The Development and Validation of a Physical Activity Recall Questionnaire for Adults Using a Cognitive Model of the Question-Answer Process

This thesis is submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy in the Centre for Sport Science and Health at Dublin City University

Submitted by Niamh Martin B Sc (Sport and Exercise Science)
Supervised by Dr Catherine Woods
Submission Date July 2005
Volume 1 of 2

DECLARATION

I hereby certify that this material which I now submit for assessment on the programme of study, leading to the award of Doctor of Philosophy, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work

Signed Niamh Martin ID NO 51179032

Niamh Martin Date 08/08/05

ABSTRACT

Research has revealed that regular physical activity (PA) positively affects health. Accurate measurement of physical activity levels among adults is thus required. Recall questionnaires can obtain this information but retrieval of accurate information is known to be difficult. Accuracy can be enhanced by developing questionnaires that use cognitive techniques and recall cues. This research designed a valid and reliable past-seven-day Physical Activity Recall Questionnaire (PARQ), using a model from cognitive psychology. This model explains the memory processes used to answer questions and identifies recall cues that can enhance accuracy.

A draft version of PARQ was developed in study one. Items were generated following discussions in eight focus groups, conducted in both urban and rural settings. Following analysis of transcripts, descriptions of physical activity-related terms and a variety of recall cues, such as psychological and physiological descriptions, were included in the design of the draft PARQ.

The cognitive processes used to complete the questionnaire were assessed through a series of cognitive interviews, in study two. Identification of cognitive problems in the questionnaire was facilitated by applying Tourangeau's (1984) model of the question-answer process, during data collection and analysis. Problematic areas, detected in the comprehension and retrieval stages of answering the question, were amended. This produced an improved version of PARQ that was ready for psychometric evaluation.

The test-retest reliability of PARQ was comprehensively assessed in study three. Initially, intraclass correlations (0.95) indicated that PARQ is a highly reliable PA measurement tool. High intensity activity and moderate intensity are equally reliable (0.94). Reliability was better when PARQ included a personal calendar (PC) recall cue (0.95) against when the PC was not included (0.83). The order of questions in the questionnaire was assessed and found not to significantly affect the reliability of PARQ. The reliability of PARQ in measuring total, high and moderate intensity physical activity was further asserted, with similar results emerging following analysis by Pearson correlation coefficients and the Coefficient of Repeatability. Assessment of specific types of PA showed that the category with the least amount to be recalled was the most reliable (family activity = 0.97). As the volume of recalled activity increased, correlation scores reduced.

This reflects Conway's description of the burdensome retrieval of frequently performed memories, due to the structure of autobiographical memory

The final study examined the criterion validity of PARQ, firstly with an RT3 accelerometer (r=0 52) The performance of a submaximal walking treadmill test by each of the participants enabled the assessment of the individual intensity questions A time-matched heart rate monitor and accelerometer concurrently assessed the criterion validity of high intensity (r=0 55) and moderate intensity (r=0 4) Concurrent validity of PARQ with the Stages of Change model of intentional behaviour change was also established, with regularly active participants reporting more activity than the not-regularly active participants The ability of PARQ to provide valid information about the duration and intensity of different types of activities, means that it is capable of proving information about energy expenditure This is important in terms of measuring the dose-response relationship between health and physical activity participation

The personal calendar acts as a recall cue that successfully probes the memory to retrieve accurate information The application of the cognitive model to the design of PARQ has produced a reliable and valid working questionnaire It directly provides accurate information about minutes of PA participation and indirectly provides information about energy expenditure, both of which are relevant to health-related guidelines, published by the American College of Sports Medicine and the Centre for Disease Control

Acknowledgements

To Mam and Dad, for the constant, unconditional and endless encouragement and support Am here at last because of you

To Catherine, thank you for your help, guidance and efforts in getting me to my final destination

To Emer, Patrick, Aine, Shay, Fergal and Patria, for taking the risk and giving more encouragement and support than is allowed to be given to one person Thank you all!

To Deirdre, Susan, Aggie, Siobhain, Alison and Niamh, for giving so freely of your time to keep me sane Thank you (to the power of infinity) for your concerted efforts in keeping it real, giving me plenty to laugh about, buying me nice things, not being offended by the cheap tack given in return, propping (me) up (against) a counter, the holidays, the lifts, the phone calls, for listening to me go on and on and on and for giving me a good kick when I needed it But mostly for the laughs and the good times

To Eimear, Norah and Karen for soooo much over the past three years! Couldn't have asked for a better bunch to share the experience with Thanks for the laughs, the comments and the invaluable help

To those that took the time to give their professional opinion, your input is very much appreciated – Olivia Hurley, Siobhan McArdle, P J Smyth, Jacinta O'Brien, Suzanne Guerin, Ann Rowlands, Gordon Willis, Nanette Mutrie, Matthew Lowther

To Aisling, Javier and Paul for all your time and effort in helping me to organise myself! Also, for going that extra mile when you didn't have to – thanks so much

To Ciaran Woods, for sharing your artistic talent and professionalism in creating the masterpiece

To all of the participants involved in this study that gave so willingly of their time and energy

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1- INTRODUCTION

1.1. Focus of this Thesis

This research addresses the methods used to design and validate a self-administered past-seven-day physical activity (PA) recall questionnaire. It was undertaken in response to the absence of a recall PA questionnaire that was appropriate to the lifestyle and temperament of the Irish adult population (aged 18-55 years) and which demonstrated acceptable reliability and validity properties. This thesis will describe the novel method used to design this physical activity questionnaire, with the application of models from the area of cognitive psychology that explain, (a) the structure of memory (Conway, 1996; Conway & Rubin, 1993) and (b) the way in which information to a question is retrieved from memory (i.e., the question-answer process) (Tourangeau, 1984; Tourangeau, Rips, & Rasinski, 2000). Finally, the methods used to assess the reliability and validity of the new questionnaire are also outlined in the subsequent chapters.

The following sections provide some background information about the basis of this current research question. Firstly, the relationship of PA to physiological and psychological health benefits and the current recommended minimum PA participation guidelines are highlighted. This is followed by an illustration of the cognitive factors that affect the accurate recall of PA participation and which therefore need to be considered in the design of a recall questionnaire.

1.2. Background

1.2.1 Current Minimum PA Participation Recommendations for Health Benefits

Physical Activity (PA) has been defined as "any bodily movement produced by skeletal muscles that results in energy expenditure", (Caspersen, Powell, & Christenson, 1985, p.262). Engagement in regular amounts of PA has been shown to positively affect the morbidity of, among other diseases, cancer, diabetes mellitus, coronary heart disease and obesity (Paffenbarger, Lee & Hyde, 1993; U.S. Dept. of Health and Human Services, 1996; Friedenreich, Courneya & Bryant, 2001). Psychological benefits such as improvements in depression, stress management and anxiety are also recognised outcomes of a regularly active lifestyle. Evidence-based research from the last decade has shown that health benefits are evident from participation in moderate intensity activity (Thune & Furberg,

2001, Dunn, Trivedi, & O'Neal, 2001, C M O , 2004) This represented a paradigm shift away from a focus on exercise and physical fitness to a more health-centred physical activity concept and led to the publication of the current recommendations given by the American College of Sports Medicine (ACSM) and the Centre for Disease Control (CDC) in the U S A These recommendations state that for the purpose of acquiring associated health benefits, "every adult should accumulate 30 minutes or more of moderate intensity physical activity over the course of most, preferably all, days of the week" (Pate et al , 1995, p 402) Other recommendations have also been proffered by the ACSM (2000) in terms of the minimum amount of energy that should be expended in physical activity for the purpose of maintaining health "The ACSM recommends a target range of 150 to 400 kcal of energy expenditure per day in physical activity The lower end of this range represents a minimal caloric threshold of approximately 1,000 kcal per week from physical activity and should be the initial goal for previously sedentary individuals" (p 151)

The health benefits incurred are proportionate to the baseline aerobic fitness level of the participant, as well as the Frequency, Intensity, duration (Time) and Type, (i e , the FITT principles), of the activity engaged in This is referred to as the dose-response curve (Philips, Pruitt, & King, 1996, Lamonte & Ainsworth, 2001) An illustration of the dose-response curve, which details the nature of the relationship between PA and health benefits is given in Figure 1 1 overleaf Curve A (Sedentary group) describes the dose-response pattern upon which the current PA recommendations are based, in that it indicates that most of the benefits are attained during regular participation in low to moderate intensity of PA (Bouchard, 2001) It also shows that more relative benefits are to be obtained when the sedentary group become active For individuals with a higher baseline fitness (Groups B and C) health benefits can also be incurred, when activity is performed for longer and at a higher intensity, than what they were previously doing

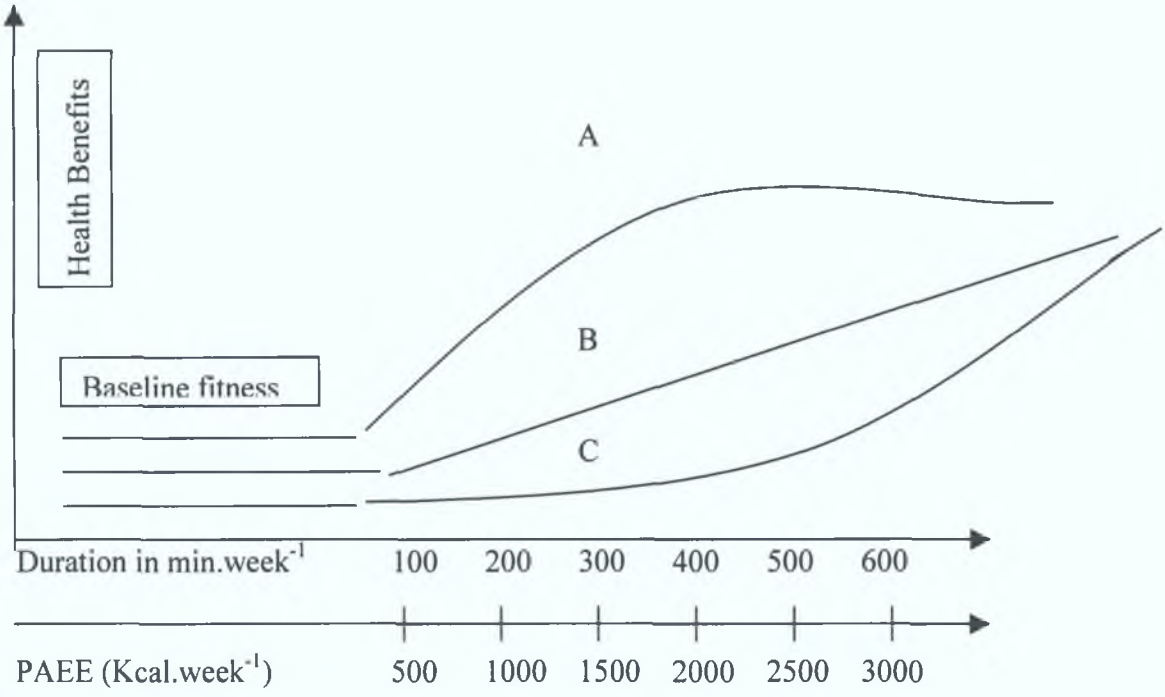

Figure 1.1. Schematic representation of dose-response curve and expectant health benefits for each group by baseline activity level. A= Sedentary, B = Moderately active, C = Very active. Note. From "Physical activity and health: introduction to the dose-response symposium", by C. Bouchard, 2001, MSSE, 33 (6), p.S349.

Although the optimum amount of PA for health benefits remains an unknown entity (ACSM, 2000), the dose-response curve represented in Figure 1.1. above, highlights the value of knowing how much physical activity individuals are engaging in and whether they are reaching the minimum ACSM/CDC recommended participation guidelines for health. The self-report questionnaire represents an inexpensive and productive method of assessing whether participants are meeting minimum participation levels. To accomplish this task however, the questionnaire must be specifically designed to enquire about activities of at least moderate intensity, that are inclusive of activities of daily living and not just sport and exercise. This has not always been evident in PA questionnaires to date (Kriska & Caspersen, 1997).

There are limitations to using recall questionnaires as the primary or sole method of data collection. Many factors associated with cognition and questionnaire design, affect the accurate recall of autobiographical memories, (i.e., memories pertaining to the past self). These factors are collectively known as questionnaire response bias (Willis & Schechter,

1997) More research is required to examine the cognitive and questionnaire design issues that affect the recall of information about PA from autobiographical memory

1 2 2 Cognitive Aspects of Memory Recall and Questionnaire Design

Questionnaire response bias is a collective term for factors that influence the quality of data received from the respondent and includes (i) recall bias and (ii) measurement bias (Sudman & Bradburn, 1982, Willis et al , 1997) Recall bias refers to the cognitive factors that influence the question-answer process (i e , the process of recalling information from memory in response to a question) These cognitive factors include (i) the retrieval strategy used, (ii) the cognitive burden of the type of information to be recalled, and (in) the presence or absence of memory cues Measurement bias refers to issues concerned with the general design and layout of the questionnaire, e g , (i) structure of questionnaire, (ii) question wording and (iii) question order (Schwarz, 1999, Smith, Jobe, & Mingay, 1991, Sudman et al , 1982) The combination of these cognitive and design features known to affect questionnaire response bias stimulated an interdisciplinary approach to questionnaire design between cognitive psychologists and survey methodologists The subsequent series of Cognitive Aspects of Survey Methodology (CASM) seminars were established to address issues of recall and measurement bias in questionnaires (U S Department of Health and Human Services, 1999)

1 2 2 1 Addressing Recall Bias

One outcome of the inaugural CASM meeting was the development of a cognitive model that provides a theoretical basis for the question-answering process (Tourangeau, 1984) This model has four stages that detail how the respondent (1) encodes (comprehends) the question, (2) retrieves the information from memory, (3) judges whether it is accurate and relevant to the question asked, and (4) decides whether to edit the answer (i e , decides if the answer reached will suffice and is socially desirable) This model provides a theoretical basis for explaining how PA information is stored and retrieved and has been successfully used in the design and pre-testing of food frequency questionnaires (Subar et al , 1995), general health and population-based surveys, i e , The National Health Interview Survey (National Centre for Health Statistics, 1991) and in questionnaires designed at the U S Census Bureau (Gerber, 2002) and the National Institute of Health (Willis, 2002) Only one PA questionnaire (the Lifetime Physical Activity Questionnaire)

has been developed and pre-tested using this cognitive model (Friedenreich, Courneya, & Bryant, 1998) This was an interview-led questionnaire for lifetime activity This thesis describes the novel application of this cognitive model to the design of a questionnaire that requests information about PA participation over the past-seven days

Research is required that explores the method in which PA information is structured within autobiographical memory (ABM) and, in particular, what recall cues can assist the respondent in accessing that information A model of ABM, such as Conway's model (1996, Conway & Rubin, 1993), offers a theoretical basis for explaining how memories are structured and accessed An aided recall procedure (i e , a method for providing memory cues to help the respondent access the information) that reflects the structure of ABM has proven effective (Belli, Shay & Stafford, 2001) It is referred to as a "personal timeline" or "personal calendar" and provides the respondent with multiple cues that are personally relevant, such as when, where and what happened The personal calendar is believed to reflect the structure of the multiple pathways that access memories in ABM This thesis explores the use of a personal calendar in aiding recall and also uses a model of ABM structure (Conway, 1996) to explain the difficulties in accessing various types of PA information that represent different levels of cognitive burden to the respondent

1 2 2 2 Addressing Measurement Bias

The design of the questionnaire (e g , layout, question wording and order) can affect the quality of information recalled (Sudman et al , 1982, Tourangeau et al , 2000, Warnecke et al , 1997) Recall of daily activity was enhanced for schoolchildren (N=24, age 11-16 years) for structured days that contained landmark structures, e g , going to school, break-times, evening activities (McKenna, Foster & Page, 2004) Decomposing the day into known sections (e g , before school, at school, after school, after dinner) also enhanced the memory of children (N=22, 8-12 years), compared to when the day was not decomposed (Baranowski, 1985) More research is required to examine the contexts in which an adult's day can be decomposed This thesis explores the contexts that adults perceive themselves to be active during the day

The characteristics and language of the target population are an important consideration in the design of a questionnaire, as it reduces the potential for misunderstanding and ambiguous wording (Warnecke et al , 1997, Ainsworth, Sternfeld, Richardson, & Jackson, 2000b, Sudman et al , 1982) For PA questionnaires, understanding

of key words like "physical activity", "exercise" and "intensity" among respondents is paramount, yet this aspect of questionnaire design has rarely been explored (Tudor-Locke et al., 2003). Misunderstandings are further compounded when culture and language differences are considered, highlighting the necessity for population-specific questionnaires that contain culturally relevant expressions and examples (Warnecke et al., 1997; Sallis & Saelens, 2000). This thesis has used qualitative research methods to identify culturally specific interpretations and understanding of PA-related words.

The effects of question order have primarily been explored in opinion and attitude surveys. The main type of effect found occurs when one of the questions is a general one on an issue and the other is more specific on the same issue. Findings suggest that answers for the specific question do not change if it is placed first or second but that answers for the general question are affected if it comes after the specific question. Explanations for this effect suggest that after answering the specific question, some respondents assume that the general question excludes the specific part already answered (Kalton & Schuman, 1982; Schuman & Ludwig, 1983; Strack, Schwarz, & Gschneidinger, 1985). Research is required to assess whether the order of questions in a PA recall questionnaire effects the quality of information retrieved. This thesis explores whether the order in which questions about moderate and high intensity physical activity are asked affects the reliability of information reported.

1.3. Identification of Research Areas

The research in this thesis evolved subsequent to the identification of disparities in the literature. Firstly, the emergence of the interdisciplinary CASM interest group has provided evidence-based assertions for using cognitive-based methods to enhance recall and improve data quality in the areas of food frequency, opinion polls and health surveys (Subar et al., 1995; Gerber, 2002; Willis, 2002; National Centre for Health Statistics, 1991). The application of a cognitive model of the question-answer process (Tourangeau, 1984; Tourangeau et al., 2000) has never been applied to the design and pre-testing of a self-administered past-seven-day PA recall questionnaire. Of particular interest to the PA-Health relationship are the problems associated with recalling activities of moderate and high intensity. Knowledge of the structure of ABM (Conway, 1996) may assist in understanding results and provide a rationale for using specific design features intended to aid the recall of accurate memories about moderate and high intensity physical activity.

Secondly, the understanding of terms associated with PA among target populations have been inadequately explored and the necessity for cultural specificity in questionnaire wording is often overlooked (Tudor-Locke et al , 2003, Warnecke et al , 1997) This suggests that the comments and descriptions of the properties of PA and how they are generally expressed in the vocabulary of the target population of this questionnaire (i e , Irish adults, 18-55 years old) need to be explored Documentation of methodologies undertaken in PA qualitative research are inadequately depicted in publications (Biddle, Markland, Gilbourne, Chazisarantis, & Sparkes, 2001) This thesis documents two different methodologies (i e , focus groups and cognitive interviews) which were employed to explore the descriptions of PA-related terms by the target population

Finally, the health-related significance of monitoring participation in activity of at least moderate intensity, as proposed by the ACSM/CDC (Pate et al , 1995), requires the validation of individual questions about high and moderate intensity The lack of one gold standard criterion instrument that is capable of validating the FITT principles of lifestyle PA obtained in a questionnaire, suggests that the concurrent use of two instruments should be employed (Ainslie, Reilly, & Westerterp, 2003) Criterion instruments such as heart rate monitors and accelerometers have known limitations in measuring lifestyle PA (Nichols, Morgan, Sarkin, Sallis, & Calfas, 1999, Jakicic, Winters, Lagally, & et al , 1999, Welk & Corbin, 1995) Many studies have used a single criterion instrument to validate a new PA questionnaire (Craig et al , 2003, Lowther, Mutrie, Loughlan, & McFarlane, 1999, Richardson, Ainsworth, Wu, Jacobs, & Leon, 1995) Other studies have used more than one piece of equipment concurrently (Ainsworth et al , 2000b, Aadahl & Jorgensen, 2003) However, these instruments were not synchronised (matched) for time and were thus analysed individually This provided two separate validity scores for the questionnaire The concurrent use of two time-matched instruments should compensate for one another's limitations and provide a more comprehensive method of measuring PA and validating a new questionnaire The concurrent use of two time-matched instruments (heart rate monitor and accelerometer) was undertaken in this research

1 4 Research Design

1 4 1 Use of Qualitative and Quantitative Methodologies

This thesis consists of two qualitative and two quantitative studies The decision to use both of these methodologies was a product of the aim and objectives of the relative studies Quantitative and qualitative paradigms contrast in all aspects of research, i e , theoretical framework, selection of sample, collecting, interpreting, analysing and reporting data In essence, qualitative research tends to use inductive reasoning to develop hypotheses from observations made in the field, while quantitative research usually has preconceived hypotheses against which results are assessed (Flick, 2002, Thomas & Nelson, 2001)

1 4 1 2 Qualitative Research

The qualitative approach permitted a method of understanding the meaning and description of specific terms and phrases for participants in a particular setting (Strauss & Corbin, 1998, Creswell, 1998) This was a particular requirement for the focus groups in study one, where interpretations and descriptions of key words and phrases related to PA were needed to generate items and questions for use in the questionnaire The focus group facilitated the discussion that satisfied the requirements of study one, which was to discover what aspects of PA were considered salient and particularly memorable by the participant From this study, items could be generated for inclusion in the questionnaire and possible recall cues could be identified to assist the respondent to remember accurately Study two also required the opportunity for observation and exploration in the field through one-on-one interviews This study sought to ensure clarity and understanding of the layout and meaning of words used within the draft questionnaire and this was made possible by qualitative enquiry

The analysis of qualitative data occurs simultaneously with data collection This permits the researcher to be critically aware of what they are observing and to enable the focus to be kept on the purpose of the original investigation (Strauss et al , 1998, Barbour & Kitzinger, 1999) The general phases associated with qualitative research are (i) data collection, (ii) analysing and categorisation, and (iii) interpretation and theory construction (Flick, 2002) In study one, data was interpreted using a constant comparison technique that facilitated understanding of a complex topic that is not feasible through quantitative methods (Tudor-Locke et al , 2003, Creswell, 1998) Qualitative research is usually

presented through words, descriptions and images while quantitative research is usually represented through numbers (Thomas et al , 2001)

Data collection procedures for qualitative research are, interviews, documents, observations and audio-visual equipment The type of data collection is however usually determined by the nature of the research question In this thesis, interviews in the form of focus groups and individual cognitive interviews were used to collect data required to produce a draft version of the new questionnaire

1 4 1 2 1 Focus Groups

These are a qualitative method of enquiry that enable, (i) exploration and discovery, (ii) context and depth, and (iii) interpretation of a given topic in a group environment (Morgan, 1998b) They were used in study one of this research, for the purpose of generating items that are relevant to the new questionnaire In particular, focus groups facilitated the exploration of the understanding of terms related to lifestyle physical activity and moderate and high intensity These terms are specific to the context of the ACSM/CDC minimum participation guidelines for health benefits (Pate et al , 1995) and thus it was important to investigate how these terms were described in the vocabulary of the target population Focus groups have previously been used for the development and refinement of instruments, to identify domains, to obtain natural vocabulary for item generation for questionnaires and to assess cultural appropriateness (Shaughnessy & Zechmeister, 1997)

Although demonstration of their use in PA questionnaire design is not prolific in literature (Sallis et al , 2000, Friedenreich et al , 1998, Durante & Ainsworth, 1996), this type of investigative format has previously been used to elicit information about complex behaviours and perceptions, like smoking (Frankland et al , 1999) and physical activity (McKenna et al , 2004, Tudor-Locke et al , 2003) The advantage of a focus group for the topic of physical activity lies in its ability to potentially provide a richness and complexity of responses that are generated as a result of group dynamics, discussion and contrasting opinions (Morgan, 1998a) This type of data was required for the purpose of study one, i e , to generate items for a physical activity questionnaire that reflect the activity experiences of the target population, in a language they are familiar with

1.4.1.2.2. Cognitive Interviews

These were used with effect by McKenna and colleagues (2004) for the purpose of exploring the recall of PA among young adults. The cognitive interview (CI) is a qualitative method used to pre-test a questionnaire (Willis, 2002; Willis, 1999). It can identify potential difficulties that may arise as a result of human behaviour when answering a question. The process of cognitive interviewing aims to structure the pre-testing situation by employing a model of cognitive theory to construct the interview, e.g., a model of the question-answer process (Tourangeau, 1984; Tourangeau et al., 2000). This model provides an insight into cognitive problem areas in the questionnaire associated with understanding and retrieving information. The cognitive interview has been used to pre-test an interview-led Lifetime Physical Activity Questionnaire (Friedenreich et al., 1998) and other behaviour questionnaires, e.g., food (Subar et al., 1995) and cultural differences in question wording (Warnecke et al., 1997). This thesis details the novel application of cognitive interviewing to pre-test the functionality of a self-administered past-seven-day PA recall questionnaire.

