extension: Direct Line: (01334) 46
Fax: (01334) 463042

University of St Andrews

School of Psychology

18 April, 2000

Dr. Hugh Morris
University of St.Andrews
School of Psychology Ethics Committee

Dear Dr. Morris,

Re: Measuring activity levels as a health outcome

Thank you for the approval of this project.

I do think the issue that the payment of £25 to participants on the completion of all tasks may inhibit withdrawal from the study is a reasonable concern, as is the unfairness of paying nothing where withdrawal is due to allergy to the tape. I have, therefore, taken advice. Following other studies using a similar pattern of payment, it has been proposed that, where participants have a genuine reason for withdrawal such as a reaction to the tape or a change in their circumstances, they will be paid a sum proportionate to their participation time.

I hope that this meets the concern of the committee.

Yours sincerely,

Rachael Powell.

St Andrews, Fife KY16 9JU, Scotland
Switchboard: (01334) 476161 Extension: Direct Line: (01334) 46
Fax: (01334) 463042

University of St Andrews

School of Psychology

UNIVERSITY OF ST ANDREWS
SCHOOL OF PSYCHOLOGY ETHICS COMMITTEE

11 July, 2001

Rachael Powell
School of Psychology
University of St Andrews

Dear Rachael

Re: Assessing Walking Limitations Caused by Illness

The above-named project has been read and approved by the Convener of the School of Psychology Ethics Committee.

If, during the course of the proposed research, any important condition were to alter, then the Committee would wish to be informed.

Yours sincerely

Dr Hugh Morris
Convener

Dictated but not read

St Andrews, Fife KY16 9JU, Scotland
Switchboard: (01334) 476161 Extension: Direct Line: (01334) 46
Fax: (01334) 463042