1.4.1.3. Quantitative Research

The deductive properties of quantitative research were more appropriate to the purpose of studies three and four, i.e., assessment of the reliability and validity of the questionnaire. The objective method of data collection and statistical analysis procedures associated with the use of validated instruments suited the testing of the hypotheses in these studies. Other features of the quantitative paradigm were appropriate to the requirements of achieving objective results from these studies. These included designing the study so that influence from the researcher is excluded as far as possible and participants are usually selected from a random sample, thus ensuring objectivity of the study (Flick, 2002; Shaughnessy & Zechmeister, 1997). Use of a quantitative paradigm for these two studies meant that the results could be extrapolated for generality to other samples of similar characteristics.

A summary of the objectives of both paradigms highlights the differences and merits of each of them and their relative appropriateness for the studies outlined in this thesis, "quantitative methods intend to meet three goals; description, explanation and prediction", (Shaughnessy et al., 1997, p.20), while "the objectives of qualitative methods are primarily description, understanding and meaning", (Thomas et al., 2001).

1 5 Aim and Objectives of this Thesis

The aim of this thesis was to design, pre-test and validate a self-administered past-seven-day recall PA questionnaire, that was culturally specific and used cognitively-based models of the structure of ABM and the question-answer process, to enhance the quality of data recalled Four studies were undertaken to complete this task and the aims and objectives of each are outlined below

1 5 1 Study One

The aim of this study was to generate items for questions in the initial draft of the new physical activity recall questionnaire, through a series of focus groups The specific objectives were

i To explore definitions, examples and contexts of "physical activity" and associated terms, such as "moderate" and "high intensity" (to enable appraisal of compliance with the 1995 CDC/ACSM guidelines)
ii To explore the physical and psychological descriptors of PA and intensity that are salient and relevant to the target population
iii To use Tourangeau's (1984, Tourangeau et al , 2000) cognitive model of the question-answer response process to identify how PA is understood and recalled by the participants and to identify possible recall cues for inclusion in the questionnaire

1 5 2 Study Two

The aim of this study was to qualitatively pre-test the draft version of the new physical activity recall questionnaire, by means of a cognitive interviewing technique (Willis, 2002, Gerber, 2002) The four cognitive processes of the question-answer process model (see p 4), (Tourangeau, 1984, Tourangeau et al , 2000) were assessed during the pre-testing of the questionnaire The specific objectives of this study were

i To identify which cognitive processes are affected by the problems, ambiguities and misunderstandings present in the draft version of the new physical activity recall questionnaire
ii To identify what retrieval processes participants use to calculate the frequency and duration of activities at moderate and high intensity

iii To amend problematic areas identified within the questionnaire and to prepare it for psychometric evaluation

1 5 3 Study Three

The aim of this study was to assess the reliability properties of the new physical activity recall questionnaire, in a test-retest method This involved evaluating the reliability of the component parts of total physical activity, i e , moderate and high intensity activity The specific objectives of this study were

i To assess whether the personal calendar recall cue has an effect on the reliability of the questionnaire
ii To comprehensively ascertain the reliability of total physical activity
iii To ascertain and compare the individual reliability scores of moderate and high intensity physical activity
iv To assess the reliability properties of different types of activity, (e g housework, sport) within these intensities
v To ascertain whether a question-order effect is apparent in influencing the reliability score of the questionnaire
vi To ascertain whether PARQ is reliable for participants at all SOC, i e not regularly active and not regularly active

1 5 4 Study Four

The aim of this study was to assess the validity of the new physical activity recall questionnaire The specific objectives were to assess

i The criterion concurrent validity of the new questionnaire, with the RT3 accelerometer as the criterion measure
ii If the strength of the relationship between questionnaire data and the criterion measure (RT3), will lessen as time between occurrence of activity and time of recall increases
iii The criterion concurrent validity of the new questionnaire with the simultaneous use of temporally-matched heart rate and RT3 accelerometer data (minutes of activity) as the criterion method
iv If high intensity activity (minutes of activity) will demonstrate better concurrent validity than moderate intensity (minutes of activity)

v The construct validity of the new questionnaire against the Stages of Change model of behavioural intention

1 6 Delimitations of this Thesis

The lack of an objective gold standard physical activity measurement instrument was a delimitation of this study This meant that the new questionnaire was being compared with a measurement instrument (a triaxial RT3 accelerometer), that was less than optimal in measuring the diverse properties of PA (Nichols et al , 1999, Welk et al , 1995, Jakicic et al , 1999) A second delimitation of the validity study was the lack of an integrated heart rate monitor and accelerometer unit as they are not yet commercially available (Ainslie et al , 2003) Instead, separate heart rate and accelerometer equipment were temporally matched, to ensure that both instruments were measuring the exact same space in time, and then used concurrently Although numerous precautions and validity checks were employed to ensure that the procedure was as accurate as possible, this methodology is far from optimal in ensuring minimal measurement error

The natural decay in memory associated with the passing of time was a delimitation, as it meant that an element of systematic bias was present in the retest condition (time two) that was not present in the test condition (time one) In this study, a four-day overlap methodology was used for the test-retest reliability This meant that what was the previous day's recall at time one was now four days away at time two The rationale for using a four day overlap is discussed in Chapter Five, however the temporal decay in cognitive processes is a natural occurrence that can not be altered (Linton, 1982, Wagenaar, 1986, Eysenck & Keane, 2001)

The complete novelty of different aspects of this thesis limited the comparison of many of the results found here to other similar studies Novel aspects included (i) the use of the cognitive interview technique to pre-test the questionnaire, (ii) the use of the cognitive models of autobiographical memory structure and the question-answer process to design the questionnaire and interpret results, (iii) assessment of a question-order effect, (iv) assessment of the impact of a personal calendar recall cue on reliability, and (v) the concurrent use of two time-matched criterion measures to validate the questionnaire

1 7 Structure of this Thesis

This thesis consists of seven chapters Chapter Two explores the relevant literature pertaining to, (i) the relationship between PA and health, (ii) methods of measuring PA (iii) other PA questionnaires in use, and (iv) the current theoretical models used to identify, explain and overcome cognitive difficulties associated with questionnaire response bias The development of the draft version of the physical activity questionnaire through focus groups is outlined in Chapter Three while Chapter Four details the cognitive pre-testing procedure used to prepare the questionnaire for psychometric evaluation The methods, results and discussion of the reliability and validity studies are the focus of Chapters Five and Six Finally, Chapter Seven links the findings of all of the individual studies and provides recommendations for instances in which this questionnaire should be optimally used, as well as guidelines for the design, pre-testing and validation of other PA questionnaires

1 8 Final Note

To enable ease of reference in the succeeding chapters, the questionnaire developed as an outcome of this research will hereafter be referred to as PARQ (Physical Activity Recall Questionnaire) Furthermore, the final version of the questionnaire (Appendix 9) also contains questions about sedentary activities The development and reliability of these questions are not the focus of this thesis and so are not referred to in the following chapters

CHAPTER 2 - LITERATURE REVIEW

2 1 Introduction

Self-report physical activity (PA) questionnaires have been used by epidemiologists and exercise scientists to study the relationship between physical activity participation and certain diseases, such as site-specific cancers (Thune et al , 2001, Friedenreich, Courneya, & Bryant, 2001), cardiovascular and coronary heart disease (Boreham, Twisk, Murray, Strain, & Cran, 2001, Lee, Sesso, & Paffenbarger, 2000) and depression (Dunn et al , 2001, Paffenbarger & Lee, 1994) They have also been used to assess the effectiveness of interventions in increasing physical activity (Winett, 1998, Sallis et al , 1985) and patterns of participation (Livingstone et al , 2001, Fahey, Layte, & Gannon, 2004, Kelleher et al , 2003)

Self-report physical activity questionnaires are heavily reliant upon memory recall processes to obtain information about activities done at some time in the past Respondent's often supply information that is less than accurate because of limits on their ability to access and retrieve relevant information from memory (Bradburn, Rips, & Shevell, 1987, Friedenreich et al , 1998, Durante et al , 1996) Improvements may be possible with the inclusion of memory probes and cues and by using cognitive models to assess strategies used to recall information This strategy has been used in other questionnaire areas, such as food frequency (Subar et al , 1995) and social-economic behaviours (Belli, Shay, & Stafford, 2001) and has also been advocated for use in the PA domain (McKenna et al , 2004, Durante et al , 1996) Using a cognitive model of recall is a novel approach to the design and pre-testing of a self-administered past-seven-day physical activity questionnaire and is the primary research focus of this current study

Validation of self-report PA questionnaires is limited by the lack of a gold standard criterion instrument that reflects the output of the questionnaire, in terms of measuring the Frequency, Intensity, Time and Type (FITT) aspects of PA (Sallis et al , 2000, Welk, Blair, Wood, Jones, & Thompson, 2000, Montoye et al , 2001) The concurrent use of two criterion instruments that, together, can capture all FITT information has been advocated as a novel method of validating a PA questionnaire (Bassett, 2000) and has been undertaken in this research

The following section will review current literature that examines, (i) the relationship between PA and health, (ii) methods of measuring PA (iii) other self-report questionnaires

used in PA research, and (iv) the current theoretical cognitive models used to identify, explain and overcome cognitive difficulties associated with questionnaire response bias. A final section examines a model of intentional behaviour change, that provides an insight into when adjustments in behaviour and behavioural intention are likely to occur. The Stages of Change model was included in the criteria for recruitment in order to gain representation from people who are at various stages of being and intending to become physically active.

2.2. Physical Activity (PA) and Health

Physical inactivity has been recognised as a serious, global problem that contributes to high rates of unnecessary illness and premature death (Booth, 2000; Dubbert, 2002). In particular, people who are inactive are more likely to be susceptible to developing diseases of the cardio-vascular and circulatory systems, while regular participation in moderate intensity physical activity is known to reduce the incidence of these and other infirmities, such as depression, anxiety, osteoporosis and cancers (Blair, Cheng, & Holder, 2001; Friedenreich et al., 2001; Friedenreich & Rohan, 1995; U.S.Department of Health and Human Services, 1996) CMO report, 2004 . The incidence of co-morbidities are likewise positively affected, e.g., lipid profiles (Carroll, Cooke, & Butterly, 2000), hypertension (Lesniak & Dubbert, 2001) and diabetes (Kelley & Goodpaster, 2001). A summary of the physiological and psychological benefits of participation in regular PA are outlined below in Table 2.1.

Table 2.1. *Some physiological and psychological benefits of regular physical activity*

Reduces risk of developing	Improves
Heart Disease	Blood Lipid Profile
Type II Diabetes	Blood Pressure
Cancer (Breast, Colon)	Weight
Cardiovascular Disease	Strength and Flexibility
Obesity	Psychological well-being
Depression	Stress/Anxiety
Hypertension	Self-Esteem
	Functional Capacity

Note. This table represents a summary of the evidence-based research referred to in the previous paragraph and Section 2.1.

Recent evidence-based research has shown that health benefits are evident from participation in moderate intensity physical activity (Blair & Connelly, 1996; Paffenbarger et al., 2000). Results from cohort and case-control studies that investigated the independent effect of occupational and leisure-time PA on breast cancer (N=41) and colon cancer (N=48), indicated that the incidence of colon and breast cancer is positively affected by participation in PA of at least moderate intensity (Thune et al., 2001).

The extent of health benefits incurred depends upon the baseline aerobic fitness of the individual, as well as the Frequency, Intensity, duration (Time) and Type, (i.e., the FITT principles), of the activity engaged in (Lamonte & Ainsworth, 2001). The combination of these factors and their relationship with general health status is referred to as the "dose-response curve". This curve describes the health benefits that moderate intensity activity can confer on previously sedentary people. It also indicates the health improvements that individuals with a higher baseline fitness can obtain, following participation in more intense activities of a longer duration (see Chapter One, p.2. for a more detailed explanation of the dose-response curve). A dose-response effect has also been established between the incidence of clinically diagnosed depression and PA of at least moderate intensity. A longitudinal study, over a 27 year period, of male Harvard Alumni (N=21, 569, aged 25-74) showed a gradient effect in terms of incidence of depression. Men who engaged in activity of at least moderate intensity were at 17% less risk of developing clinically recognised depression, while those that engaged in high intensity activity were 28% less likely to be diagnosed (Paffenbarger et al., 1994).

The emergent data which advocated the health benefits available from participation in moderate intensity activity, led to the publication of the current recommendations by the American College of Sports Medicine (ACSM) and the Centre for Disease Control (CDC). These recommendations state that for the purpose of acquiring associated health benefits, "every adult should accumulate 30 minutes or more of moderate intensity physical activity over the course of most, preferably all, days of the week" (Pate et al., 1995). This represents a paradigm shift away from a previous focus on exercise and physical fitness, to a more health-centred PA concept. The debate about the optimum amount and level of PA required for achieving maximum health benefits is still ongoing (Wilmore, 2001; Lee & Paffenbarger, 1996). However, a comprehensive profile of PA, that provides information about the FITT principles, is necessary for assessing whether the minimum requirements for health benefits are being met.

Profiling the participation rates of Irish adults in PA has been recently undertaken in two national surveys The Economic and Social Research Institute's (ESRI) survey (Fahey et al , 2004) of sports participation found that less than 40% of Irish adults take part in regular PA (N=3,080, F=51%) In 2003, a National Health and Lifestyle Survey (SLAN) showed similar participation levels for a range of age groups and social classes According to these results, 28% of all adults engage in no activity at all, while 51% of all adults engage in some form of regular physical exercise (Friel, Nic Gabhainn & Kelleher, 2003) While the results from these surveys provide an indication of participation rates, they do not provide a comprehensive profile of PA in Ireland that is relevant to the ACSM/CDC health-related guidelines This is because the ESRI survey focussed solely on sport and did not enquire about participation in other types of lifestyle PA, e g , occupational, commuting, housework (Fahey et al , 2004) Results from the SLAN survey also have limitations in gauging the health status of the population The threshold for participation in "regular exercise" was defined in the SLAN survey as "mild exercise, four times per week for 20 minutes or moderate exercise three times per week for 20 minutes" This is not relevant to or compatible with the evidence-based *minimum* guidelines established by the ACSM/CDC (Pate et al , 1995) and thus unreliably informs about (i) levels of PA in Irish adults and (ii) extrapolations of possible health benefits These findings also highlight the requirement for an Irish questionnaire that elicits FITT information about PA and is relevant to ACSM/CDC guidelines

Paramount to estimating potential dose-response health benefits in a sample population, is the assessment of energy expended during participation in PA Physical activity energy expenditure (PAEE) is known to vary according to the Frequency, Intensity, Type and Time (FITT) of the activity that is being performed (ACSM, 2000) The next section describes some of the current methods available to measure the FITT aspects of activity that contribute to the calculation of PAEE

2 3 Physical Activity Energy Expenditure (PAEE)

General health and well-being are partially determined by the relative contribution that both food and physical activity make to the overall energy balance equation The energy balance can thus be a positive or negative quantity, depending upon the relative contribution of the following

Equation 1. Energy Balance Equation

Energy Intake – Energy Expenditure = Energy Balance

The term "Energy Intake" refers to food consumed while "Energy Expenditure", encompasses resting metabolic rate (RMR), thermic effect of food (TEF) and physical activity (PA). Therefore a more detailed energy balance equation can be presented as:

Equation 2. Detailed Energy Balance Equation

Food Intake – [RMR + TEF + PA] = +/- Energy Balance

As the energy expenditure determined by RMR and TEF are mostly stable, the expenditure variable that exerts the most influence on the prevalence of obesity and other diseases, is physical activity (ACSM, 2000). It is this aspect of energy expenditure that is relevant to the issue of dose-response and requires measurement in order to determine the extent of the relationship between lifestyle physical activity and general health (Lamonte et al., 2001). Variables of interest when measuring PAEE are defined by those of the FITT principle. A combination of these principles provides a better reflection of the total energy cost of participation in the activity to the individual, in comparison to the assessment of one variable alone (Lee et al., 2000). Consequently, the measurement tool used to measure PAEE should have the capabilities to capture information pertaining to each of the FITT variables.

A lack of a gold standard measurement instrument, i.e., an instrument that can accurately measure all of the FITT variables for lifestyle PA, has occurred because of the challenge of capturing information about something that is so complex and prevalent in everyday life. Lifestyle PA can be undertaken in a variety of situational contexts, e.g., "occupational activities, discretionary activities, household tasks…and activities for physical fitness and health promotion", (Conway, Seale, Jacobs, Irwin, & Ainsworth, 2002). The diversity of these situations and the complexity associated with capturing information about some or all of the FITT properties of PA, has given rise to a variety of instruments that either directly measure PAEE or indirectly estimate it by extrapolating their results (Table 2.2).

Table 2 2 *Instruments used to measure physical activity energy expenditure (PAEE)*

Direct Measurement of PAEE	Indirect Measurement of PAEE
Calorimetry	Anthropometric measures
Doubly labelled water (DLW)	Metabolic measures
Motion Sensors (Accelerometer, Pedometer)	Heart rate telemetry
Observation	Self-report questionnaires
Diaries	Surveys
Logs/ Records	Fitness measures

Direct Measurement - The instruments that directly and objectively measure PAEE, include calorimetry, DLW and motion sensors However, these instruments are (i) relatively expensive and therefore not suitable for large-scale application, (ii) prone to instrument malfunction, and (iii) do not differentiate among types of activities, or intensity in the case of DLW (Ainslie et al , 2003, Freedson & Miller, 2000, Bassett, 2000) Other features of a surveillance instrument required for the measurement of PAEE in a population-based study, include accessibility and user-friendliness, so that sample numbers can be maximised and the burden to participants is minimised It should also permit freedom of movement, be socially acceptable and not be immediately noticeable so that it does not alter normal physical activity patterns (Allor & Pivarnik, 2001, Lamonte et al , 2001, Armstrong, 1998) The ability of some of the measurement instruments to satisfy all of these requirements is outlined in Table 2 3 overleaf

Table 2 3 *Comparison of different measurement instruments suitability to measure PAEE*

Properties	DLW	Accelerometer	Pedometer	Heart Rate	Self-report
Frequency	No	Yes*	No**	Yes*	Yes
Intensity	No	Yes*	No	Yes*	Yes
Type	No	No**	No**	No**	Yes
Time (Duration)	No	Yes*	No**	Yes*	Yes
User-friendly	No	Yes	Yes	Yes/No	Yes/No
Cost (€)	Very High	High	Low	Medium	Very Low
Can be re-used	No	Yes	Yes	Yes	No
Suitable for all ages	No	Yes	Yes	No	No
Large scale population use	No	No	Yes	No	Yes
Alters behaviour	No	No	Yes	Yes	No
Objective	Yes	Yes	Yes	Yes	No

Note DLW = Doubly Labelled Water Cost Very High = > €1,000, High = €400-900, Medium = €100-300, Low = €10-99, Very Low = < €10

* = Can estimate when accompanied by individualised regression equations, calculated following further testing in the laboratory

** = Possible when participant is compliant and makes written note of details of individual activities

Table 2 3 above provides a summary of the findings of various research studies that have used these instruments to facilitate the measurement of PAEE (Ainslie et al , 2003, Freedson & Miller, 2000, Bassett, 2000) Limitations to using heart rate (HR) monitors in the field have been identified Variations in baseline fitness level, age and gender can affect the HR-VO_2 relationship (VO_2 is the rate of oxygen consumed and is a direct measure of energy expenditure) Use of HR in the field has also been restricted by its inability to monitor FITT principles, without first conducting further laboratory tests to calculate individual regression equations This is necessary because HR is similarly influenced by other features such as emotion, caffeine, fatigue, dehydration and temperature (Davidson, McNeill, Haggarty, Smith, & Franklin, 1997, Haskell, Yee, Evans, & Irby, 1993, Strath, Bassett, Swartz, & Thompson, 2001) Further testing is similarly required for accelerometers in order to specify intensity of activity, as this is not readily given in the output from the instrument Accelerometers are also known to under and overestimate PAEE because of their inability to accurately measure upper body and static movement (Campbell, Crocker, & McKenzie, 2002, Chen & Sun, 1997, Leenders, Sherman, Nagaraja, & Kien, 2001) Pedometers and DLW are not acceptable methods of PA surveillance as

they do not measure any of the FITT principles and in the case of DLW, is very expensive to administer (Conway et al , 2002, Welk et al , 2000, Starling et al , 2002)

Indirect measurement - The use of a self-report questionnaire is reputedly the most practical means of ascertaining comprehensive information about physical activity behaviour, and indirectly about energy expenditure, in population based studies (Sallis et al , 2000, Kriska et al , 1997, Montoye, Kemper, & Saris, 1986) The relative merits and disadvantages of using self-report are quite diverse and also dependent upon the form of self-report employed (i e , interview, diary, questionnaire) Typically, self-report questionnaires are ideal for large-scale surveillance as they provide a speedy method of amassing a variety of pertinent data, are relatively inexpensive to distribute and analyse and can provide information on all aspects of PA, i e , frequency, intensity, type and duration (Sallis et al , 2000, Montoye et al , 1986)

The PA questionnaire that seeks information about intensity of activity can provide an estimate of whether the respondent is reaching the minimum recommended guidelines of participation for health, as stated by the ACSM/CDC (Pate et al , 1995) This has been done with other questionnaires, e g , The International Physical Activity Questionnaire (IPAQ), (Craig et al , 2003), 7-day Exercise Record, (Baranowski et al , 1999), Kaiser Physical Activity Survey, (Ainsworth et al , 2000b) In these questionnaires, PAEE was estimated either by (i) extrapolating the information about time spent in activity of at least moderate intensity and converting it to units of energy expenditure that have been established for each intensity (ACSM, 2000, Pate et al , 1995) These are 3 0-5 9 METs (4-7 kcal min $^{-1}$) for moderate intensity and > 6 0 METs (>7 kcal min $^{-1}$) for high intensity Alternatively, PAEE was estimated from the compendium of physical activities, which is a collection of daily and sporting activities that have been assigned MET values, following the calculation of MET values for 605 activities in field and laboratory settings (Ainsworth et al , 1993a, Ainsworth et al , 2000a) Limitations of using this compendium to estimate PAEE for general populations, reflect that data from the compendium was extracted from a homogeneous sample of adults The energy cost of PA does not account for environmental conditions and individuals with differences in body weight, age, gender, efficiency of movement (Ainsworth et al , 2000a)

The focus of this current research is to design a self-report recall questionnaire that can capture as much information as possible about physical activity, including energy expenditure The outcome data of this new questionnaire will provide an indication as to

whether respondents are meeting the established minimum health-related guidelines (Pate et al , 1995) The following section provides a detailed exploration of the type of information about PA that is provided by some of the physical activity questionnaires that are currently used in research and discusses whether they are adequate in meeting the requirements of a comprehensive physical activity-health surveillance instrument

2 4 Questionnaires used in Physical Activity Research

The following review compares some of the many questionnaires currently available for assessing amounts and types of PA participation For the purpose of maintaining clarity, questionnaires that have been selected for inclusion in this review are comparable to the proposed design of the questionnaire in this research (i e , the design of a self-administered, past-seven-day recall questionnaire, capable of measuring PAEE) This selected sample of questionnaires also provides a contrast between questionnaires that were constructed prior to, and after, the publication of the revised minimum ACSM/CDC recommended participation guidelines for health (Pate et al , 1995) Self-administered questionnaires with a recall period confined to within the past month and that satisfied the following inclusion criteria were selected

i Clarity in defining the variable of interest, i e , "physical activity"
ii Variety of FITT properties of PA identified
iii Test-retest reliability assessed
iv Validated against an objective criterion method (e g , accelerometer, heart rate monitor, calorimetry)

Questionnaires were identified through a review of published literature available on PubMed and Science Direct literature databases and also by a manual search of relevant journals and communication with the authors of PA questionnaires The inclusion criteria of the search, as described above, identified eight questionnaires from over 40 published PA questionnaires available for the adult population Details of each of the eight questionnaires are outlined overleaf in Table 2 4

Table 2 4 *Comparison of physical activity questionnaire features*

Questionnaire	Type	Administered	Recall Period	Intensity	Variable
Seven-Day Physical Activity Recall Questionnaire (Sallis et al , 1985)	L, O, H	Interview/self	Past 7 days	M, H, VH	F, I, T, Ty
Godin Leisure-time Exercise Questionnaire (Godin & Shephard, 1985)	L	Self	Usual week	L, M, H	F, I, Ty
International Physical Activity Questionnaire (Craig et al , 2003)	L, O, S, H	Self	Usual/Past 7 days	S, M, H	F, I, T, Ty
Kaiser Physical Activity Survey (Ainsworth et al , 2000)	H, O, S	Self	Sport=Year, House = Week	At least moderate	F, I, T, Ty
Lipid Research Clinics Questionnaire (Ainsworth, Jacobs & Leon, 1993)	L, O	Self	Usual (no time)	S, M, H	F, I, T
Paffenbarger PA Questionnaire (Paffenbarger, Wing & Hyde, 1978)	L	Interview/self	Past week or year	L, M, H	F, I, T, Ty
Scottish Physical Activity Questionnaire (Lowther et al , 1999)	L, O, S, H	Self	Past 7 days	At least moderate	F, I, T, Ty
Baecke Questionnaire of Habitual Physical Activity (Baecke, Burema & Frijters, 1982)	L, O	Self	Usual (no time)	S, M, H	F, I, T

Note *Type* L=Leisure, O=Occupational, H=Housework, S= Sport *Intensity* S=Sedentary, L=Light, M=Moderate, H=Heavy, VH=Very Heavy *Variable* F=Frequency, T=Time, I=Intensity, Ty=Type

The diversity of questionnaires available for the assessment of PA is apparent from this review The eight questionnaires outlined above are comparable in terms of (i) their clarity in defining the variable of interest (i e , physical activity), (ii) FITT principles that are measured, and the methods in which (iii) reliability and (iv) validity of the questionnaires were assessed

2 4 1 *Defining PA in Questionnaires*

The lack of understanding of fundamental words in a PA questionnaire has been highlighted as a possible cause of inaccurate reporting Tudor-Locke and colleagues (2003) investigated the meaning and understanding of phrases such as "physical activity", "leisure" and "exercise" among a group of African-American women aged at least 40 years (N=196), in a series of focus groups Results showed that words like "leisure", "exercise" and "physical activity" were poorly and not uniformly understood by this group Descriptions of these terms by the women differed greatly from the academic definition of the same term For example, these women reported that "physical activity" is typically considered to be "structured exercise" and not incidental activities of daily living This conflicts with the intended meaning of many recently developed PA questionnaires that include activities like housework and childminding (e g , Kaiser PA Survey for women)

In the national SLAN survey of health related behaviours in Ireland (Friel et al , 2003), the phrase "physical activity" was not defined Therefore, it was unclear what specific meaning the investigators attached to it and whether this matched the understanding of the target population Also, in the survey, "physical activity" was used synonymously with the word "exercise" In academic terms, "exercise" is defined as " a subset of physical activity that is planned, structured, repetitive and purposive", (Caspersen et al , 1985) Ambiguous wording of PA related terms can cause misinterpretation of the question, which may affect the accuracy of results (Durante et al , 1996, Tudor-Locke et al , 2003) It is possible therefore that the results of the SLAN survey are not representative of the true participation levels in Ireland and thus extrapolations of these results to predict future health consequences would be questionable

Of the questionnaires referred to in Table 2 4 the Paffenbarger Physical Activity questionnaire (Paffenbarger, Wing, & Hyde, 1978) offers no explanation of what is meant by "physical activity" Furthermore, the Lipid Research Questionnaire (Ainsworth, Jacobs, & Leon, 1993b) states that the information requested about physical activity is "self-

explanatory" These questionnaires also omit to offer the respondents any examples of physical activity events to provide some guidance about the type of answer that is required In comparison, other questionnaires, such as the Godin Leisure-Time Exercise Questionnaire (Godin & Shephard, 1985) and the Kaiser Physical Activity Survey for Women (Ainsworth et al , 2000b), afford examples of activities for different intensities with the Godin Leisure-Time Exercise Questionnaire also providing examples of associated physiological changes with varying intensities The Scottish Physical Activity Questionnaire (Lowther et al , 1999) provides more comprehensive directions to the respondents by offering them an explanation of the types of activities to be considered (exercise, sport and general), while also giving examples of activities within each of these domains

Studies undertaken to construct and generate items for inclusion in a PA questionnaire, particularly details of qualitative studies that explore topic understanding, cultural relevance and understanding of the questionnaire items, are rarely published (Biddle et al , 2001) Research relating to the methodology used to develop and design PA questionnaires is similarly limited (Durante et al , 1996, Friedenreich et al , 1998) Problems associated with subjective interpretation of the question and concerns related to cultural differences understanding of PA, suggest that researchers should describe the processes undertaken to design the questionnaire Warnecke and colleagues (1997) used a model of the cognitive response process (Tourangeau, 1984) to explore question wording in culturally diverse populations for interview surveys (N=426, aged 18-50 years) Six different ethnic and cultural backgrounds were investigated Findings showed that race/ethnicity influenced interpretation of the question, in the absence of specific cues to explain These results highlight the need for culturally specific questionnaires which have been designed and pre-tested by involving the target population and include a definition of the variable of interest, to reduce misunderstanding and ambiguity

2.4.2 FITT principles measured in PA Questionnaires

(a) Type

Questionnaires developed prior to the 1995 ACSM/CDC revised position stand are likely to be less sensitive to moderate intensity physical activities and thus not reflective of the lives of most women (Ainsworth, 2000). This also makes them less relevant to the revised health-related paradigm. Questionnaires developed previous to the 1995 position stand, have tended to focus on occupational and sport activities that demand an intense exertion, and often fail to describe the extent of physical activity in women and minority populations like the elderly and various ethnic groups (Sallis et al., 2000). This is reflected in the questionnaires outlined in Table 2.4. above, where all three of the "post 1995" questionnaires include housework as an activity but only one of the questionnaires developed prior to 1995 includes it. Inaccurate profiles of physical activity are often depicted for women and minority groups who are inadvertently perceived as being "inactive" or "sedentary". Inaccurate PA profiles can result from the occurrence of a "flooring effect", which occurs when the lower end of the physical activity continuum is not accounted for. When this happens, the lowest score available is too high for some respondents. This causes an inaccurate inflation in one category (sedentary/inactive level) and a complete omission of representative data in another (low-moderate intensity). This has been a fundamental failure in the self-report questionnaires developed prior to the ACSM/CDC health guidelines, that focussed upon sport and other structured activities (Tudor-Locke & Myers, 2001).

A review of eight leisure-time activity questionnaires found that the gender-gap in participation is related to the rigidity of the definition of activity and the context of the questionnaire (Stephens, Jacobs, & White, 1985). If, for example, the questions related only to sport, males were found to be dramatically more active than women, while differences in activity levels appeared less diverse between the genders than when more moderate intensity activities such as walking and jogging were also included. This may be indicative of the era in which the questionnaires were constructed (early 1980's), before the shift away from fitness to a more health centred paradigm. During this time, moderate-high intensity activities were analogous with health and fitness. Post 1995, there has been a suggestion that "...the lack of consideration of the [physical activity] context can lead to underestimates of prevalence", (Sallis & Saelens, 2000, p. 4). The process of classifying PA

into various settings enables a more holistic, thorough and pragmatic opportunity to measure energy expenditure because PA is being measured in as many regular, daily settings as possible If the circumstance in which the activity is being measured is not appropriate to the population that is being investigated, measurement errors are inevitable (Armstrong, 1999) The development of a new PA questionnaire should therefore profile PA in a comprehensive method, by being inclusive of a variety of contexts of PA that are representative of the target population It should also be relevant to the 1995 ACSM/CDC health-related guidelines by being inclusive of activities of at least moderate intensity

(b) Intensity

Each of the eight questionnaires in Table 2 4 include questions about activity of at least moderate intensity The Seven-day PA Recall Questionnaire (Sallis et al , 1985) is the only questionnaire to include a "very high" intensity category in addition to "moderate" and "high" A description of the "very high" category was given as "how you would feel if you were running", while "moderate intensity" was "how you would feel while you were walking", and "high intensity" was described as "somewhere in between both" (Sallis et al , 1985) These descriptions were provided as guidance to respondents as an alternative to providing symptoms of physical exertion It was decided by the authors that symptoms would be misleading because they are dependent upon individual fitness and environmental conditions A lack of documentation to state otherwise, suggests that this opinion was not formed following the result of investigative research among the general population More research is required to understand how people understand and describe the physiological and psychological responses to participation in PA of different intensities The Kaiser PA Survey for women (Ainsworth et al , 2000b) and the IPAQ (Craig et al , 2003) use only one physical descriptor each to describe the response to PA These are "sweating" and "breathing" respectively, with the difference between moderate and high intensity separated by the word "heavy" in front of the descriptor, e g , heavy breathing Other questionnaires, such as the SPAQ (Lowther et al , 1999) and Godin Leisure-time PA Questionnaire (Godin et al , 1985) provide a more comprehensive list of possible responses to PA to assist their respondents

(c) Frequency and Duration

Each of the eight questionnaires in Table 2.4. measure frequency and all but three also measure duration of activity. Both of these properties are important for the purpose of assessing whether respondents are participating in enough PA to incur minimum health benefits (Pate et al., 1995). Information about the duration of activity performed is particularly relevant to the 1995 guidelines, as it is the total amount of activity that the participant engages in that is indicative of health benefits incurred. It is perhaps significant then, that the three questionnaires that do not specify information about duration, i.e., Baecke Questionnaire of Habitual Physical Activity (Baecke, Burema, & Frijters, 1982a), Godin Leisure-time PA Questionnaire (Godin et al., 1985) and Lipid Research Clinics Questionnaire (Ainsworth et al., 1993b), were developed prior to the 1995 ACSM/CDC guidelines. For the purposes of PA health behaviour surveillance, questionnaires should include questions pertaining to duration and frequency of participation.

2.4.3. Reliability of PA Questionnaires

One issue raised by researchers in reliability testing has been (i) the use of a sample which is non-representative of the general population, which provides higher values due to the homogeneity of the group (Booth, Owen, Bauman, & Gore, 1996; Pols, Peeters, Kemper, & Grobbee, 1998; Richardson, Leon, Jacobs, Ainsworth, & Serfass, 1994). Reliability testing of the Minnesota Leisure Time Physical Activity Questionnaire by Richardson and colleagues (1994) used a study sample consisting of faculty staff and students from the university community and their families and friends (N=78, age 20-59, F=64%). The authors of that research noted that this degree of homogeneity in educational status within the sample was a limitation of the study, as it was unrepresentative of the educational level of the general population. Participants used in the test-retest reliability of the SPAQ (N=34, age 33+/-12 years, Male=26%) were all recruited from an exercise promotion project or aerobics class (Lowther et al., 1999). The results from this convenience sample may be difficult to generalise to sedentary or less regularly active people. Reliability will be affected if the context of the questionnaire is inappropriate for the sample participants and it is also difficult to generalise the results for other populations that have not been tested (Ainsworth, Richardson, Jacobs, Leon, & Sternfeld, 1999). For questionnaires that are intended for population-based studies, a sample that is more representative of the general population is preferential.

The time-frame between the test-retest administrations varies between questionnaire reliability assessments (See Table 2 5) A low reliability co-efficient should be indicative of the poor measurement characteristics within a questionnaire and therefore indicate its worth in prospective research However, a low reliability may also result from a true variation in the activities of the respondents since the last time the questionnaire was completed, due to seasonal influences or change in work/family commitments In order to eliminate avoidable measurement errors such as seasonal change in participation rates, test-retest administrations should be completed within a short space of time, and ideally overlap by 3-4 days in order to account for the possibility of activity variation In a study that looked at test-retest timeframes, Booth et al (1996) compared the retest reliability of adults (N=116, age 18-79, Male=47 4%) in recalling their participation in PA over two separate two week periods, against the retest reliability of adults (N=115, age 18-79, Male=48 7%) over the same 2 week period In the second group, participants were administered the second questionnaire only 3 days after the first one

The results of the study indicated a higher correlation (r = 0 86) for the same two-week period, compared to the two different two-week periods (r = 0 58) It is possible that the less favourable results for the different-period group have resulted from a natural variation in actual physical activity and not due to any error associated with the actual questionnaire Findings from this study suggest that variability in actual physical activity patterns will account for poor reliability scores, when an overlapping period is not accounted for in the retest timeframe A further recommendation from this study is that the time between test and retest should also not be too short This is to prevent respondents merely recalling what they previously wrote, as opposed to recalling the activities performed, although the optimum recall timeframe does not appear to have been tested (Booth et al , 1996, Durante et al , 1996)

Details of the reliability methods undertaken for the eight questionnaires discussed earlier m this review are outlined in Table 2 5 overleaf The Scottish Physical Activity Questionnaire (SPAQ) is the only measurement tool listed here to have employed an over-lapping technique in its test-retest reliability study, using a 4 day overlap (Lowther et al , 1999) The International Physical Activity Questionnaire (IPAQ) used an eight day retest period (Craig et al , 2003) This meant it had no overlapping time as it was recalling the past-seven-days All others have used a one month time-frame, including the Paffenbarger PA Questionnaire which was also retested again after 9 months The influence in seasonal

variation and different PA patterns over longer time frames is evident from the variation in correlations reported, i e , r = 0 72 (1 month) and r = 0 43 (9 months), (Ainsworth et al , 1993b)

The reliability of individual intensities of activity has not been consistently reported for the eight PA questionnaires in Table 2 5 The IPAQ stated that the high intensity had higher values than moderate intensity, but did not detail the exact scores (Craig et al , 2003) Neither the Kaiser PA Survey for women or the SPAQ conducted an analysis by intensity, because the questionnaires were not designed to extract information about individual intensities (Ainsworth et al , 2000b, Lowther et al , 1999) Results from both the Seven-day PA recall questionnaire and Godin Leisure-time PA Questionnaires (see Table 2 5 overleaf), indicated that moderate intensity is recalled less reliably than high intensity (Sallis et al , 1985, Godin et al , 1985) No explanations are offered for the relatively poorer performance of moderate intensity, thus further research is required to investigate the specific properties of moderate intensity that are involved in the recall of this level of intensity This type of information would be important because of the prevalence of moderate intensity activity in everyday life and due to its relevance to the ACSM/CDC guidelines (Pate et al , 1995)

Table 2 5 *Reliability and Validity properties of published questionnaires*

Questionnaire	Study	Validity Criterion	Validity Score	Sample	Test-Retest Timeframe	Reliability Score	Sample	Recall Period
Seven-Day Physical Activity Recall (Sallis et al , 1985)	Sallis et al , 1985	Accelerometer (kcal)	Pearsons r Total=0 53 Moderate=0 11 Hard=0 67	N=53 Male=54%	2 weeks	Pearsons r Total=0 34 Moderate=0 07 Hard=0 3 Very hard=0 6	N=53 Male=54%	Past 7 days
Godin leisure-time exercise questionnaire (Godin & Shephard, 1985)	Godin & Shephard, 1985	VO_{2MAX}	Pearsons r Total activity =0 24 Strenuous=0 38 Moderate=0 03	N=306 Male=54%, Age 18-65	2 week	Pearsons r Total = 0 74 Moderate = 0 46 Strenuous = 0 94	N=53 Age 18-65	Usual week
	Jacobs et al , 1993	Accelerometer 4 week diary VO_{2MAX}	r=0 32 (total PA) r=0 36 (total PA) r=0 56 (total PA)	N=77, Male =34% Age=20-59	1 month	Light = 0 24 Moderate = 0 36 Strenuous = 0 94	N=77, Male =34% Age=20-59	Usual week
International PA Questionnaire-Long (Craig et al , 2003)	Craig et al , 2003 Across 12 countries	Accelerometer (total counts)	Spearmans p =0 33	N=744 Male= ~50% Mean age=25-49	8 days	Spearmans p =0 77 (average of all participating countries)	N=294 Male= 50% Mean age=25-49	Usual and past 7 days

Table 2 5 (continued) *Reliability and Validity properties of published questionnaires*

Questionnaire	Study	Validity Criterion	Validity correlations	Sample	Test-Retest time	Reliability score	Sample	Recall Period
Kaiser Physical Activity Survey (Ainsworth et al , 2000)	Ainsworth et al , 2000	Accelerometer PA records VO_{2peak}	Spearmans rho Caregiving = 0 3 Occupational = 0 44 Caregiving = 0 03 Occupational = 0 64 Caregiving = 0 01 Occupational = 0 04	N=50 All women Age=18-60	1 month	Intraclass House &Caregivmg r= 0 8 House r = 0 79 Sport r = 0 84 Occupational r = 0 85	N=50 All women Age=18-60	Caregiving =1 week All others = year
Lipid Research Clinics Questionnaire (Ainsworth et al , 1993)	Ainsworth et al , 1993	Pearsons r Accelerometer (kcal) VO_{2MAX}	 Total = 0 2 Total = 0 49	N=73 Male=36% Age=20-59	1 month	Pearsons r Total activity =0 85	N=73 Male=36% Age=20-59	Usual activity
Paffenbarger PA questionnaire (Paffenbarger et al , 1978)	Ainsworth et al , 1993	Accelerometer (kcal) VO_{2MAX}	 Total = 0 29 Total = 0 6	N=73 Male=36% Age=20-59	1 month 9 months	0 72 0 43	N=73 Male=36% Age=20-59	Past week or year

Table 2 5 (continued) *Reliability and Validity properties of published questionnaires*

Questionnaire	Study	Validity Criterion	Validity Score	Sample	Test-Retest time	Reliability Score	Sample	Recall Period
Scottish Physical Activity Questionnaire (Lowther et al , 1999)	Lowther et al , 1999	Accelerometer (kcal)	Pearsons r Total = 0 34	N=30 Female=70% Age=35+/- 13 5	4 day overlap	Pearsons r =0 99, coefficient of repeatability = 53 minutes	N=34 Female=74% Age=33+/- 11 1 years	Past 7 days
Baecke Questionnaire of Habitual Physical Activity (Baecke et al , 1982)	Baecke et al , 1982				3 months	Pearsons r Work = 0 8 Leisure = 0 74 Sport = 0 9	N=306 Age = 20-32 Male = 45%	Usual activity (i e no time frame)
	Jacobs et al , 1993	Accelerometer (kcal)	Pearsons r Total=0 08 Work=0 03 Leisure=0 11 Sport=0 02	N=73 Male=36% Age=20-59	1 month	Pearsons r Total-0 93 Work=0 78 Leisure=0 86 Sport=0 9		

2 4 4 Validity Of PA Questionnaires

Cardio-respiratory fitness is weakly related to physical activity questionnaires that measure habitual or recent (past week) activity (Boreham et al , 2001, Paffenbarger, Blair, Lee, & Hyde, 1993) This means that if the outcome of interest in a new questionnaire is related to information about physical health, rather than fitness, then equipment used to measure changes in the parameters of physical fitness over time (e g , VO_{2MAX}), is not appropriate Other instruments, that are more sensitive to monitoring activity levels that have been recently performed, at a lower intensity and are beneficial to health, include the accelerometer and heart rate monitor (Paffenbarger et al , 1993) A new questionnaire should therefore be validated with a measurement device that is related to the outcome of interest, so that its results can be evaluated appropriately (Bouchard, 2001)

Four questionnaires have used a measure of cardio-respiratory fitness for validation (See Table 2 5) Of these, three were developed prior to the 1995 ACSM/CDC paradigm shift from fitness to health (Pate et al , 1995), i e , Godin leisure-time exercise questionnaire (Godin & Shephard, 1985), Paffenbarger PA questionnaire (Paffenbarger et al , 1978) and Lipid Research Clinics questionnaire (Ainsworth et al , 1993) The focus of these questionnaires on activity more relevant to improving fitness rather than health is demonstrated in the superior correlations for VO_{2MAX} (a measure of cardio-respiratory fitness) over the accelerometer In contrast, the Kaiser Physical Activity Survey for women, which was developed in 2000 and includes questions about housework and caregiving, found a lower correlation between VO_{2peak} and the caregiving component of the questionnaire (r=0 01), compared to an accelerometer (r = 0 3) Accelerometers are a reliable method of measuring daily lifestyle activities, particularly those performed at a moderate intensity All of the questionnaires developed since 1995 have been validated with an accelerometer and demonstrated a correlation of between 0 3 and 0 4

A limitation of accelerometers as a criterion measure is their tendency to underestimate PAEE in field studies, particularly for activities that involve upper body movement, an increased gradient or increased velocity (Nichols et al , 1999, Jakicic et al , 1999, Welk et al , 1995) The output data of accelerometers is in units of energy expenditure and time, thus it can directly provide information about two aspects of FITT (i e , duration and intensity) The output also provides information about the magnitude of force exerted by the participant, called vector magnitude counts The Tritrac-R3D

accelerometer was tested against a portable metabolic (VO_2) instrument to measure PAEE at moderate intensity, for several recreational and household activities (Hendelman, Miller, Baggett, Debold, & Freedson, 2000) Participants (N=25, F=60%, age=30-50 years) completed walking, played golf and performed indoor and outdoor household chores Results indicated a strong relationship between VO_2 (METs) and accelerometer (vector magnitude counts) for walking (r=0 89) and combined other activities (r=0 62) This study also noted an underestimation of PAEE (Kcals) by 30-50 % in activities other than walking In contrast, Nicholls et al (1999) assessed the reliability and validity of the Tritrac-R3D accelerometer during walking and running Results showed that Tritrac-R3D consistently over estimated energy expenditure (Kcals) during horizontal treadmill walking and was not sensitive to changes in gradient These studies highlight the limitations of Tritrac-R3D in measuring PAEE in activities that are high intensity (or on a gradient), involve upper body movements and when validated with Kcals instead of vector magnitude counts

In the context of the current research, i e , a questionnaire designed to measure intensities across the continuum of moderate–high, the criterion validation equipment should reflect the multi-dimensional (FITT) properties of a behaviour like PA Limitations to equipment that are available, such as accelerometers (Freedson et al , 2000, Baranowski & de Moor, 2000, Bassett, 2000, Bassett et al , 2000) and heart rate monitors (as described in Section 2 3 , p 7) means that no one instrument has been predominant in validating new PA questionnaires The combining of two pieces of equipment that can compliment each others limitations, e g , an accelerometer and a heart rate monitor, has recently been suggested by Ainslie et al , (2003) as a novel method of validating questionnaires This method requires further investigation and is an objective of this current research

In summary, the choice of questionnaire for a specific study should be relevant to the research question, as well as the characteristics of the target population The evaluation of its reliability should include an overlapping period of days, so that the psychometric properties of the questionnaire are assessed and not the genuine variability in participation levels The new questionnaire should also have been validated with an appropriate criterion measure that can adequately measure lifestyle PA, at different intensities and that is, in itself, a valid instrument Questionnaires that do not suit the exact requirements of the researcher have been adapted to suit the requirements of the researcher The Baecke Habitual Physical Activity Questionnaire (Baecke, Burema, & Frijters, 1982b) has been adapted for recent epidemiological studies, to produce two different questionnaires These

are (i) the Kaiser Physical Activity Survey for women (Ainsworth et al., 2000, see Table 2.4.) and (ii) the ARIC/Baecke Questionnaire to assess leisure-time PA (Richardson et al., 1995). Similarly, the SPAQ was adapted from the Stanford Usual Physical Activity Questionnaire (Blair et al., 1985).

Adapting a previously validated questionnaire for another population requires the researcher to ensure validity and suitability of the new questionnaire to the target population. However, documentation of the processes undertaken to develop new and adapted PA questionnaires have been scarce (Friedenreich et al., 1998). The difficulty in recalling certain types of information about PA has prompted the application of a model from cognitive psychology to help in the design stage of the questionnaire. This cognitive model has been applied to the design of other types of questionnaires, e.g., socio-cultural (Warnecke et al., 1997), food (Subar et al., 1995) and socio-economic (Belli et al., 2001). It has also been recently applied to the design of a Lifetime PA questionnaire (Friedenreich et al., 1998) and advocated for use in the design of a short term recall (e.g., past week) PA questionnaire (Durante et al., 1996). This represents a novel approach to designing a past-seven-day PA recall questionnaire and is the focus of this current research. The following section outlines the structure and processes of memory, involved in recalling information about physical activity.

2.5. The role of autobiographical memory (ABM) in the recall of physical activity

Questionnaires that request details from the respondents' personal memory, (i.e., personally experienced and remembered events), are classified as being dependent upon the autobiographical memory for answers (Conway, 1996). Autobiographical memory (ABM) is defined as "... a memory for information related to the self", (Brewer, 1996). It is further explained by Conway and Rubin (1993) as "a memory for the events of one's life [that] relate to the self, emotion, personal meanings and goals", (p. 103). Memories from the autobiographical memory structure take longer to retrieve than other types of memories (e.g., a scientific formula, a poem) because they need to be constructed from various aspects of the personal self rather than merely reproduced. In this way, autobiographical memories form a compilation of constructions of knowledge about the self (Conway et al., 1993; Conway, 1996).

For Conway (1996), autobiographical memories are "...conceived as temporary mental representations constructed and maintained by a set of central processes such as the

central executive of the working memory", (p 67) According to Baddeley , working memory is a replacement concept for the short-term memory (Baddeley, 1996) It has three components, the central executive which deals with cognitively demanding tasks, is the key component The other two components (the phonological loop and the visuo-spatial sketch pad) are used by the central executive for specific purposes, i e , the presentation of word order and the storage and manipulation of visual and spatial information Conway (1996) and others (Anderson & Conway, 1993, Tourangeau et al , 2000, Williams & Hollan, 1981) believe that ABM's are constructed on the basis of knowledge sampled from the autobiographical knowledge base, a process initiated by the presentation of a pertinent "cue", e g , a face, a song, a smell, a question

The type of information contained within ABM varies in relation to (i) specific events, (ii) general knowledge about categories of events, and (iii) knowledge about extended or major intervals in ones' life, (Tourangeau, Rips & Rasinski, 2000, Conway, 1996, Conway & Rubin, 1993) Theorists tend to agree on how such information is organised For example, they agree that ABM reflects upon both the *structure* of the memory system and also the *processes* operating within the structure (Eysenck et al , 2001) Areas of contention between theorists however centre around the different types of units that store the autobiographical information and also the method in which different units are connected within the autobiographical structure to produce a memory (Tourangeau et al , 2000) The following section reviews firstly (i) models of the structure of autobiographical memory and secondly, (ii) the process of accessing memories within that structure (particularly in relation to answering a question in a questionnaire)

2 5 1 *Autobiographical memory structure*

The concept of "self" plays an influential position in encoding (understanding) and retrieving autobiographical information from memory (Eysenck et al , 2001, Tourangeau et al , 2000) One function of human memory is to assess the progress of personal goal attainment These goals are believed to be structured in a hierarchy that forms part of the "working self" (Conway, 2001) Autobiographical memory formation is a function of prior knowledge, personal importance and affect When an event is judged to be highly self-relevant, encoding processes co-ordinate to rapidly integrate the event and its memory details into long term memory (Conway, 1996) Events lower in self-relevance may not promote such co-ordination of encoding processes, thus leading to less extensive

integration of information. This results in a memory that is incomplete or lacking in detail. ABM is a type of memory that persists over hours, days, weeks, months and years and it retains knowledge of the self at different levels of abstraction (Conway, 1996).

For the purpose of designing this past-seven-day recall PA questionnaire, which is concerned with activity at different intensities (to facilitate comparison with ACSM/CDC (1995) guidelines), it is necessary to understand how recently performed activity that lasted minutes to hours is structured within memory and to understand which "self-related" cues are most effective in optimising recall of those activities.

2.5.1.1. Tulving's model of autobiographical memory structure (Tulving, 1983)

This model of ABM structure suggests two distinct memory systems. These systems are (i) "episodic memory", which contains units of specific autobiographical events, i.e., a particular holiday, and (ii) "semantic memory", which contains more general knowledge about situations, i.e., holidays in general. Knowledge in semantic memory is organised in terms of general concepts rather than actual events and is thought to exist at a very abstract level. Episodic memory on the other hand is organised in a very loose temporal way, without any general knowledge. Figure 2.1. illustrates application of Tulving's model to an everyday physical activity, i.e., walking to work. For example, thinking about walking to work triggers a memory of a particular day when the person walked to work and met John on the way. Information relating to the earlier event is stored with information about the later event (as they both relate to the subject of "walking to work"). When the event needed to be recalled later, the temporal (time-related) properties associated with episodic memory would assist the person to assess which event came first. Tulving's model suggests that further general information about this and other walking events must be completed by the semantic memory, e.g., if I was walking to work, I must have had my laptop and umbrella with me. Thus, semantic memory acts as a supporting mechanism to episodic memory by filling in the memory with the usual information that is associated with general events. In this way, even though episodic and semantic memories are completely distinct, both are required to answer a question about a specific event.

Tourangeau, Rips and Rasinski (2000) reported that experimental data using this model is limited as it was criticised for providing a very simplistic account of the structure and storage of autobiographical memories. It has also been criticised for failing to explain fully how specific episodic memories link to the general and abstract information in the

separate semantic memory system (McCauley, 1984; Hintzman, 1984). Later models (see below) suggest that abstract information and specific memories are in fact part of the same structure, a theory that Tulving himself later also subscribed to (Tulving, 1984).

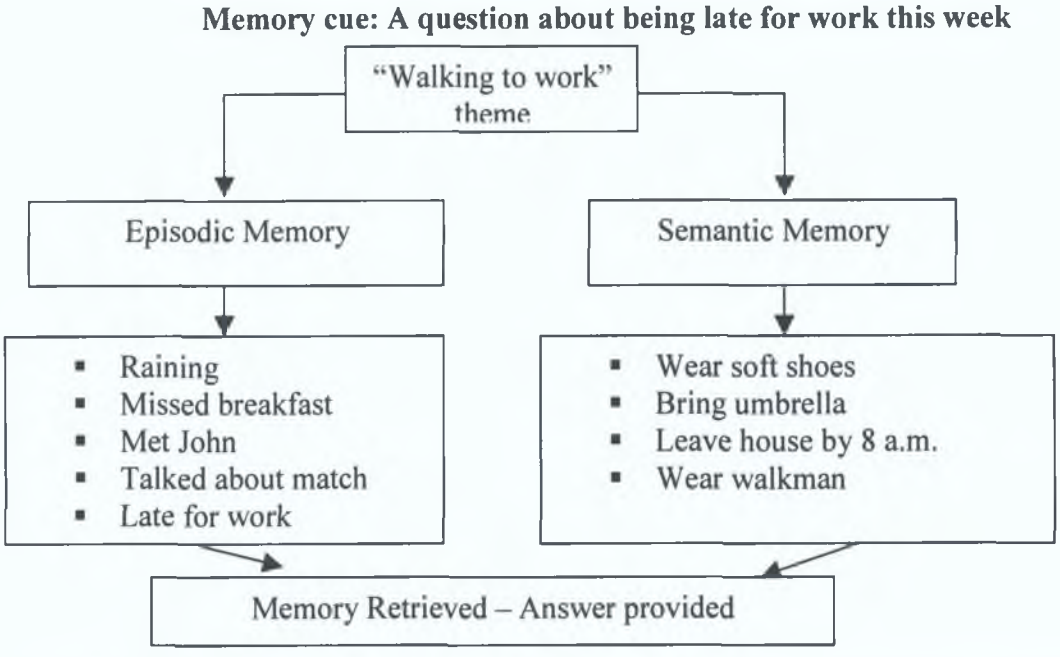

Figure 2.1. A schematic representation of Tulving's (1983) model of autobiographical memory structure and retrieval process.

2.5.1.2. Kolodner's model of autobiographical memory structure (Kolodner, 1985)

In contrast to Tulving (1983), Kolodner suggests a more integrated model of ABM structure. This model proposes the existence of distinct sections within ABM that accommodate (i) general information about categories of events, and (ii) information about specific and unique events. However, Kolodner suggests that a hierarchical connection exists between events.

Kolodner (1985) outlines her theory of ABM structure in a computer model called CYRUS. The design of CYRUS is analogous to that of a filing cabinet, in that it contains the title of an event on the outside, while inside are folders which have the description of many different types of specific events filed within them. Following a six year diary study of his own memory, Wagenaar (1986) also proposed the presence of a "filing cabinet" structure of autobiographical memory, containing a large number of cards with different tags. Wagenaar's study involved recording details of daily events for six years and then recalling a random selection of them, over the same six year period. Results from his study

suggested that the detailed information on the "filing cards" selected during the retrieval of a memory, encouraged the further retrieval of another selection of cards with more specific information In Kolodners' model, each "file" is called an Event Memory Organisation Packet (E-MOP) and contains information about various events or categories of events

A feature of Kolodner's model (1985) is the top-down direction for accessing retrieval of complete autobiographical memories Each category of event has a collection of indices of related details to that category, with each of these indices having sub-indices Figure 2 2 overleaf applies Kolodner's model to a football player asked to recall a goal he scored The initial index is accessed via cues from the question asked, i e "Have you scored any penalties in any Cup games this year?" The cue "goal" leads to an E-MOP of "goal types" This leads to sub-ordinate E-MOPs of various categories of types of goals, which leads to further subordinate E-MOPs of individual incidences of goals within each of those categories Having identified the "penalties" E-MOP, the player then accesses a subordinate E-MOP that corresponds specifically to Cup games as opposed to league games, before he arrives at an answer In this way, access to relevant information is hierarchical and obtained, via links, to associated information about the event A further feature of Kolodner's model is that, similar to Tulving (1983), there are two distinct types of units The higher E-MOPs contain generic information about categories of events while the lower E-MOPs contain specific details about the individual event Unlike Tulving's model though, Kolodner proposes a network of links between the generic and specific units

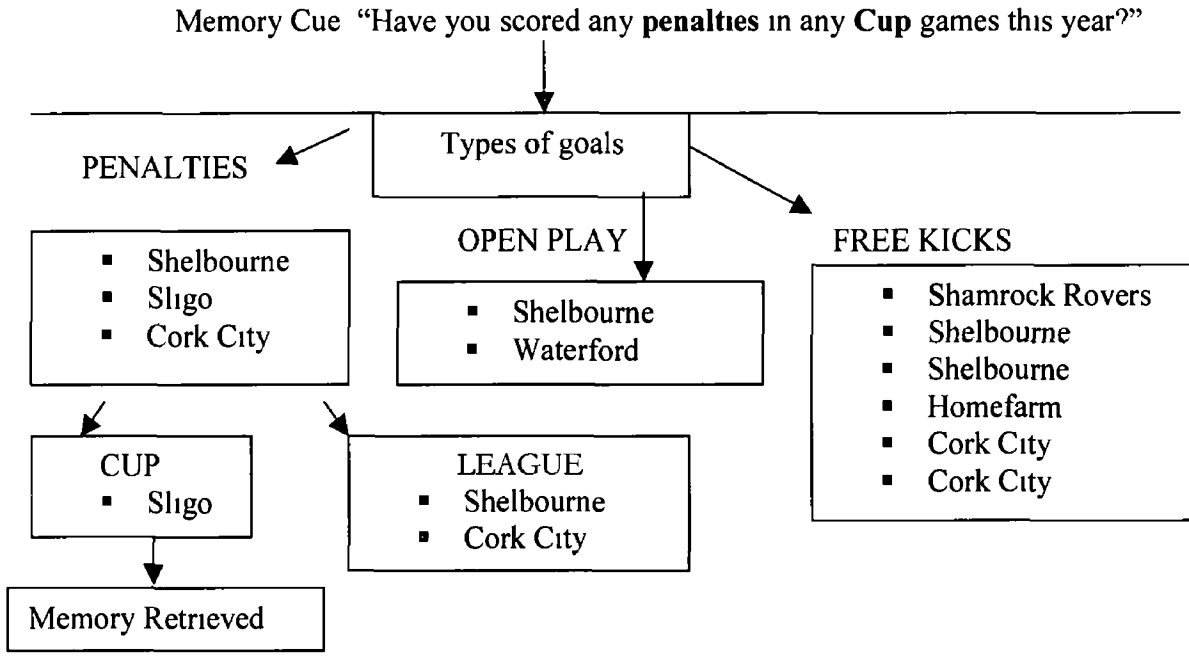

Figure 2 2 Schematic representation of Kolodners (1985) model of autobiographical memory structure and retrieval process

Empirical data (Barsalou, 1988, Linton, 1982) have supported Kolodner's suggestion that access to memories is via general index In a cued-recall study at the end of the first semester, Barsalou (1988) requested participants (N=78, age 18-22) to recall specific events from the previous summer Results produced disrupted and incomplete memories Participants were more likely to naturally refer to general summaries of events, without specific details of individual events Linton (1982) reported a six year diary study of her own memory, in which she wrote detailed accounts of events as they occurred and later used them as cues to encourage recall of the dates of those specific events Part of the criteria of events being recorded was that they had to be unusual and salient, in order to permit remembering at a later date However, at a later date, when given the detailed accounts of some events and asked to recall the dates and order of the events, she was only able to access very generalised memories, without the specifics Both of these studies indicate that initial access to autobiographical memory is via general index, as advocated by Kolodner (1985) However, they also proffer criticisms of the model's simplicity of design and its lack of flexibility in the retrieval of specific memories Results from Linton's diary study in particular shows that Kolodner's model does not appropriately take account of

temporal (time-related) dimensions between events This element of storage related to the time of an event is explained in Conway's model of ABM (1993,1996)

2 5 1 3 Conway's model of autobiographical memory structure (1993, 1996)

Similar to Kolodner (1985), Conway has assigned a name to the knowledge structures within ABM These structures are known as Autobiographical Memory Organisation Pockets (A-MOPs) However, Conway proposes a three-tiered structure of ABM, as opposed to two as described by Kolodner (1985) and Tulving (1983) This model is also hierarchical in structure, illustrating levels of memories that vary in their scope and detail of events (see Figure 2 3 overleaf) However, Conway's model advances on Kolodner's model by proposing that memories are organised in terms of temporal and thematic pathways and that it is these pathways that are the guiding forces in retrieving memories Firstly, a description of Conway's (i) proposed structure of ABM is discussed, followed by (ii) the processes of retrieval within that structure

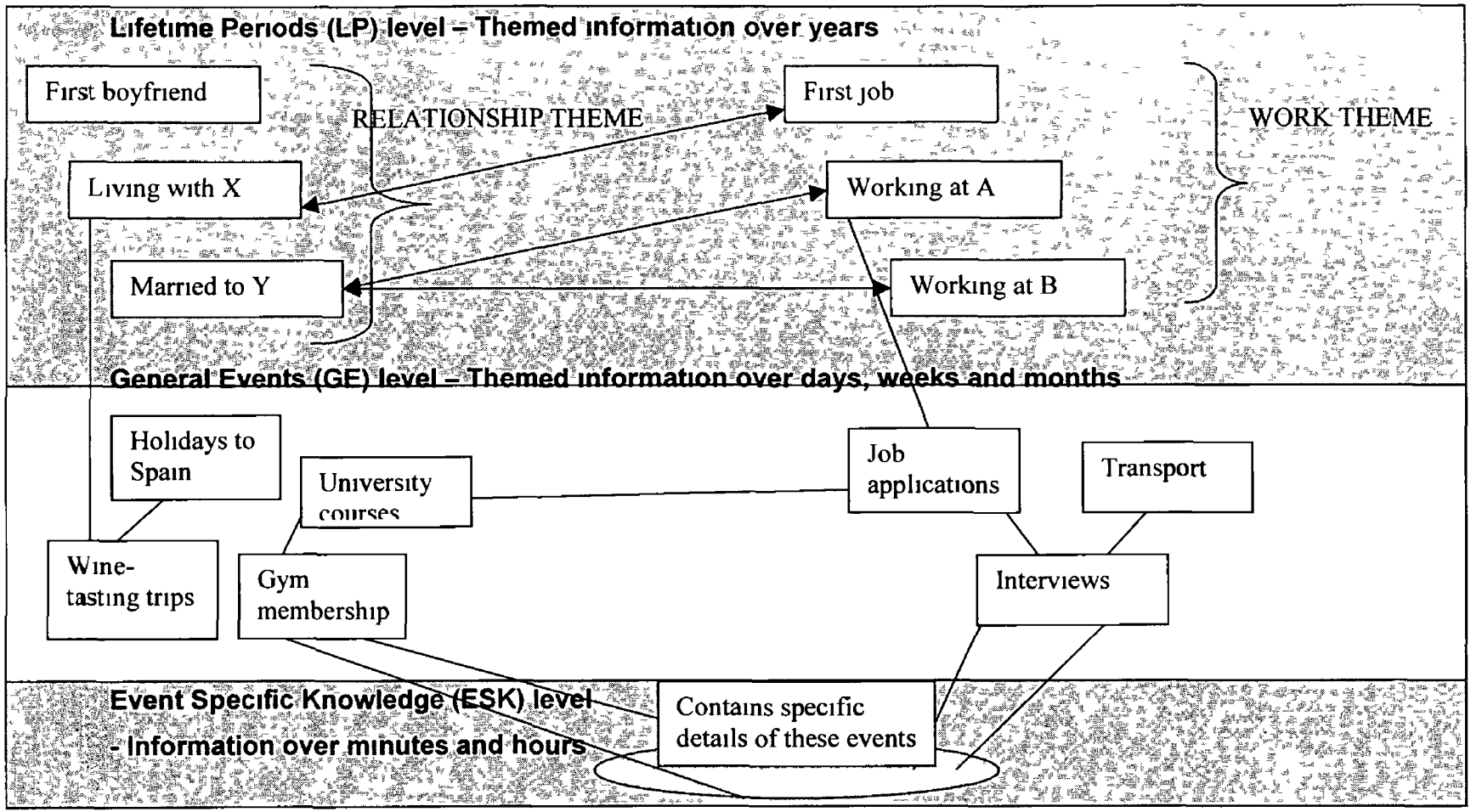

Figure 2 3 A schematic representation of Conway's (1996) model of autobiographical memory structure Adapted from "Autobiographical knowledge and autobiographical memories" In D C Rubin (Ed), Remembering our past, p 67-93

(i) Structure of Conway's ABM model

The first level is referred to as "Lifetime Periods" (LP) and represents a general or abstract level of autobiographical knowledge that primarily contains thematic knowledge relating to extended time periods, e.g., when I was at school X, while I was at job Y. Within a theme, events are also sequentially organised by time. For example, the index marked "relationship theme" might contain, "Met my boyfriend in November 2000, bought a house together in 2002 and got married in 2004". Likewise, events in LP for a given time period might extend across various themes, e.g., "During the time that I lived in London, I worked at firm X (work theme) and was engaged to person Y (relationship theme)". Information at the LP level is defined in terms of continuing over years and decades.

The second level of this model is referred to as "General Event" (GE) and contains knowledge of repeated events (e.g., going to the shops) or things that have occurred over an extended period of time but are shorter in duration relative to LP, (e.g., a short holiday). A feature of this level is the summarised version of events that have occurred, which Belli (1998) suggests is important in being able to efficiently store information that is highly frequent and therefore potentially cognitively burdensome. Summarised information of frequently performed similar events are abstractly contained at the GE level, thematically ordered via recurring temporal patterns of incidence, e.g., weekly basketball matches during a competitive season. All events at this level are thematically or temporally related to each other and are also related to the LP level, in that incidences of recurring events form part of, or are situated within, incidences of more extended events, i.e., travelling to away games by coach (General event), during competitive basketball seasons (Lifetime period). Information at GE level is defined in terms of lasting days, weeks and months. Conway (1996) suggests that GE provides a basic level of representation that is not too general like LP, or too specific like the next level on the hierarchy, Event Specific Knowledge (ESK).

The third level is referred to as "Event Specific Knowledge" (ESK) and information at this level tends to take the form of images, feelings and highly specific details of GE and LP events. Conway and Rubin (1993) define ESK as "...memories that indicate the retention of sensory details of actions in a general event or lifetime period" (p. 107). Conway (1996) proposes that events at ESK level contain perceptual and contextual information that provides a sense of reliving a past event. According to Conway, they exist as a collection of salient images, feelings, thoughts and emotions that are organised at the point of retrieval, by attaching to relevant thematic and temporal events of higher order

autobiographical knowledge structures, i e at GE or LP levels Information in ESK is defined in terms of events that have lasted seconds, minutes and hours

Support for Conway's three-tiered hierarchical structure of autobiographical memory is evident in the diary studies of Linton (1982) and Wagenaar (1986) and also from an experimental study on amnesic patients (Conway, 2001) Linton's (1982) six-year diary study, described above, indicated that entry to ABM was via GE level, rather than ESK level, lending support to a form of hierarchy Access at the GE level for information provides a summarised version of events Then depending upon the external cues offered, a more labour intensive search of the next level, ESK, may be initiated in order to enrich the memory retrieved, with more specific information

This explanation of access to the ESK level in Conway's model of ABM is supported by Wagenaar's (1986) study In a six-year diary study of his own memory, Wagenaar found that " the probability of recall depended upon the number of retrieval cues as well as on the nature and combination of those cues", (p 249) In his diary, he noted the "who", "what", "when" and "where" of all events and later, when given one or a combination of those cues, he would have to complete the memory by providing the missing information Results from his study indicate that " double and triple cueing resulted in better performance than independent cue utilization" (p 249) A limitation of Wagenaar's study was that he was the sole participant and thus issues of familiarity bias and generality to other people are relevant Wagenaar's study was at least a progression from Linton's (1982) six-year diary study of her own memory, where she recorded brief descriptions of daily events, with descriptions not exceeding 180 letters Unlike Wagenaar, her dependent variables were dating accuracy and response speed In Linton's study, the effectiveness of different cues was not possible as the cues the author wrote on her cards daily were those that she deemed personally salient and unusual Also, the only question she asked herself at each memory test was always in relation to "when" Linton (1982) has stated that she has a bias for being extremely good at remembering dates and so the generality of her findings are perhaps limited

Conway (2001) also presented neuropsychological evidence of a hierarchical structure, following an EEG study with amnesic patients (N=5, age 18-35) that mapped changes in neurological impulses during remembering In this study, participants were monitored while (i) they prepared to recall memories (pre-retrieval phase), (ii) recalled memories to cued words (retrieval phase), (iii) made a manual response to indicate when

they had a detailed and specific memory in mind and then hold it for several seconds (hold memory in mind phase). Patterns of activation were taken at three different intervals, representing (a) early, (b) middle and (c) late in the three phases. Results indicated when patients could access ESK, they were always also able to access GE and LP. However, if LP information was lost, a simultaneous loss of GE and ESK information also appeared to occur. The methods employed in this study (using amnesic people) provided a method of isolating the retrieval process of very specific memories. This may have facilitated the removal of interference from other memories. However, while these results provide physiological support for the hierarchical structure theory of Conway's (1993, 1996) model, further investigation is required with non-amnesic participants.

Conway's model represents the most comprehensive explanation of the structure of ABM, detailing the various levels of structure and the characteristics of memories stored at those levels. It presents a more flexible approach to the access of memories within ABM, via time-related and thematic indices, than is explained by either Kolodner (1985) or Tulving (1983). Physiological evidence, in the form of EEG's, has also given support to the basis of Conway's theoretical model (Conway, 2001). Therefore, Conway's (1993, 1996) model of the ABM will be used to explain the structure of, and access to, memories relating to physical activity in this thesis.

(ii) Process of retrieval of memory from ABM

Events within each level of the structure of ABM are organised temporally and thematically, so that particular memories may be indexed depending upon the context, detail and type of the external cue that initiated the memory retrieval (Brown, Rips, & Shevell, 1985). Memory retrieval can therefore be top-down, sequential or parallel, as memories may overlap themes or temporal zones. The following section details a model of the *process* of memory retrieval that reflects his theoretical structure of ABM (Conway, 1993). Following presentation of a cue from the environment, the task of memory retrieval launches a 3 stage cyclical retrieval process (Conway, 1996; Williams & Hollan, 1981). The first stage of this retrieval process involves recognition of the demands of the task at hand and the creation of a memory description to fit that task. Then secondly, that memory description is used to search the indices of the autobiographical structure in order to retrieve the correct memory. The extent of the search and the richness of the memory retrieved will depend upon the scope of the initial cue and the perceived demands of the task. This leads

to the third and final part of the retrieval process proposed by Conway (1996) i e , an assessment of whether the memory retrieved is correct and the most appropriate This may necessitate another cycle of retrieval until " the appropriate LP, GE and ESK are represented in a temporary structure in working memory" and the process of cyclical retrieval is terminated (Anderson & Conway, 1993, p 1180) This model is similar to a model of the cognitive process used to retrieve information to answer a questionnaire (Tourangeau, 1984, Tourangeau et al , 2000) Tourangeau's model provides a theoretical framework for explaining the question-answer process, which occurs when a respondent answers a question and is discussed below

2 5 2 *Model of the question-answering process for questionnaires (Tourangeau (1984), Tourangeau, Rips & Rasinski (2000)*

Questionnaires are designed to yield and collect information or experiences from respondents, by means of eliciting personal opinions or recalling personal events from the autobiographical memory (Reiser, Black, & Abelson, 1985) The establishment of the CASM interest group in 1984 initiated the studying of cognitive processes, specifically memory recall in questionnaires Only one adult PA questionnaire (Lifetime PA Questionnaire) has applied this cognitive model during its design and pre-testing phases (Friedenreich et al , 1998) Until recently, it has been the practice of many physical activity researchers to extrapolate the findings of general literature on memory recall and then apply them to their own situation (Durante et al , 1996, Baranowski, 1988) The application of a cognitive model has yet to be applied to the design of a short-term recall PA questionnaire (Friedenreich et al , 1998) and is the purpose of this current research

There are four stages in the question-answer process described by Tourangeau (1984, Tourangeau et al , 2000) These are detailed in Table 2 6 (see p 35) The stages are not necessarily sequential, distinctive or mandatory Instead, the engagement of specific cognitive processes to access the autobiographical knowledge base will depend upon (i) the type of question asked, (ii) how accurate the answer needs to be and (iii) how quickly it needs to be answered (Tourangeau et al , 2000) For example, these authors suggest that for factual questions, judgement (stage three) may replace retrieval (stage two) when people answer questions about information that is automatically known to them For example, people do not actively need to search and retrieve information that is known to them like their age or name

Table 2.6. *Stages of the response process*

Stage	Specific Processes
Encoding (Comprehension)	▪ Attend to questions and instructions ▪ Represent logical forms of question ▪ Identify information sought ▪ Link key terms to relevant concepts
Retrieval	▪ Generate retrieval strategy and cues ▪ Retrieve specific, generic memories
Judgement	▪ Assess completeness and relevance of memories ▪ Draw inferences based on accessibility ▪ Integrate material retrieved ▪ Make estimate based on partial retrieval
Response	▪ Map judgement onto response category ▪ Edit response

Note. Adapted from "The Psychology of Survey Response" by R. Tourangeau, L. Rips and K. Rasinski, (2000), Cambridge University Press, USA, p.8.

There are many different types of questions that are asked in questionnaires, which in turn leads to the possibility of different types of answers that could be given (e.g., factual, attitude or frequency questions). Each of these types of questions make varying demands on the cognitive response process (Tourangeau, Rips & Rasinski, 2000). As the purpose of this study is to design a questionnaire that is concerned with obtaining information from the respondent about the duration, frequency and intensity of specific physical activities, the remainder of this review will focus upon the cognitive response processes associated with this type of question. The following section is a review of Tourangeau's (1984) and colleagues (Tourangeau, Rips & Rasinski, 2000) model of the question-answer response process.

2 5 2 1 Encoding (Comprehension) Stage

The purpose of this stage is to identify the information being sought, i e to assign a meaning to the question asked by attending to the cues, syntax and context of the enquiry or instructions In order to develop a reliable and valid questionnaire, the meaning that is intended from a word or sentence must be relatively stable across all respondents How the respondent will interpret the question is a function of many variables, including (i) competency in the language, (ii) knowledge of the subject matter, (iii) knowledge of the context in which the question is written and also (iv) according to the amount of time and effort that is invested in interpreting the question (Tourangeau, Rips & Rasinski, 2000)

Problems at this stage of the response process have been identified through explorative focus groups and cognitive interviews that examine the respondents understanding of the meaning of the question (Willis, 2002) Difficulties in comprehension tend to occur as a result of the complexity and ambiguity of sentences (syntax) as well as the use of presuppositions and a vagueness in the meaning of the question (semantics), (Gerber, 2002) Understanding the meaning of a question is compromised when a lengthy or complex question exceeds the respondent's ability to process it, resulting in a misinterpretation of that question This causes an overburdening of the respondents working memory (Just & Carpenter, 1992) and may result in a slowing down of the cognitive processes, including the retrieval of information

Issues of ambiguity in questions arise when words or terms used may have more than one meaning The context of the question may provide some indication of what meaning is intended, however the ambiguity often arises because of the innate structural complexity within the question In terms of questions specific to physical activity, definitions of PA have previously appeared inconsistent in their descriptive content, or even absent altogether from questionnaires (Caspersen, 1989, Blair, Jacobs, & Powell, 1985) Recurrently, terms such as "fitness" and "exercise" are used synonymously with descriptions of PA, e g , The Sports Participation and Health among Adults in Ireland Survey (Fahey et al , 2004), Survey of Lifestyle and Nutrition (SLAN) Survey, Friel et al , 2003) This may result in misunderstandings among respondents and thus, misclassification of exercise behaviours Inaccuracies are thus produced in the relationship between physical activity for health and physical activity for fitness (Blair et al , 2001)

The level of respondent comprehension of terms used in PA questionnaires is not well documented and therefore the resultant effect on reliability and validity is an unknown quantity (Durante et al , 1996) Findings from Tudor-Locke et al 's (2003) study that investigated the understanding of common phrases like "physical activity", "intensity" and "leisure", found diversity in the interpretations of each word across a sample of African-American women (N= 196, aged 40+ years) To illustrate the ambiguity and misinterpretation in a PA questionnaire, examples of questions from two commonly used PA questionnaires are shown below

1 "How many city blocks or their equivalent do you normally walk each day? (Let 12 blocks equal 1 mile)" (Taken from "Paffenbarger Physical Activity Questionnaire", Paffenbarger, Wing & Hyde, 1978)

2 "For at least the last three months, which of the following activities have you performed regularly?
   a) Jog or run at least 10 miles per week
   b) Play strenuous racquet sports at least 5 hours per week
   c) Play other strenuous sports at least 5 hours per week
   d) Ride a bicycle at least 50 miles per week
   e) Swim at least 2 miles per week

(Taken from "Stanford Usual Physical Activity Questionnaire", Sallis et al ,1985)

Both of these questions incorporate all four sources of comprehension difficulty identified by Tourangeau et al (2000), i e , ambiguity, complexity, vagueness and pre-supposition In terms of *ambiguity* (a word that may have more than one meaning), question one above asks about the amount of "blocks" that the respondent walks This term is limited to respondents that are familiar with the use of the word "block" in the context of distance The word may have a different meaning for another respondent who works in the construction industry and deals with another type of "block", (i e , a building block) The word "regularly" in question two above also has the potential to elicit different meanings among respondents, for what is "regular" to one person, may be completely irregular for another In the questionnaire, authors needed to define what they mean by the word "regular", to prevent misinterpretation

The issue of *complexity* is highlighted in question two above, where a list of specific examples and durations are given Ambiguity within the question is again a feature, due to the complexity of the layout and syntax of the question Having defined themselves as taking part in activities "regularly", the respondent may then wonder if they are "regularly" active only if they meet the criteria of activity outlined in one of the subsections, or do they need to answer "yes" to all of the subsections in order to satisfy the term "regular"

Both questions one and two above also offer *vague* expressions (imprecise ranges of application) and *presuppositions* This may cause the respondent to misunderstand the question While presupposition is a normal part of a sentence, it may lead to difficulties in questionnaires when the presupposition fails to apply In particular, question 1 asks the respondent to "Let 1 mile equal 12 blocks" Firstly, having pre-supposed that the respondent knows what a block is, the question attempts to clarify the term by equating it to a mile A further presupposition is thus imposed on the respondent by assuming they know automatically how long a mile is What if the respondent is only familiar with metric kilometres? In question two, the term "last three months" may be perceived as vague as it does not specify whether this should be the *last three calendar months*, or *three months from the present date*

Although the above scenarios are essentially hypothetical as little research to date has fully explored the issues involved in the comprehension of PA questionnaires (Durante et al , 1996), it highlights the necessity for further research within this area Due to the paucity of PA related data, the hazards associated with questions from other questionnaires and surveys may be theoretically applied to PA For example, some respondents may choose one way to make a vague, complex or ambiguous question more personally meaningful, while others could choose a different way This might lead to adverse variability in the responses as well as the reliability and validity psychometric properties of the questionnaire research (Sudman et al , 1982, Belson, 1981)

2 5 2 2 Retrieval stage

The purpose of this stage is to recall relevant information from long-term memory in order to answer the question asked In terms of questionnaire design, the characteristics of the question presented to the respondent can affect the quality and accuracy of the retrieval process, leading to the construction of either a complete or incomplete memory (Jobe, Tourangeau, & Smith, 1993) Question characteristics that can facilitate an appropriate

retrieval strategy for the respondent include (i) the provision of specific retrieval cues to help generate recall that are of good quality and in sufficient volume, (ii) enhance the distinctiveness of the event, (iii) a good degree of fit between the terms used in the question and the events original encoding and (iv) a limited time frame between the time the event occurred and the date of recall Discussion and empirical evidence for each of these characteristics is now presented

(i) Retrieval cues

The most effective type of cue will depend upon the goal of the retrieval task, i e , the focus of the question and while it is obviously impossible for a questionnaire designer to accommodate cues that are individually distinctive to all respondents, research has indicated that some cues are more effective than others in enhancing retrieval of information about events Results from Wagenaar's (1986) six-year diary study indicated that " double and triple cueing resulted in better performance than independent cue utilization" (p 249) The "when" cue performed the worst when singular cues were given This result may have implications for questionnaire design if, for example "when" and "where" cues are given together, to help the respondent retrieve a more accurate memory

One study examined (N=54, age 37 +/-8 2, Female=74%) participants recollections of going to the doctor during a given time frame (Brewer, 1988) Different types of cues were compared to see which was the most effective for retrieving accurate memories Brewer compared answers with medical records and found that the most effective cue stated what went on (i e , reason for visit), while less effective cues were to mention other people (i e , Dr's name), where the event occurred (i e , which surgery) and when it occurred (i e , date of event) Brewer's conclusion for his findings was that if there are a large number of similar events, a significant cue would be to state what went on during the event (e g , had an injection) In terms of questionnaire design, the provision of various examples and the context of occurrence (e g , household activity – vacuuming the floor), may provide some cueing assistance to the respondent

(ii) Distinctiveness and original encoding of the event

Linton (1982) found that with the passage of time during her six-year diary study, some events began to lose their distinctiveness and some eventually became completely forgotten She attributed this to the possibility that as the number of trials or experiences

increased, so too did the level of semantic knowledge about that type of event Simultaneously, a decrease in the episodic knowledge occurred, making the specifics of each event less discernible Kolodner's (1985) model of autobiographical memory structure provides a level of support for this rationale However, Conway (1993) suggests that highly frequent similar events, are harder to remember than unique ones, particularly if the event had no personal significance and therefore made little impact on the "self" while the memory was being encoded Types of details that make the event "distinctive" to the person, enabling more effective encoding of the event and thus easier retrieval include, emotionality, saliency and sensory information (Conway, 1996)

In the context of a past-seven-day PA recall questionnaire, encouraging access to Event Specific Knowledge (ESK) in the autobiographical base is preferable, as information at this level is high in specificity and sensory detail This is due to its recent, and in some cases its infrequent occurrence (Conway et al , 1993) Access to ESK is through the General Event (GE) level (Anderson & Conway,1993) This study by Anderson and Conway directly examined which type of memory detail would provide the fastest access to the memory as a whole Fifteen participants (11 women, mean age 19 years) had previously recorded ten details including, the beginning detail, the end detail and a distinctive detail about 36 events as they occurred The ten details were presented incrementally to the participant, with the first cue being either the beginning, end or distinctive detail The outcome measure was the time to verify the memory Results indicated that the personally distinctive details provide the fastest access to knowledge at GE The authors concluded that GE is organised thematically and chronologically in terms of contextualising distinctive details, in order to discriminate one memory in GE from another

Implications for PA questionnaire design suggest that cues should be used that can index GE, which in turn provides cues that access the more detailed and accurate ESK Research should thus explore which cues and sensory details are encoded during and specific to the performance of PA and are personally distinctive to the majority of participants (so that connections with the sense of self are made) There is little documentation to suggest that this has previously occurred (Durante & Ainsworth, 1996, Friedenreich et al , 1998) Reiser, Black and Abelson (1985) explain that in order to gain the best access to specific memories, the presence of original context cues should ideally be at hand Specifically, the apposite cues may include verbal, visual, olfactory or other stimuli that were in attendance at the encoding stage In other words, congruence of original

conditions will facilitate accurate recall from memory for the purposes of responding to a questionnaire Consensus on personally distinctive cues and how PA is understood and encoded by the general public could be achieved through focus group studies (Tudor-Locke et al , 2003) Exploration of how physical activity and its distinctive cues are encoded were undertaken in this thesis

(iii) Temporal issues and forgetting

Few theories of forgetting have been investigated for "real-time" events and have tended instead to look singularly at laboratory tasks (Tourangeau et al , 2000) The pioneer of forgetting curves (Ebbinghaus, 1885, cited in Eysenck & Keane, 2001) showed an exponential rate in forgetting, but only in relation to non-sense syllables Explorations of memory for real-life situations was conducted in two diary studies (Wagenaar, 1986, Linton, 1982) During her 6 year study, Linton found that fewer than 1% of items were forgotten during the calendar year in which they were written, but after the first year, forgetting occurred at a rate of 5-6% She suggests that this indicates that real-world episodic memories are more durable than laboratory memories However, as Linton defined "forgetting" in terms of whether she could accurately date the event only and not for remembering any other information, the low percentages reported may not be transferable to everyday memories Wagenaar (1986) on the other hand highlighted the methods used in both his and Linton's study as conducive to encouraging rehearsal of the event because of the possibility of retrieving the same memory on a few occasions This he feels would affect the retention curve by encoding the memory further into memory, so that retention curves " do not describe a property of memory, but rather a property of how our lives are structured", (p 236) The possibility of "rehearsing" events or activities may have implications for the recall of activities performed a while ago and stored in the Lifetime Periods level The recall of activities performed during the past seven days should not be affected because of the relative recency of occurrence and the lack of comparative opportunity to frequently recall it

2 5 2 3 Judgement stage

For questions concerned with calculations or summing of events (as in PA questionnaires), this stage assists in gathering all information retrieved and judging whether the sum of memories is complete, accurate and falls within the scope of the question asked

(Tourangeau et al , 2000) Questions relating to time spent performing physical activity are referred to as "Temporal-duration" and "Temporal-rate" types of questions and will be the focus of the remainder of this review

Temporal-duration questions are intent on finding the time between the beginning and the end of an event Temporal-rate questions ask respondents to count or estimate the number of time units in which a particular event has occurred (Tourangeau et al , 2000) Some events may consist of one continuous time unit (e g , playing a basketball game), while others may extend over many discontinuous time units (e g , how long it took to cycle to work in one week) In the context of the current research question, i e the design of a 7-day recall physical activity questionnaire, the outcome of interest is the amount of collective time that the respondent has spent doing different types of physical activities, during a specific time frame This may involve recalling a combination of both continuous and discontinuous events Figure 2 4 outlines the timeline of a recalled discontinuous event, e g , time spent cycling

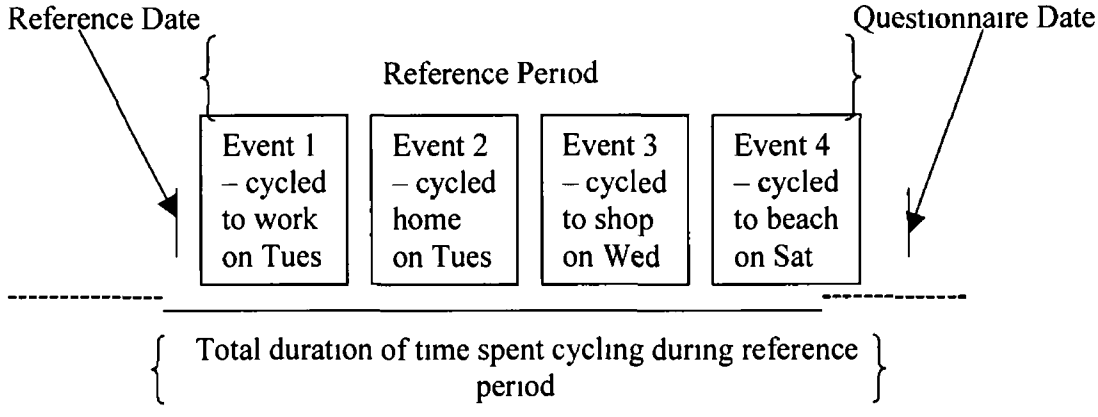

Figure 2 4 Illustration of the timeline of a discontinuous event Adapted from "The Psychology of Survey Response" by R Tourangeau, L Rips and K Rasinski, (2000), Cambridge University Press, USA, p 102

Having recalled the relevant information to working memory, the respondent is not yet ready to provide an answer For questions relating to duration and rate, the respondent must add, average, combine or in some way summarise the information to provide an answer (Subar et al , 1995) Calculation strategies are outlined by Tourangeau et al (2000) and include

i Episodic enumeration – i e recall and count of individual incidences

ii Rate-based estimation – i e , recalling generic information, usual rates

iii Exact tally – i e , recall current tally of events

iv General impression – i e , make a guess from question cues and perceived "normal" values

Use of one or all of these calculation strategies will depend upon the nature of the question With specific reference to PA recall questionnaires, behavioural-temporal questions are a common feature as respondents are often questioned about their involvement/participation in terms of "How many activities ? How often ? How long ?" However, in terms of obtaining acceptable reliability and validity properties for a new questionnaire (i e the specific objectives of this research), designing the questionnaire to facilitate the use of episodic enumeration strategy would seem to be prudent, in that it removes the variability associated with guessing and estimations

The most habitually used process of recalling data for temporal-rate information is by means of episodic enumeration, although this appears to be determined by many factors, including the type of information sought and the number of events to report (Burton & Blair, 1991, Blair & Burton, 1987, Sudman, 1980) Blair and Burton (1987) conducted a telephone questionnaire (N=387, age 28 +/- 14 7 years), which investigated whether respondents used any other processes other than episodic enumeration to calculate various behaviour rate questions Questions ranged according to expected measures of frequency and included things such as frequency of ATM use and incidences of grocery shopping The results revealed 12 distinct processes, with episodic enumeration accounting for only 28% of the answers Rate processes accounted for the majority of the answers and as rate demands increased, a simultaneous rapid decline in the likelihood of enumeration was observed, in favour of rate based estimation When between one and three events had to be recalled, the authors note that 84% of respondents used an enumerative strategy However, when between six and ten events were recalled, episodic enumeration accounted for only 15% of the strategies used, decreasing to zero % for 11 or more events A conclusion of the results from this study would indicate that when economy of effort is considered, it would seem more probable that rate based processes of memory recall are engaged in

In a food frequency questionnaire recall study, the accuracy of respondents to recall what they had eaten during either the previous two or four week period was assessed, comparing their answers to a self-written diary for the same period (Smith et al , 1991) Results indicated that recall of what was eaten and when, were poorly matched with the diary records, even when tested immediately after the data collection period In addition,

many items of food that were not written in the diary were written in the test questionnaire, because " foods that were routinely eaten were given in answers" (p 11), regardless of whether they had been actually eaten during the data collection period This suggests a tendency to rely on generic rate estimations for frequently consumed items, rather than taking time to individually recall This finding has repercussions for PA questionnaire design, as the type of information requested of respondents can often relate to frequently performed activity, e g , climbing stairs

The main determinant in deciding which calculating strategy to use is perceived effort of recall In turn, perceived effort is determined by relative accessibility in memory of the required information Following on from this, accessibility is assessed according to the task variables at hand These can also be a source of response error and are categorised into either a) event characteristics, or b) question characteristics (Tourangeau et al , 2000, Burton et al , 1991, Blair et al , 1987, Sudman, 1980) and are outlined in Table 2 7

Table 2 7 *Sources of response error and variables influencing the use of episodic enumeration strategy*

Event Characteristics	Question Characteristics
Saliency of event	Question relevance
Recency of event	Question order
Regularity of event	Question wording
Perceived threat of event	Question structure
Omission and Telescoping of event	

2 5 3 2 1 Event Characteristics

(i) Saliency

This refers to the uniqueness of the event The greater the personal cost or benefit of the event, the more likely the respondent is to remember it, e g , winning the league cup, winning a prize, getting married This feature relates to Conway's (1996) concept of autobiographical memory being a representation of the self and how memories are stored in terms of reference to attainment of personal goals Less salient events however, usually involve more habitual, everyday occurrences which do not stand out from the norm These events are not regarded as important to the construct of self and are therefore not strongly encoded into memory at the time of occurrence (Conway, 1996) This makes them more

difficult to remember Such events would include usual daily routines, e g usual work schedule, things usually done at home, at weekends

Behavioural events are usually salient across two dimensions (Sudman, 1980) In a study to compare the properties of similar events that were easily remembered or difficult to remember (N=2,138), Sudman found that the economic/social costs or benefits, e g something that was fiscally extraordinary, like buying a car, will stand out in the memory, while daily/ weekly grocery bills will typically not vary radically Secondly, the continuing consequences of the event will affect the saliency of the event For example, if there are continuing reminders that the event happened, like photographs of the event or the presence of a car or frequent recounting of the event to others, this results in a rehearsal of the event in memory and further, better quality encoding of the event in memory (Conway, 1996)

In terms of PA questionnaires, further research is required which will examine what properties of physical activity and moderate and high intensity are described as salient by the target population of the questionnaire Descriptions of PA by the general population could reveal features that make the performance of these activities personally salient These features can then be included as recall cues within the questionnaire

(ii) Recency and regularity of event

This factor is similar to the saliency issue in that it suggests that recall of irregular events from memory is more accurate than routine-type incidences Episodic enumeration becomes effortful and time consuming for events that are not exceptional and are not prompted by sufficient or appropriate retrieval cues It is presumed therefore that when this happens, episodic enumeration will be abandoned for a more favourable and less cognitively demanding (but also less accurate) process (Burton & Blair, 1991) Conway's (1996) model of autobiographical memory provides an explanation for this effect, in terms of irregular, recently performed events being akin to events held at Event Specific Knowledge (ESK) level These events are rich in sensory detail that take the form of feelings, images and highly specific details of actions (Conway & Rubin, 1993) As time since the event increases however and the frequency or regularity similarly escalates, details become less distinct and events begin to merge Memory details constitute a personal summary of an episode rather than a complete, literal and objective record (Conway, 1996) Recall of these types of activities are similar to events held at the General Event (GE) level which contains information about a type of event at a more abstract level

Events at this level may be prone to forgetting because the events are too similar to each other

The merging of frequent events from ESK to GE level has been given some credence by an Electroencephalograph (EEG) monitoring study (N=25, Male=60%) that performed a neuro-physiological examination during memory recall It indicated that event specific, sensory-perceptual details of recent experiences (past minutes and hours) do not last in memory unless they are linked to other more permanent systems in the autobiographical structure This integration may occur by frequent rehearsal or attachment of a strong sense of "self" to the event, (Conway, 2001) In their six-year diary studies of real-life memories, both Linton (1982) and Wagenaar (1986) found that memories for events tended to merge or be forgotten as time passed or were similar to other events Questionnaires that request information about the performance of regular PA may be equally prone to forgetting or inaccurate recall Further research is required to investigate the properties of activity that is most likely to be regularly performed among the general population and at which intensity

(iii) Perceived threat of event

The influence of this factor will be determined by the respondent's interpretation of the event and also whether they view it as being personally threatening to their sense of self or how personally important the question topic is to the respondent (Krosnick, 1991) Individuals who are concerned deeply with a specific topic which they do not deem to be threatening to their identity, e g health, divorce, poverty etc, will be motivated to a greater degree to make an effort to interpret, retrieve and provide a response to a question In contrast, respondents with no interest in the topic or who deem it to be personally invasive or threatening, will be less likely to be motivated to make the effort to recall accurately from their memories (Krosnick, 1991)

Autobiographical memory represents a knowledge of the experienced "self" where personal goals play a major role in the formation, access and construction of specific memories (Conway, 1996, 2001) The perceived threat of a question to the sense of self, as determined by information in autobiographical memory, is an important consideration when constructing a questionnaire According to Conway's (1996) model, the respondent may fail to answer or give inaccurate answers if the content of the question does not correspond to or threatens information about the sense of self that is represented in

autobiographical memory This may in turn affect the validity of the questionnaire Typically, PA is not perceived to be a personally threatening topic (Sudman, 1980) In this study (N=2,138), Sudman assessed peoples "uneasiness" to talk about certain leisure activities Only 1 3% of the respondents considered physical activity an uneasy topic to talk about, compared to other topics such as smoking marijuana (42%), intercourse (41 5%), intoxication (29%) and income (12 5%)

For non-threatening questions, specific memory errors, referred to as omission and telescoping, are viewed as the greatest detriment to accurate reporting and are discussed below (Sudman & Bradburn, 1973)

(iv) Omission

This refers to the failure of the respondent to report an event that should have been included in the time-frame asked Given the effect of memory decay with time, as described earlier, respondents are less likely to be accurate about specific events for a given time-frame when the gap between the time of event occurring and the recall query is substantial The further back the event is in time encourages the merging of episodic (singular, short-term) memories with semantic (general, long-term) memory, particularly for highly frequent events This may increase the opportunity for inaccurate reporting Tourangeau et al (2000) suggest that long retention intervals may encourage respondents to infer or estimate answers, instead of trying to recall accurately

The ability to accurately place and date a particular event can be performed in two ways (Tourangeau et al , 2000) Firstly, the respondent can utilise a "temporal tag" which may be present in the memory trace being used to retrieve the information These tags usually represent some important or significant personal or public landmark, which allows the respondent to place the recalled information into some relative timeframe Alternatively, the respondent can use associated events to access the pertinent data In this way, each individual event is linked to a group of events that are connected by similar causal factors (Brown et al , 1985) Accessing information about one event may induce recall of information about the incident in question, in this way temporal and thematic tags attached to information in memory can assist in recalling the information required Conway's (1996) model of autobiographical structure also advocates temporal and thematic linking of events (see Section 2 4 2)

For questionnaire purposes, mechanisms to reduce error by omission or under-reporting events for a specific time-frame would include helping the respondent to place the event in a particular contextual time-frame and also to limit the length of the recall period The use of a personal calendar has been used in other questionnaire topics to help the respondent limit their recall to the relevant timeframe and reduce the opportunity for omission (Belli, 1998) It is also a method of reducing telescoping and so is discussed in more detail below

(v) Telescoping

This is another commonly reported error and involves reporting events which have occurred prior to the assessment period in question, but are moved forward to that time period (Sudman & Bradburn, 1973) This has the effect of falsely over-reporting and thus creating inaccurate answers In their study of subjective dates in long-term memory, Brown and colleagues (1985) found that respondents tended to use telescoping, particularly in relation to high profile events In this study, respondents were given a list of public events (some high profile, some low profile) and asked to date them Events that were higher in public awareness were usually reported as happening more recently than less salient events, thus being subjected to telescoping Durante & Ainsworth (1996) have made an assumption re physical activity from these findings, i e , that the recall of vigorous activity will most probably result in telescoping, more so than light/moderate activity, as being higher in awareness, it is the more salient of the two

The negative effect of telescopmg on accuracy could be improved by providmg relevant retrieval cues These could involve helping to place the respondent in the relevant time frame by encouraging them to locate personal landmarks during this time with the use of a personal event history calendar (Means & Loftus, 1991) Unfortunately this method has only been used in face-to-face interviews and not in self-administered questionnaires (Belli, 1998) This is therefore an area requiring more investigation, as self-report questionnaires are the assessment tool of choice in epidemiological research and is also the particular focus of this research Event history calendars are thought to improve the quality of memory retrieved because it encourages thematic and temporal links through sequential and parallel retrieval processes, similar to the natural structure and process of autobiographical memory (Conway, 1996)

By asking the respondent to date specific landmark personal events, the accuracy and completeness of the questionnaire information is thought to improve as it allows the respondent to use the landmarks as cues for other events as well as placing events in a specific timeframe, thus reducing the possibility of telescoping (Belli, 1998) To compare the effect that a personal event history calendar (PC) had on the accuracy of recalled data , Belli, ,Shay & Stafford (2001) used an interview-led questionnaire about social and economic behaviours from the previous two years to compare the effect In one condition, the respondents (N=307) did not use a calendar, while the second condition did (N= 309)

Results that in comparison to the questionnaire without the PC, the PC condition led to better quality retrospective reports for the reference period of the questionnaire Research to date has focussed upon the use of landmark personal events in studies focussing on exploration of lifetime events (Belli et al , 2001, Belli, 1998, Means et al , 1991, Thompson, Skowronski, Larsen, & Betz, 1996) However, remaining specific to the research question of this study, i e recall of past seven day physical activity, a calendar pertaining to autobiographical memory at the general event (GE) and event specific knowledge (ESK) instead of lifetime periods (LP), would possibly be a more relevant source of cues for recent day-to-day events Further investigation is also required to evaluate the effect of using a PC with a short-term recall PA questionnaire, compared to when the PC is not included

2 5 3 2 2 Question Characteristics

(i) Question relevance

Caspersen (1989) states firstly that the " science of epidemiology is concerned with carefully quantifying the rate of health-related states or events that occur within the population being studied" (p 423) and secondly that " epidemiological investigations should employ physical activity measures that correspond to the dimension that relate to the disease or health outcome under investigation", (p 439) Clearly identifying the parameters of enquiry with regard to health aspects or specific diseases, would therefore seem to be an imperative component of a reliable and accurate questionnaire For example, in order to obtain the most accurate responses pertaining to the study of coronary heart disease in the adult population, the survey questions should be tailored to reflect the important types of activities that impact on the incidence of heart disease, i e , activities associated with caloric

expenditure and also activities which correspond to various aerobic intensities (Caspersen, 1989).

(ii) Question Order

Kalton & Schuman (1982) observed respondents answering general questions on a particular issue which were intertwined with more specific questions on the same issue. Two different studies by these researchers, dealing with topics such as abortion and driving standards, found that there was no difference in answers to the specific questions, regardless of whether they were placed before or after the general questions. Conversely, answers to the general questions differed by 7 % when the question was asked after the specific question as opposed to before it. While no definitive theory has been proffered for this discrepancy, the authors suggest it is possible that the respondents believe that the answer to the general question should not include the specific information already answered in the preceding specific question. When this occurs, the answer has a " subtraction effect"(Kalton & Schuman, 1982).

Question order effects are also thought to occur when respondents are presented with a list of specific items and asked to tick their favourite items or items they use most. Sudman (1980) reports that when a list becomes large, items at the top of the list are more likely to be read more carefully than items in the middle or at the end. It is unclear as to whether this is due to fatigue or due to a lack of motivation on behalf of the respondent, however items placed later in long lists are usually reported less while items listed early on long lists tend to be mentioned more often (Sudman, 1980). In general, the effect of question ordering on respondent accuracy has not been shown to be of great significance. Research on question order in physical activity questionnaires appears absent from the literature and thus requires further investigation to see whether the reliability of data is adversely affected.

(iii) Question Wording

As previously discussed in the "Encoding" section (2.5.2.1.), ambiguous wording and complex syntax can adversely affect the respondents understanding of the questionnaire. Wording of the question also has the potential to affect the response strategy that is used to calculate an answer, particularly if a vague quantifier like "minutes usually spent" or "approximately" are used (Tourangeau et al., 2000). These words imply that a rough estimate will suffice, while more direct wording like "how many?" or "exactly" may

encourage the respondent to use an individual enumeration strategy (Blair & Burton, 1987) In this study, the effect that the wording of the question "how often?" against "how many times?" was compared to determine which retrieval strategy was used to arrive at an answer (N=384, randomly selected from a telephone book) Results indicated that the "how often?" question was less likely to encourage respondents to specifically enumerate, resulting in estimated answers Currently, there is little direct evidence to indicate that the decomposition of a question improves accuracy, however Tourangeau et al , (2000) suggest that the recall and count of specific incidences may be more likely when complex questions are decomposed into smaller, more specific units, particularly in situations where a large number of events need to be recalled

Questions, which the researcher may presume to be "standard", may be viewed differently by respondents who vary according to educational, socio-economic and ethnic backgrounds (Warnecke et al , 1997) Although the question has been asked in a language common to both respondent and researcher, the interpretation and understanding of that sentence/question may be different for both Fundamental to achieving a valid, accurate measurement tool is to ensure uniformity across all administrations of the questionnaire (Sudman, 1980) Piloting of the proposed tool among the specific intended population will assist the researcher to identify common problems and potential pit-falls with the wording and terminology In this way, standardisation of wording can be enhanced and misinterpretations of questions can be minimised

(iv) *Question Structure* (Open vs Closed)

Generally, researchers will have a preference to use closed questions because they are easier to process and make cross-study comparisons They also reduce coder variability as they are usually already categorised, e g , according to frequency In this way, they have the further advantage of being a form of aided recall, used to enhance accurate reporting, if they are very specific in their probing (Sudman & Bradburn, 1973)

In certain cognitively demanding tasks, respondents are known to select processing strategies that balance effort and accuracy (Durante & Ainsworth 1996) Answering questionnaires is one such cognitive task, where the process of episodic enumeration has previously been described as the optimum strategy for producing accurate answers rather than the semantic approach of rate based estimates (Baranowski, 1988) With regard to the influence of question structure (i e , either open or closed-ended) on respondent accuracy,

the provision of response categories (as for closed questions) has varying significance Response categories can provide magnitude cues, thus assisting the respondent to place the question in temporal context and allowing for a more accurate response However, Burton & Blair (1991) report that the provision of category responses may discourage episodic enumeration by encouraging respondents to answer, e g in groups of 10 minutes or to the nearest half hour This also indicates to respondents that accuracy and therefore maximum effort is not a requirement

2 5 3 4 Response stage

This stage involves selecting the response by assessing the answer and editing it for appropriateness and suitability to the question asked If at this stage, the respondent feels a more appropriate answer could be achieved, the answer is rejected and the retrieval and judgement process must be re-enacted (Tourangeau et al , 2000) The cognitive burden of the retrieval process and the calculation strategy employed in the judgement phase however may influence the respondent in the "Response" stage to accept the answer reached, particularly if motivation is absent and perceived effort is high In this incidence, the respondent may choose to simply "edit" the answer reached to make it more appropriate and acceptable (Jobe, Pratt, Tourangeau, Baldwin, & Rasinski, 1997) In a study of the numbers of lifetime sexual partners, the participants (N=175) tended to report in rounded numbers of multiples of five, once the number of partners was larger than 10 The obvious sensitivity of this topic is a possible reason for rounding, in that the respondent may want to provide a lower, more socially desirable answer However, as reported in section 2 4 3, enumerative strategies are also less likely for items of a greater frequency than 10 (Means et al , 1991, Blair & Ganesh, 1991) Therefore, depending upon the type of information and the quantity of information to be recalled, the answer delivered may be the result of an accuracy-effort trade-off, regardless of the efforts of the researcher (Tourangeau et al , 2000)

2 6 Stage of Exercise Behaviour

Throughout the various studies in this thesis, a central construct of the transtheoretical model (TTM) of behaviour change, called the Stages of Change (SOC), has been used as a criterion method of classifying participants according to their current level

of participation in physical activity and their intention to change The TTM is a model of behaviour change that explains peoples readiness to adopt new behaviours The SOC model is one of the key constructs of this model and was originally developed following observation of smokers during a self-directed smoking cessation process, but has since been applied to the understanding of readiness to participate in physical activity (????Prochaska & Marcus, 1994) Evidence suggests that there are five stages which describe the current position and intention of individuals, in relation to their readiness to do physical activity (Prochaska & Norcross, 1999) The five stages are described in Table 2 9 Each of these stages represent a variable temporal period and invariable tasks that must be completed by the individual, prior to moving to the next stage Movement between the stages is believed to be in a cyclical pattern, thus explaining how some actioners may relapse to contemplation for a short while before re-establishing themselves once more as regularly active Many people tend to recycle to previous stages several times before reaching the maintenance stage and successful behaviour change is achieved (Prochaska & Norcross, 1999)

Stage	Description
Precontemplation	Not active and no intention to change behaviour in the next six months No awareness of health risks of being inactive
Contemplation	Not active but thinking about becoming active within the next 6 months Weighs up the pros and cons of changing behaviour
Preparation	Is active, but not enough to reach minimum health requirements* Is intending to become regularly active within the next 30 days
Action	Have made overt changes to their behaviour to become regularly physically active within the previous 6 months
Maintenance	Have been regularly physically active for more than 6 months and are less tempted than actioners to relapse

Note Minimum health requirements refers to the ACSM/CDC (1995) guidelines of minimum participation levels for health benefits

In the context of physical activity (PA), movement into the fourth and fifth stages of the model (action and maintenance), indicates compliance with the minimum participation levels recommended by the ACSM/CDC (Pate et al , 1995), for the purposes of gaining health benefits Participants in both of these stages are considered to be regularly active,

while participants who identify themselves as being in one of the first three stages (precontemplation, contemplation or preparation) are considered to be not regularly active enough to attain health benefits (Cardinal, 1997). Validity of the SOC model has been examined in PA studies by collapsing the five stages into two or three stages (Marcus & Simkin, 1993; Cardinal, 1997). Marcus and Simkin divided their participants (N=219, female=64%, mean age=40.6 (+/-12.9) years) into the five stages and then grouped them into three categories, i.e. (i) precontemplation and contemplation (because these are sedentary), (ii) preparation (because these are not regularly active) and (iii) action and maintenance (because these are regularly active). Having completed the 7-day physical activity recall questionnaire (Sallis et al., 1985), analysis was performed to assess the relationship between the stages of exercise behaviour and reported time spent in moderate and high intensity activity for the three levels. Results from this study indicated that significant differences existed between the three categories, in terms of time spent in activity, suggesting that whether one is deemed regularly active or not by their stage of change, is differentiated by their self-reported physical activity (Marcus & Simkin, 1993). An examination of British adolescents (N=244, Female=58.6%) with the Godin Leisure time questionnaire (Godin & Shephard, 1985) also collapsed the five SOC's into three categories and found that moderate and high intensity activity differentiated the stages, but low intensity did not (Wyse, 1995). These studies were limited as they did not fully explore each of the individual stages of exercise behaviour, and therefore caution should be attached to the application of the results. However, the collapsing of stages into categories that identify regularly active and non-regularly active people, is appropriate for the purpose of assessing compliance with the ACSM/CDC recommended minimum participation health guidelines (Pate et al., 1995).

In the context of questionnaire design, the SOC model has also been used to provide construct validity to the Scottish Physical Activity Questionnaire (SPAQ) (Lowther et al., 1999). Participants (N=94, mean age=33 (+/-10.8) years) were divided into the five stages of exercise behaviour and assessed for differences in levels of reported physical activity. Results indicated, that as expected, actioners and maintainers reported significantly more activity than any of the three not regularly active stages, thus indicating that SPAQ has concurrent validity with the stages of change model. This conclusion is reached on the assumption that the stages of change model has itself been previously widely validated for physical activity. A meta-analysis of the application of TTM to physical activity was

conducted by Marshall and Biddle (2001) This analysis included 91 independent samples from 71 reports published between 1983 and 2000 and included just one random sample study as well as six longitudinal, 54 longitudinal and 10 quasi-experimental studies (N=68,580) The selection criteria for inclusion in the meta-analysis was the application of at least one core construct of the TTM In terms of studies that used the SOC model, numbers of individuals identified at different SOC's was dependent upon the depth and relevance of the enquiry about physical activity to the ACSM/CDC recommended minimum participation criteria for health (Pate et al , 1995) Specifically, the highest number of maintenance individuals were identified when information about frequency, duration and intensity was included, compared to mainly precontemplators when it was not (Marshall & Biddle, 2001) This is possibly because of the precontemplators comparative lack of knowledge about "physical activity" and what it means to be regularly active, thus when no formal definition or explanation is provided, this group may miscalculate the amount needed to meet minimum requirements and categorise themselves at a lower stage Misinterpretation and ambiguity in understanding key phrases and questions was also a feature of a cross cultural examination of similar words and an investigation of physical activity related words, when participants interpretations were compared to official academic definitions of the same words (Tudor-Locke et al , 2004, Warnecke et al , 1997) Thus, the validity of the stages of exercise behaviour should be enhanced by clear explanations of each stage, preceded by a statement of the minimum participation guidelines for regular physical activity

The SOC of exercise behaviour has been used throughout this thesis as a selection criteria for participants in each of the studies This enabled inclusion of individuals who self-selected themselves as being not regularly active and no intention of becoming so, to individuals that are regularly active and have been for at least 6 months As the outcome measure of this thesis (i e , the Physical Activity Recall Questionnaire, PARQ) is concerned with whether individuals are meeting the ACSM/CDC required minimum participation levels of physical activity (Pate et al , 1995), participants were categorised into "not regularly active" (NRA) and "regularly active" (RA), once they had identified themselves as being in one of the five stages of change

2 8 Summary of literature review

Participation in regular PA of at least moderate intensity can positively affect health and reduce the incidence of some chronic diseases The revised ACSM/CDC guidelines detail the minimum participation levels required for the purpose of incurring health benefits (Pate et al , 1995), although the degree to which a participant's health is positively affected will increase as they expend more energy The amount of physical activity energy expenditure (PAEE) used, is related to the baseline fitness of the participant and the level of intensity and duration that the activity is performed at This is known as the dose-response curve (ACSM, 2000) Questionnaires provide the most comprehensive method of obtaining information about all of the properties that contribute to calculating PAEE (Montoye et al , 1986) These are frequency, intensity, type and time (FITT) of the activity

Recall questionnaires are known to be cognitively burdensome on the respondent as it involves remembering information about activities performed at some time in the past, from a memory system known as autobiographical memory Accuracy of the recalled information depends upon factors related to (i) the cognitive aspects of recalling information, and (ii) the questionnaire, i e , wording, layout(Schwarz, 1999) Although many PA questionnaires have previously been constructed, little research has been undertaken to identify the factors that optimise a respondent's recall process (Durante & Ainsworth, 1996) Conway's (1993, 1996) model of autobiographical memory and Tourangeau's (1984, 2000) cognitive model of the question-answer response process provide guidance in terms of how information is stored, accessed and retrieved from autobiographical memory The research topic of this current study is concerned with accessing information about physical activities that were done within the past week and lasted for a period of minutes to hours

According to Conway's model, the most accurate, sensory-perceptual and contextual information for activities of this duration and recency in occurrence is contained at the Event Specific Knowledge (ESK) level and is accessed through the General Event (GE) level Few questionnaires to date have considered the natural structure and retrieval processes of autobiographical memory when drafting questions for the respondent to answer (Belli, 1998) To reduce possible error in reporting when, where, how long and how often physical activities occur, appropriate and numerous cues should be investigated in the

design stage of questionnaire development, in order to facilitate multiple retrieval routes from memory These cues can then access the temporal and thematic indices in many directions, as advocated by Conway's model (i e , top-down, parallel, sequential) McKenna and colleagues study (2004) found that physical activity was readily recalled by type, but cues were required for properties such as intensity and duration Effective cues, identified from the focus groups and cognitive interviews of this study, were suggested to include movement descriptors, physical signals of effort and the use of daily landmarks to enhance recall of activity These cues reflect the temporal and thematic mdices that Conway's model advocates as having access to information in memory

Episodic enumeration has been identified as the most accurate form of recalling memory in certain conditions (i e , infrequent, recently performed events) Thus, a new questionnaire should ideally be developed and constructed in a manner that encourages the employment of episodic enumeration, where possible Research has also indicated that the respondent is most likely to provide accurate responses when clear, unambiguous instructions are provided, so that they can accurately answer the questions

Aided recall is an important motivating mechanism which can be used to improve responses and encourage respondents An example of aided recall mcludes a personal calendar, which may help contextualise the time-frame, reduce telescoping and provide temporal cues for other activities by landmarking events The more salient events are more likely to be recalled easier and more accurately Therefore, it is important to investigate what properties of physical activity are most salient to people This may be where it occurs, when it occurs, what is being done or how it is being done

The researcher should identify the variable of interest from the outset in order to remove any ambiguity and in order to focus the questions and the respondents on the specific health outcome under investigation Particular attention is required when the questionnaire is being designed to ensure that the content, context and structure is appropriate for the population in question This is the focus of study one in chapter three, which utilises the focus group setting to investigate and discover what words, properties and meanings Irish adults of 18-55 years of age use and attach to the concept of lifestyle physical activity It is anticipated that this study will permit an insight into the way information about physical activities are thematically stored within the structure of autobiographical memory

CHAPTER 3 - ITEM GENERATION FOR THE PHYSICAL ACTIVITY RECALL QUESTIONNAIRE (PARQ) A FOCUS GROUP STUDY

3 1 Introduction

Ambiguous wording and misunderstanding of questions are known to affect the quality of questionnaire response (Burton et al , 1991, Sallis et al , 2000) According to Tourangeau's (1984, Tourangeau et al , 2000) model of the question-answering response process, understanding the intent of the question or instruction is the first problem encountered by respondents because of the use of unfamiliar terms or words Tudor-Locke and colleagues (2003) research showed that terms such as "physical activity", "leisure" and "intensity", are synonymous with varying degrees of interpretation, understanding and application among a sample of women and are not always understood as researchers presuppose they will be Interpretations and understandings of key phrases can differ depending upon age, culture and personal lifestyle (Tudor-Locke et al , 2003, Warnecke et al , 1997) A report of the challenges of measuring PA in women by Ainsworth (2000) advised further investigation into how people define moderate intensity activity and the cues that trigger the recall of moderate intensity activities as it had been previously overlooked

Retrieval of accurate memories is part dependent upon the encoding (comprehension) conditions that are in force at the time of event occurrence, i e , how well the respondent interprets the activity to be relevant to their concept of self (Conway, 1996) According to Conway, events that do not impinge upon the current themes, plans and goals of the self, may simply not be encoded into memory Accurate retrieval is also dependent upon the type of information that is sought Infrequently performed physical activity is located in the Event Specific Knowledge (ESK) level of autobiographical memory This level contains memories that are a collection of salient images, thoughts, feelings and emotions These memories are organised at the point of retrieval by attaching to relevant thematic and temporal events of higher order autobiographical knowledge structures, i e , at General Event (GE) or Lifetime Period (LP) levels Conway (1993) suggests that GE is the preferred level of entry for memory retrieval and represents the point where autobiographical memory is maximised in terms of access to event specificity Information pertaining to recurring temporal patterns of incidence are contained here In the context of designing a questionnaire that seeks accurate information about specific physical activities

from the past, it is necessary to explore (i) the general properties and (ii) the specific properties, of physical activities stored at GE and ESK, respectively This study investigated these properties, by discussing people's exposure to and understanding of physical activity and its related terms, such as moderate and high intensity These phrases were specifically investigated because they are appropriate to ACSM/CDC health-related guidelines, which advocate at least 30 minutes of participation in physical activity (PA) of at least moderate intensity, on most, preferably all days of the week (Pate et al , 1995) One's understanding of the term ultimately depends upon one's exposure to it, the context of that exposure and the perception of that contextual experience

In summary, the purpose of this study was to explore the thematic, temporal and salient properties of physical activity and its associated intensities It was particularly concerned with how these descriptions were generally expressed in terms of the vocabulary and jargon of the sample population for which the use of the new Physical Activity Recall Questionnaire (PARQ) is intended (i e , Irish adults, 18-55 years old) Recurrent descriptors, phrases and examples of activities that emerged from these focus groups were then used for the purpose of item generation in the development of the PARQ questionnaire

3 2 Methods

3 2 1 *Recruitment Criteria*

The literature review in chapter two unearthed an imbalance between the types of questions asked in physical activity questionnaires and their suitability for certain sections of the population, e g women and ethnic minorities (Warnecke et al , 1997, Ainsworth et al , 2000b, Ainsworth, 2000) As the aim of this research was to design a questionnaire that is capable of capturing information about physical activity behaviours among the general Irish adult population between 18 and 55 years old, the sample for this study were intentionally selected to represent a variety of people within this age group

Volunteers for the study were asked to complete a screening form (Appendix one) which included the following selection criteria a) age, b) gender, c) place of residence (Dublin Vs outside Dublin), d) occupation / unemployment status and e) physical activity stage of change (SOC), which is a construct of the Transtheoretical Model (TTM) applied to physical activity (Prochaska & Marcus, 1994) The TTM is a model of intentional

behaviour change and the SOC construct of this model explains the time-related change in intention to alter physical activity behaviour. It was deemed important to gain a varied perspective about "physical activity" from people who are at various stages of intending to become and being physically active.

3.2.2. *Recruitment Procedure*

Access to participants in cities (Dublin and Limerick) was made through contact with hospitals, businesses, schools, unemployment agencies and universities. Access to rural participants in small towns (Banagher, Co. Offaly; Claregalway, Co. Galway; Miltown Malbay, Co. Clare) was made through their respective area health boards, who provided a list of contact names for local community groups.

Eight focus groups in total were conducted, four in Dublin and four outside of Dublin. Participants were selected for the focus groups by stratified random sampling (Thomas et al., 2001). This occurred by firstly dividing the volunteers by one specific criteria and then randomly selecting them for participation. In this study, the dividing criterion was physical activity SOC and participants were divided into group one (stages 1-3) or group two (Stage 4-5). This was because stages 1-3 represent those that are not yet regularly physically active, in contrast to those in stages 4 and 5. It was intended that each focus group would consist of 10 participants, with each focus group made up of five participants in stages 1-3 and five participants in stages 4-5, that were heterogeneous in terms of gender, age and occupation status, although this was not always possible because of limited availability. The interaction of active versus inactive participants in the same group was purposive and is advocated as a means of exploring different perspectives that different participants may have (Morgan, 1998a). As the purpose of this study was to ascertain different viewpoints and interpretations of physical activity and its intensities among all sections of the target population (18-55 year olds), this mixed group methodology provided a means of uncovering and exploring descriptions and understanding of these properties, by allowing the participants time to describe and discuss their own experiences with other members in the group (Morgan, 1998b).

Following selection, each participant was individually contacted by telephone to further explain the purpose of the study and to arrange a day and time that suited everybody to meet and conduct the focus group. All focus groups were conducted at a location that was convenient for the majority of participants. They were also informed at this stage that

they would receive €15 towards travelling expenses All participants received two phone calls – one to initially invite them to attend and the second on the day before, to remind them of the details of the focus group the following day Two reserve participants for each group were also contacted a week before and asked if they would be willing to participate at short notice, should somebody drop out However, even with this back-up support, group numbers ranged from 5-12

3 2 3 *Focus Groups*

In order to acquire an understanding of participant's physical activity experiences and to capture some of their verbal expressions, a series of focus groups were conducted Focus groups are a qualitative method that enable (a) exploration and discovery, (b) context and depth and (c) interpretation of a given topic in a group environment (Morgan, 1998b) They are particularly useful for the development and refinement of instruments, especially to identify domains, to obtain natural vocabulary for item generation for questionnaires and to assess cultural appropriateness (Shaughnessy & Zechmeister, 1997) Discussions are centred around key questions that are posed by the moderator Once the question is asked, the moderator recedes and allows an open forum type discussion to evolve among the participants These participants are usually unknown to each other but normally share one or more common characteristics (such as age, occupation, smoking habit, physical activity status) and each group usually consists of between 5-10 participants (Morgan, 1998a) In order to gain experience in conducting a discussion group and to ensure the suitability of the script, a pilot study was undertaken

3 2 3 1 Focus group script

Questions in the script evolved from topics identified in the literature review and also directly from the research question specific to this study, i e , how do people perceive, explain and express their knowledge or involvement in physical activity at different intensities? The final script (Appendix two) is the product of three revisions following the pilot study and appraisal by a review panel The first review panel (the participants of the pilot study) examined the content of the script and suggested the improvements mentioned above Secondly, two members of the general public, that had the profile of the target population, suggested that pictures of somebody engaging in moderate or high intensity could be included in the discussion to spark a debate about whether they were accurate

These were included in the focus groups, but were introduced after participants had voiced their own perceptions and thoughts. Finally, an experienced focus group moderator examined the script content and recommended the inclusion of more summary points at the end of questions.

The final script included questions about what is understood by terms such as "physical activity", "moderate intensity" and "high intensity". Some questions required the participants to break into groups of two or three to write down their thoughts before reporting back to the larger group. This was to promote a more detailed discussion among the participants before they reported back to the group as a whole. For most questions, participants were also asked to use specific, individual examples from their own lifestyle and experiences to demonstrate their point. The rationale for the structure and content of this script was to attempt to see what language they used to specifically describe the term "physical activity". To maintain relevance to the dose-response issue and the ACSM/CDC health related participation guidelines (Pate et al., 1995), the script also specifically enquired about descriptions and understanding of both moderate and high intensity physical activity. In particular, questions were posed to see what information seemed most salient and repeatedly engaged in when the participants were recalling specific types and intensities of activities. Probing questions were based upon obtaining information about general and specific properties of recalled activity. This corresponds to information stored at GE and ESK levels of autobiographical memory, respectively. It also permitted insight into how participants interpreted (encoded) and recalled features of PA and intensity, based upon the encoding and retrieval stages of Tourangeau's (1984; Tourangeau et al., 2000) model of the question-answer response process.

3.2.3.2. Pilot Focus Group

The aim of the pilot study was to test the interview script and assess the moderator's facilitation skills, in a focus group of people with similar characteristics to the questionnaire target population (N=8, males=50%). The pilot study was observed by an experienced focus group moderator who provided constructive feedback and suggested modifications for the interview script and moderating technique. Two recommendations were made, specific to improving moderation skills. Firstly, it was noted that participants were prematurely cut-off when offering comments that were contrary to the group's opinions. It was suggested that sensitivity to diverse views would be better if the moderator took less notes and maintained more eye contact. Secondly, a need for a summary by the moderator

after questions one and question five was identified, as these questions were important to the research question (i e , participant's understanding of specific terms like physical activity and moderate intensity) It was also suggested that participants should be encouraged to oppose or concur with the content at that time This would assist in maintaining the focus of the discussion on the intended topic area

Recommendations were also made by the observer with regards to the content of the interview script It was suggested that participants should be broken into smaller groups and be given more time to answer question two (i e , What do you understand by the term "physical activity"?) This was because it was felt that answers to this question were central to the main research question Breaking into smaller groups may better enable contributions from each participant Contributions from each group were then to be presented back to the group as a whole It was also suggested that emphasis should be placed upon the word "you" when participants were asked questions about their opinions or thoughts on something (e g , what does "moderate intensity" mean to **you**?) This was to ensure that answers were representations of their own experiences or attitudes and not those considered to be what other people do or think

The content of the answers to each of the questions were deemed to be context appropriate by the experienced observer The pilot study participants were also asked to contribute their thoughts about the content and style of the focus group Feedback from the groups stated that the meaning of the questions were clear and that they found it easy to contribute when they wanted to It was felt that the topic (physical activity) was one that was comfortable to talk about and something that everyone had some experience of

Following the review by the group and the experienced observer, alterations were made to the interview script It was also decided that the moderator would not take notes (except for quick comments at the end of a question) during the course of the focus group Detailed notes would be taken immediately afterwards from the camera and audio-recording equipment and also during the group by the assistant moderator This would enable the moderator to appear more attentive to the group and be more aware of the direction of the discussion

3 2 3 3 Focus Group Research Design

In this study, all participants were between ages of 18-55 and were unknown to the researcher Each focus group followed the same format The uniformity of methods

between focus groups is important to ensure that subsequent themes and topics that emerge between groups are directly comparable (Barbour et al., 1999). All focus groups lasted between 60-75 minutes. The format applied to each focus group is now outlined.

Each group was conducted by the same moderator (author) and assistant moderator (an exercise science postgraduate student). Before the participants arrived, refreshments were organised, seating was arranged and audio-visual recording equipment was set up. As each participant arrived, they were welcomed and a ten minute informal chat with other participants was facilitated in order to create a relaxed environment. All participants were then asked to sit at the table provided and to read through and sign their consent forms. Further explanation of the purpose and structure of the focus group was verbally given by the moderator and the opportunity for the participants to ask questions was provided. The presence of the recording equipment was explained as being necessary for the purpose of transcribing details after the discussion group was over, to ensure that no important information had been overlooked. Participant confidentiality was assured and no objections were made.

The function of the moderator is to gain insight into and understanding of a topic by gently probing, without leading, the participants in a specific discussion area. The skill of the moderator requires an ability to remain neutral, keep the questions focussed and to maintain the discussion on track, without impeding the flow of conversation (Krueger, 1998b). The moderator was assisted in achieving this balance by following a script of questions that helped to maintain the focus of the discussion. The role of the assistant moderator was to ensure the functioning of the recording equipment and to take notes on observations and comments made throughout the focus group (Krueger, 1998b). The assistant moderator was seated away from the table where the participants sat, so as not to be a distraction to the group. At the end of the discussion, the assistant moderator was invited to reflect upon the major points that had been raised. To ensure that the content reported was valid, participants were asked to verify whether the assistant moderator had accurately reflected their comments. They were also invited to clarify or add any further points if they wished.

At this point, the focus group ended and the participants were thanked for their time and contributions. The assistant moderator gave each of the participants their travel expenses and directed them to the exit. The moderator and assistant then spent approximately 10 minutes going through the relevant details of the focus group to ensure

that all of the discussion had been captured on tape and so the assistant moderator could also provide some feedback on the discussion process Within a week of the discussion, the assistant moderator also provided a written report of her comments on each of the focus groups

3 3 Data Analysis

3 3 1 *Reliability and Validity Procedure*

Systematic planning of the focus group sessions provided a visible, structured and accountable approach to undertaking the study Records of script developments, sample recruitment, focus group transcripts and analysis procedures were maintained to provide a transparent, thorough and rigorous method of data collection and analysis This served to verify results and suggested that findings were reliable by assuming that another researcher would arrive at similar conclusions for these groups, if access to documents and raw data were made available, (Krueger, 1998a)

Validity of the data was considered by purposely including various steps, such as participant verification and peer examination and evaluation of a transcript Participation verification was driven by

1 At various points during the focus groups, the moderator provided a synopsis of the previous discussion and participants were given the opportunity of clarifying or adding any points At the end of each focus group, the assistant moderator provided an overall synopsis and again the participants were afforded the opportunity to clarify any misinterpreted points
2 At the end of each focus group, the moderator and assistant moderator held a debriefing session to compare notes and thoughts about the groups

The validation of the data analysis procedure was facilitated by a group of three individuals (an educational psychologist, a social science researcher and a university lecturer in research methodology) They were each sent a copy of the second focus group and asked to code the text for emerging themes and categories in the context of physical activity and performance at various intensities Their comments and coding lists were compared against those of the researcher Any discrepancies were discussed with each individual and alternative wordings were agreed upon between the researcher and the individual For instance, examples of activities that were described by other reviewers as

"domestic", "garden-related" or "housework", were described by one reviewer as "environmental" Following discussion with the reviewer, the title of "housework" was assigned to that type of activity

3 3 2 *Content Analysis*

The selection of the method of analysis is guided by the research question The purpose of this research was to identify the phrases, contexts and images that participants associate with the words "physical activity", "moderate intensity" and "high intensity", given the potential benefits from engaging in physical activity as identified by the ACSM/CDC (1995) The outcome of this study was to generate items for a questionnaire

Strauss and Corbin (1998) suggest that grounded theory will generate or discover a theory by developing and interrelating categories of information across multiple discussion groups Categories and information will emerge from data that are "grounded" in the field and the information that is collected represents the reality of experiences of the group members Creswell (1998) suggests that the centrepiece of grounded theory research " is the development or generation of a theory closely related to the context of phenomenon being studied", (p 56) The purpose of this research is not however to "build a theory", but rather to provide insight into the vocabulary used when describing the recall of physical activity of different intensities, for the purpose of generating items for a questionnaire In this research, content analysis of the focus group data was undertaken This yielded categories of information in relation to descriptions of participation in, and recall of, physical activity at different intensities These were then compared with data from successive focus groups in a method referred to as a constant comparison technique This form of interpretational analysis is a general principal of grounded theory, in that it is a process of continuously interpreting and comparing texts from data that has been obtained from and "grounded" in the field (Cote, Salmela, Baria, & Russell, 1993)

Specific to a physical activity related area, comprehensive use of content analyses has been used in sports psychology investigations (Sage, 1989, Bain, 1989, Locke, 1989), with varying degrees of diligence in describing the analysis process (Cote et al , 1993) A flexible 2-stage approach to data organisation is proffered by Cote and his colleagues (1993) that has been endorsed by other sport and physical activity researchers as providing " procedural guidelines for the conducting of content analysis and are based on the grounded theory approach to research", (Biddle et al , 2001) This is outlined below

Stage one: Vocabulary used by participants

Organisation of the unstructured data from each of the eight focus groups in this study was facilitated by using a qualitative software package (NVIVO version 10). Initially, segments of text were coded by attaching provisional in-vivo nodes to them, which were the exact phrases spoken by the individuals themselves to identify and label meaningful pieces of text. These could later be linked to other related pieces of text. For example, words such as "pleasurable", "enjoy the feeling" and "getting a buzz" were used to describe the sensation of participation in moderate intensity activity. Quotes from respondents were labelled this way in the initial coding session but were later grouped together as "Perceptions of PA/Psychological/positive affect" in secondary coding.

This open-coding format facilitated a flexible approach to separating relevant sections of information and was the first stage of organising and analysing the vast unstructured data from the eight focus groups. The second stage permitted a more rigorous sorting and categorising of data through engagement of the constant comparison technique. This method of analysis conveyed relevance to the research question, i.e. information about encoding and recall properties of PA and intensity, by "comparing incidents applicable to each category and integrating categories and their properties", (Flick, 2002, p.231).

Stage two: Axial coding and the creation of categories

The next stage in the inductive analysis of the focus group data was to list and compare the in-vivo nodes and to place them into categories. These categories contained similar references to, or descriptions of, physical activity in general or at specific intensities. NVIVO software was used to search each focus group for marked sections of texts that had non-specific node titles but that had shared properties. This facility offered by NVIVO means that all stage one nodes relevant to a specific theme could now be grouped together into more descriptive and specific category nodes, independent of which focus group or which question within the focus group it had originated from. For example, although the question, "What do you understand by the term physical activity?" was asked in question two, participant's perceptions of physical activity that were voiced much later in a focus group were still captured and coded.

At times, stage one codes were identified as contrasting to the description of a category node title (known as a deviant case) When this occurred the analysis process either accommodated the deviant case by modifying the category node title to be inclusive of new descriptions, or as the analysis proceeded and no other cases with similar descriptions were identified, the node title was altered to exclude the deviant case If a case was considered deviant to one category node, it was then recoded into a more appropriate theme The decision to include or exclude deviant cases is part of analytical induction, where the presence of deviant cases means that the category or description must be redefined or reformulated, until a universal relation is established (Flick, 2002) For instance, participants often provided incidences of being active that did not fit into established nodes that described types or categories of activity These included descriptions such as "moving muscles", "just walking from room to room" and "climbing stairs" While these clearly indicated some form of activity, their non-specific context meant they did not appear to bear any relationship to a specific category, e g sport, housework In this instance, an "Incidental activity" node was created to accommodate non-specific activity

This stage of the analysis was concluded when "theoretical saturation" had been reached This is defined by Strauss and Corbin (1998) as " when no new information seems to emerge during coding, that is, when no new properties are seen in the data", (p 136) At this point, any new data from further focus groups will adequately fit into existing categories or nodes and there are no new emergent themes or categories An example of the coding process performed in this study is illustrated in Table 3 1

Table 3.1. *An example of 1st and 2nd stage coding from focus group two.*

Q2] So talking about physical activity, what is it then when I actually say the words, "physical activity"…what is it you actually think of or understand by those words? What are the first thoughts that come into your head? OK, Orla's just going to write them up on the board there for us, so if you just want to shout out…whoever wants to start for us there

	1ST stage	2ND stage
[C]:We've sport and enjoyment	"sport"	Type of PA
	"enjoyment"	Psychological
[O]:Em, sometimes it's not enjoyment, when you go to the gym and you're running on the treadmill, which I find really boring, it's not enjoyment for me, like	"gym" "running"	Type of PA
	"not enjoyment"	Psychological
	"boring"	Psychological
[J]:It's so boring…		
[C]:Like football	"football"	Type of PA
[R]:You can…you're taking physical activity by just walking to a bus stop, walking to work, walking anywhere	"walking to"	Type of PA
[J]:Well, I know like with me…'cause to try and motivate myself to actually go to the gym or do something like that is hard, so what I try and do…is I try and walk into work…instead of taking the bus, or you know, say in the shopping centres, I'll go up the stairs instead of taking the elevator. I try and put them into my daily routine, you know so I'll have walked 45 minutes in the morning and I won't really have felt like I've just walked that, if you know what I mean?	"gym"	Type of PA
	"walk into"	Type of PA
	"up the stairs"	Type of PA
	"walked 45 minutes"	Duration
[O]:I think that's the best way of you know, doing more physical activity is getting off the bus, back at the earlier bus stop, like I would get off in O'Connell street if I'm going to George's street, 'cause the traffic is just terrible, like. So that's even a pleasure to get off and do some walking, instead of sitting on the bus, like (laughs)	"pleasure"	Psychological
	"walking"	Type of PA
[M]:I don't find it though that enjoyable…'cause I'm living out in Ballsbridge and I'm coming out here…I get more time in bed basically if I cycle, because I like cycling, but since I've started doing it, I've really got a buzz out of the cycling part of it like, whereas if I put in things like that…climbing the stairs or something, it doesn't make me want to do more	"not enjoyable"	Psychological
	"like cycling"	Psychological
	"get a buzz out of it"	Psychological
	"cycling to"	Type of PA
	"climbing stairs"	Type of PA
	"make me want to do more"	Psychological
[A]:Housework, housework can tire you out a lot sometimes, I think…	"housework"	Type of PA
	"tire you out"	Physiological

3 4 Results

3 4 1 *Participants*

Eight focus groups were conducted in this study (N=71, Female=69%, mean age 34 5±10 4 years) Details of each focus group are given in Table 3 2

Table 3 2 *Demographic details of focus group participants*

Focus Group	N	%Male	Location	% stage 4 & 5 SOC	Age range (yrs)	% in paid work
1	9	23	Dublin	44	21-50	77
2	8	50	Dublin	37 5	20-42	0
3	8	25	Dublin	62 5	20-53	87 5
4	11	18	Dublin	63 6	20-42	72 7
5	5	0	Offaly	100	19-40	100
6	8	50	Clare	25	30-55	87 5
7	10	40	Galway	60	18-55	80
8	12	25	Limerick	66 6	19-50	75
TOTAL	71	30	-	56 3	18-55	70 4

A content analysis of the focus groups revealed the thematic, temporal and salient properties of physical activity and how they are generally expressed in terms of the vocabulary and jargon of the sample population The data also revealed that "physical activity", "moderate intensity" and "high intensity" are terms that are prone to varying degrees of interpretation, understanding and application It seems that one's understanding of these terms, e g physical activity, ultimately depends upon one's exposure to it, the context of that exposure and the perception of that contextual experience The following sections reveal the descriptions that Irish adults (aged 18-55) used to describe "physical activity" and "moderate and high intensity", in relation to how it applies to their general, daily living

3 4 2 *Descriptions of Physical Activity*

Results from this study show that mode or type of activity was the most accessible information in participant's memory when directly asked to explain what they understood by the term "physical activity" Interpretations of "moderate" and "high intensity" were provided by participants, following direct probing questions The physical and psychological descriptors of different intensities that emerged were also the product of

specific probes but were extensively discussed once initiated Information about frequency and duration were not specifically probed in this study, but were at times indirectly referred to The retrieval strategies used to calculate frequency and duration were a specific purpose of the cognitive interviews in study two (Chapter four) Other properties of physical activity that emerged without probing from the researcher were "degree of intention" and "purpose for undertaking activity" As the purpose of this study was to explore the definitions, examples and contexts of "physical activity", "moderate" and "high intensity" (to enable appraisal of compliance with the 1995 CDC/ACSM guidelines), results of this study will be discussed under these headings Other properties of PA that emerged indirectly, e g , degree of conscious intent are also discussed, where appropriate, under these headings

3 4 2 1 Type of activity

Question one specifically asked participants to articulate what they understood by the term "physical activity" (see FG script, Appendix two) All focus groups described types of activity during initial answers to this question, indicating that this is the type of information that is most accessible to participants in terms of storage and ability to recall For the most part, participants recalled these activities according to the place in which they occurred, e g , "at home" or "at work" They also recalled the activity in relation to the context in which it occurred at the time, e g , "going to/from work" or "while looking after the kids" This section looks at the properties and characteristics of the types of activities that were mentioned by participants It gives examples to support the titles of the categories, which were applied by the researcher, as a the result of recurrent phrases used by the participants during the course of the discussions Numbers in parentheses refer to quotes from the focus groups that are located in Tables within the respective sections For example, each category title has a Table number assigned to it, e g , Commuting (Table 3 3) Within the text of that section, a number m parentheses, e g , (4), means quote number four in Table 3 3 Other quotes which provide further support for categories and descriptions identified here are also located m the Appendices section (Appendix three) Numbers of Tables and Quotes in the Appendix correspond to the same numbers of Tables and Quotes in the main text below The reader is requested to simultaneously refer to both

3 4 2 1 1 Commuting (Table 3 3)

Arriving at a destination, other than by motor transport, is undertaken as an intentional act, for the purpose of enjoyment or health or due to the lack of an alternative Even when there is apparent intent, the motivation behind that intention is often variable (1) This persons' reason for walking home is in the context of having just completed an hour of highly intensive Boot Camp activities and it appears she uses the opportunity to wind down after such an intensive outburst of energy Although this could be seen as the end of a structured form of exercise, rather than commuting, she still chooses to "always walk home" because it makes her feel better

Deliberate choices were also made in everyday commuting activities (2) This participants' intention to actively engage in physical activity is obvious Although she initially began her journey on the bus, she consciously decides to get off the bus early and walk the rest of the way, declaring it a more pleasurable alternative than "sitting on the bus" Although a contributory factor in motivating her to get off the bus is the "terrible traffic", it still indicates an intentional behaviour to engage in physical activity

There are others who believe that having access to a car reduces the opportunity and motivation to either walk or cycle to wherever they are going They may even consider that they have no other choice but to walk or cycle (3) Having a car is associated with convenience (4)

In general, the participants of these discussion groups felt that there was always an option to using a car or taking the bus to get to work or the shops or wherever This was true for those who owned cars and those who did not (5) Walking or cycling to work, shops etc is seen as a practical form of physical activity It can be seen as a necessity for some people, for different reasons In one females' experience, "it's a means to an end, you have to get to work, so you do it naturally" [Focus Group 3 (FG3), (Age) 29, SOC 5], while for others, " I suppose the only chance they get to exercise is to (walk) to work and back home, 'cause they're sitting in their offices all day" [FG2, Female (F), 23, SOC 5]

To summarise, commuting by foot or bike to a given destination is seen as a viable form of physical activity Depending upon the participant, it's functional purpose is varied Sometimes it is done because of a lack of other alternative, ("need to do it" or "a means to an end") At other times it is an intentional act for the purpose of gaining enjoyment, "I've really got a buzz out of the cycling part of it", [FG2, F, 31, SOC 4]

Table 3 3 *Full quotes from the raw data that detail properties of physical activity while commuting*

No	Quote
1	"Yeah but then I'll always walk home because I think you need to unwind slowly, so I think the walk home kind of stretches the muscles a bit more and I feel better actually after walking home again" [FG2, F, 32, SOC 3]
2	"I think that's the best way of you know, doing more physical activity is getting off the bus, back at the earlier bus stop, like I would get off in O'Connell street if I'm going to George's street, 'cause the traffic is just terrible, like So that's even a pleasure to get off and do some walking, instead of sitting on the bus" [FG2, F, 23, SOC 5]
3	"I'd say I'm a good bit more active than a lot of my friends who have cars, like 90% of my friends would have cars, so they don't you know, need to walk as much as I would So that's a factor I think too" [FG3, F, 29, SOC 5]
4	"I used to walk to walk to the shop, I used to cycle everywhere, but when I bought a car, I just got lazy" [FG4, M, 29, SOC 5]
5	"I guess the walking instead of taking the car, so like walking to the shops to get something instead of taking the car to get it " [FG6, F, 30, SOC 3]

Note FG=Focus Group, F=Female, M=Male, Number= Age(Years), SOC=Stage of Change

3 4 2 1 2 Sports and Exercise (Table 3 4)

In many instances across the different discussion groups, certain recurring activities of a sporting nature occurred to people when they thought of "physical activity" As the participants were asked to first discuss in smaller groups, their thoughts are attributed to the larger focus group rather than any specific individual (1)

Table 3 4 *Full quotes from the raw data that detail properties of physical activity while doing sports or exercise*

No	Quote
1	"Walking, gym, playing football, sport in general, dancing" [FG3]
2	"They think that everybody was just born that way and that they are just amazing, whereas they forget that people actually had to start and had to learn the skills to be learnt in different sports and there's skills to be developed but they think that people have a talent and that they just appear like this you know" [FG8,F,50, SOC 5]
3	"I've just started a class in Scottish country dancing which can be very physical! So there's at least one night a week where I'm really puffing and panting!" [FG3,F,41, SOC 5]
4	"When it comes to set dancing, [it] is a great form of physical activity" [FG5, M, 53, SOC5]
5	"I love going to the gym, I love the feeling I have afterwards" [FG3, F, 37, SOC 3]
6	"I think a lot of people do physical activities for the social element as well, to meet new people and that sort of thing" [FG8, F, 27, SOC 5]

Note FG=Focus Group, F=Female, M=Male, Number= Age(Years), SOC=Stage of Change

As a first impression of "physical activity", these words provide an image of an entity that encompasses many properties. Particular properties were emergent throughout the discussions and clearly linked many of the named activities. These included requiring the participant to:

- travel to a specific place or venue
- have access to special equipment
- usually (though not always) requires financial input
- usually (though not always) requires trained supervision

More specifically, activities such as gym, football, swimming, tennis and camogie were mentioned and satisfy all of the above conditions. Each of these activities could also be considered to be structured in their format. There may be certain rules or regulations which must be adhered to for successful completion of the task, particularly if it relates to a "competitive" environment [FG5], "team training" [FG7] or "team sport" [FG8]. The word "sport" in particular is associated with a certain skill requirement (2).

Similarly "gym" is associated with "having to keep up a fitness regime to get results all the time" [FG8; M; 29; SOC 5]. The imagery consistent with the word "regime" is that of a regulated activity which is undertaken in order "to get results". This participant also suggests that "people need places like the gym to have a set of boundaries that, you know, I've gone to the gym, I've done my work on Monday, Tuesday, Wednesday, so they can see…to look at a bit of paper and see that this is exactly what they did…". Again, "boundaries" is suggestive of a structured activity that works within regulations or rules.

Words such as "walking" and "dancing" have been included within this category because depending upon the context in which they are undertaken, they may also possess one or more of the properties of this "sport and exercise" category. For example [FG8; F; 49; SOC 3], states that "walking has become very popular like altogether, power walking you know?", suggesting that it is a purposeful walk that is different from everyday walking. A reason for doing this more formal, intentional walking is that people who do it are "gone real health conscious and they're always walking and it is, like, to lose weight", [FG1; F; 22; SOC 2]. Equally dancing can be described in different contexts (3,4).

In summary, physical activities that are categorised as "Sport and Exercise" can be described according to certain properties they all possess. Firstly, each of them are structured to some extent with either formal (e.g., team sports) or self-imposed (e.g.

walking) regulations. The participant willingly and intentionally undertakes the activity, usually for some personal reason, for example enjoyment (5), or social (6). Except for activities like walking, additional properties associated with doing this type of activity involve needing money, a specific venue and equipment and perhaps a qualified instructor. Words used by participants to describe the collection of activities in this section tend to be "Sport" and "Exercise".

3.4.2.1.3. Leisure (Table 3.5.)

Activities described in this category may have some similarities to those discussed in the sport and exercise section, in that they could equally be referred to as "dancing", "walking", "playing football". The uniqueness of activities in this "Leisure" group however lies in the context in which the activities are undertaken. For the most part, these activities might be described as being more relaxed and less structured than those activities described in the preceding section. They don't require a formal setting and can often be spontaneous. There is an element of fun and sociability about these activities and there is not always an expectation of physically benefiting from participation (1), which was a component of "sport and exercise" when participants referred to "losing weight" and being "health conscious".

A description by one participant of "walking", shows how a change in intention or purpose can make the same activity seem different. A change in intensity or effort exerted seems to move it out of the leisurely pace and into a pace where she "expects to get something from it". This provides an insight into how participants perceive and recall doing the same activity, in different contexts. The purposive intent of walking by herself has similar properties to the sport and exercise category, while the walking with friends suggests a more leisurely context that is not time dependent and has a more sociable context. One participant in FG8 refers to "social walking", [F; 50; SOC5]. Descriptions from participants in FG5 make similar inferences (Appendix three, (2)).

Other activities mentioned by participants also exhibit properties of spontaneity and informality, e.g. "the lads kind of go and kick a ball around" [FG1; F; 21; SOC 5], and "things like going for a stroll every night" [FG8; M; 29; SOC 5]. Dancing is also seen as "a good form of physical activity", [FG4; F; 20; SOC 5] but can vary in its properties according to its context. As seen in the sports and exercise category above, it can be quite structured in "classes" but can also be informally participated in, in the context of

"nightclubbing" Although this type of activity does require a venue, it is not a pre-requisite of attending this venue that patrons must dance Therefore those that do engage in dancing, do so voluntarily, intermittently and in general, without instruction from a third party, i e a teacher Dancing in this context, despite its leisurable qualities is seen as quite an intense activity (Appendix three (3)) However, it appears that the sociable aspect of dancing at these venues is as important and therefore separates it from the more formal structured classes of the "sport and exercise" category, as discussed by participants in focus group three (3)

The varying contexts in which dancing can be seen as an activity is highlighted in this case by a female who says that she doesn't do "dance and clubbing" However, she had previously described another incidence of dancing that she does ((3), Table 3 4) Clearly, she does not equate both types of dancing as being in the same context This indicates that like walking, dancing can be performed in different contexts Within these contexts, it will be perceived as a form of activity by some participants but not by others This highlights the need for separating the more leisureable contexts of activities like walking and dancing from the more structured and purposeful forms of the same activity In a questionnaire format this may be achieved by creating two separate categories, i e "sport/ exercise" and "leisure"

In summary, although the same activities can be attributed to both "Sport and Exercise" and "Leisure" categories, it is the context in which they are undertaken that distinguishes their individual properties It is a feature of leisure activities that they are often undertaken in a less formal environment, are primarily motivated by enjoyment/fun rather than physical benefit (although this can be a secondary outcome) and are equated with a less intense input than those in the "Sport and Exercise" category There does not appear to be any specific term or phrase that the participants use to describe activities in this category, other than "for leisure" or "for social"

Table 3 5 *Full quotes from the raw data that detail properties of physical activity while doing leisure*

No	Quote
1	"For me it would be like the difference in going for a walk by myself and going for a walk with my friends When I go for a walk with my friends, we are just taking it easy and chatting but when I go for a walk by myself, I'm pushing myself always that little bit more because I want to get something out of it" [FG6, F, 30, SOC 3]
3	(Person 1) "Nobody at all mentioned dance and clubbing like does anybody do that because I don't" [F, 41, SOC5] (Person 2) "No, I used to but it's an age thing as well" [F, 29, SOC5] (Person 1) "That must be so intense" (Person 3) "I used to do it a lot more" [F,30,SOC2] (Person 2) "You'd be dancing the whole night" (Person 3) "Well you'd go to the nightclub, but you mightn't necessarily be dancing at all, you'd be sitting and chatting" (Person 4)" I don't go to nightclubs that much, I go to the pub more But yeah, it is something I didn't think about Yeah, I think a lot of people my age do get a lot of exercise that way" [M,20,SOC5]

Note FG=Focus Group, F=Female, M=Male, Number= Age(Years), SOC=Stage of Change

3 4 2 1 4 Domestic (Table 3 6)

Frequent reference was made to activities that occur in relation to maintaining order in the house and its environs (gardens) In general terms, these specific activities were described as "housework" or "doing work/things around the house" and "gardening" More specific activities that applied to this category were mentioned during the discussion groups (Table 3 6)

Table 3.6. *Full quotes from the raw data that detail properties of physical activity while doing domestic activities.*

No.	Quote
1	"There is always washing and ironing and delph and this and something to be done...If I sit down I'd always be thinking, God if I wasn't sitting I could have the ironing done, you know". [FG5; F; 35; SOC 5]
2	"I suggest to my wife...that going up and down the stairs how many times a day to do whatever chore needs to be done is physical activity". [FG1; M; 50; SOC 5]
4	"What kind of housework?" [Moderator] "Well, the shopping. The carrying of bags, the hoovering and gardening and that kind of thing" [FG4; F; 40; SOC 2]
5	"I'd just say if you're going to town shopping or that, usually because you're doing a lot of moving around and you end up being tired at the end of the day, so it must be you're doing physical activity". [FG2; F;35; SOC 30]
6	"Working against a resistance makes something like hoovering much harder...you can just hoover on a floor like this [wooden], you might pick up loads of stuff, but there's no resistance. But as soon as you're on a carpet and especially if you are trying to lift something that's in a carpet, rather than just on the carpet, then you start to feel like you're working against it. It's the same to my mind about light intensity is gardening when you're trimming something...but working against a weight..." [FG3; F; 41; SOC 5] "Like washing a window like you're taking out some sort of vengeance". [FG3; F; 30; SOC 2]
7	"My mother has a guest house in the summer and em, she always loses weight over the summer, like she's running around so much, even though she's not doing any gym or anything else, but it's just the house, it's more intense though" [FG2; F; 31; SOC 4]

Note:FG=Focus Group; F=Female; M=Male; Number= Age(Years); SOC=Stage of Change

A common perception pervaded that this type of activity is a means of "getting things done" (1, 2) and a form of work, perhaps reflected in it's description "chore" (2).The impression from these descriptions is that the activities undertaken are not done with any particular relish or for the purpose of enjoyment. There is intent in the execution of the activity but it appears to be only for the purpose of necessity, while the word "chore" as used by two of the participants, suggests an image synonymous with responsibility and obligation.

Examples of activities which were included under the labels of housework and gardening were varied but have recurred consistently across all groups (Appendix three, (3)). Another activity concerned with domesticity is that of "shopping" (4). Although this activity is not strictly carried out within the home, it has been included under this category because of its relation to domestic living. It is concerned with gathering and carrying the groceries and household items that are required for the cleaning, washing etc, back to the house. Contextual reference for shopping as a physical activity are also given (5).

Each of these examples concerning housework, gardening and shopping were given in relation to examples of types of activities that are performed at various intensities It appears that the intensity of the activity is often variable and is dependent upon the situation in which it is being performed (6) For one participant, doing "intense" housework was the cause of "losing weight" (7) This sentiment was not one generally expressed by other participants and is suggestive of a deviant case that does not typically represent participant's sentiments about this type of PA

In summary, the defining characteristic of a domestic activity is that it is "work" specifically concerned with the house (or its environs) The degree of intention involved in doing the activity is dependent upon the participant and the intensity (discussed later), but in general ranges across a continuum of intent and consciousness The purpose of doing the activity is however usually exclusively practical and concerned with "doing things" in order to maintain the garden, house and its contents The most popular adjectives used by the participants to describe activities that possess the properties identified in this "domestic" category are "housework", "gardening" and "shopping"

3 4 2 1 5 Occupational (Table 3 7)

The properties of these activities are similar to domestic activities in that they range across a continuum of intensity and intention They are distinct in their purpose Each activity that is performed is relevant to that person's occupation or is done to facilitate the completion of a task that is relevant to that person's job The type of activity and the intensity at which it is performed "depends upon your job, I suppose", [FG8, F, 49, SOC 3] There is awareness among the participants in the groups that some jobs would be less active than others while some jobs require a sustained amount of energy (1)

Table 3 7 *Full quotes from the raw data that detail properties of physical activity while doing occupational activities*

No	Quote
1	"[It depends upon] your occupation as well I work dealing with desk forms all day, he's a builder, Padraic is a teacher and his intensity might be classed as light, but it'd be a lot more intense than what I'd be doing I was working in a clothes shop and that was quite intense because you were lifting lots of boxes and you were up and down the stairs all day but now I'm sitting down at phones all day dealing with customers" [FG7, F, 21, SOC 5]
2	"Our jobs [nursing] then as well, we'd be very much a mix throughout the day" [FG3,F, 53, SOC 2]
3	(Person1) "Having a very active job or something like that"' (Person2) "Yeah that's a good one"' (Person1) " A lot of us are not exercising regularly, but we would be going to work everyday and being active in our jobs but we're not necessarily exercising everyday" [FG3]
4	"It's like it's actually I've never really been in a gym so I don't know (laughs) but it'd remind me of like I don't think we need gym membership because it's just your lifting books like this, you're walking, you're sitting down we're running upstairs, we're running downstairs it's very physically active" [FG1,F, 29, SOC 3]

Note FG=Focus Group, F=Female, M=Male, Number= Age(Years), SOC=Stage of Change

Even within one's own job, the participants have highlighted that intensity or particular activities can vary across the day and between days (2) In question 2, when the participants were asked to discuss what came to mind when they saw or heard the words "physical activity", two people made immediate reference to occupational activities (3)

The perception of physical activity in the workplace for some participants, can be synonymous with the type of activity experienced in a gym or sport setting (4) By their own admissions, neither of these participants are familiar with the event that they are comparing their work against However, it is their perception that their work activity contains many of the properties associated with a more structured, acknowledged form of activity They appear to be using "the gym" and "sport" as their criterion measures against which they measure the amount of physical activity they do at work Notably, both of these participants have very different occupations, which vary in their physical demands, i e , one is a librarian while the other is a builder and farmer

In summary, the participants usually refer to this work as "my work", "my occupation" or "my job" Activities in this category vary in type and perceived intensity depending upon the particular occupation, the time of day and the participants comparative experience with other jobs Occupational activities are performed on a continuum of consciousness, depending upon the activity and its relative intensity

3 4 2 1 6 Family (Table 3 8)

This category emerged following recurrent references to the effort involved "when you have kids" The types of activities vary according to the age and numbers of children involved but it was typically felt that "you are constantly active when you have kids because you're constantly going", [FG5, F, 39, SOC 5] Adjectives most frequently used to describe the physical activity associated with children are "constant" or "always on the go" This appears to be true for people with one child or for those with more than one (1)

Table 3 8 *Full quotes from the raw data that detail properties of physical activity while doing family activities*

No	Quotes
1	"I find more because I've only the one, that I'm always on the go trying to keep her entertained, do you know what I mean?" [FG5, F,35, SOC 5]
2	"My niece and nephew are up during the evenings, so I'm running around playing chasing with them and you know they have you wrecked after half an hour" [FG1, F,21, SOC 5]
3	"Most evenings I'd be more active at home bringing kids here, there well they're grown up now, here and there I'd spend more time cleaning up, tidying up and running after them" [FG3, M,51, SOC 5]
4	"This is Barbara and she has three children and at the moment that's what her physical exercise is chasing them" [F, 42, SOC 2] "This is Caroline She's got two children [and] that keeps her busy" [M, 30, SOC5]

Note FG=Focus Group, F=Female, M=Male, Number= Age(Years), SOC=Stage of Change

The types of activities performed when "you have kids" typically refer to "running around after them", either in the context of entertaining them (2) or in other contexts (3) Both of these examples represent very different types of "running" In the case of (2), she is physically engaged in playing with her niece and nephew and clearly feels the physical consequences of her actions, i e , "wrecked" The type of "running" number (3) is referring to, involves bringing his children to various places by car, yet he still perceives this as being "active"

Other activities that involve looking after children include organised activities like "playing football or curves in the back garden" and "playing with the children", which indicate an obvious intent in becoming active At other times, "caregiving" or "looking after kids" appears to be made up of normal everyday duties, like lifting up or carrying young children, making meals for them etc In FG4, two mothers were perceived by two single people (one male and one female) as being very active because of their children (4)

In summary, Family activity is typically described by the participants as "looking after the children" Properties of this category include activities that are varied in intensity and consciousness, depending upon the actual activity engaged in Looking after children is commonly agreed upon by the participants to involve "constant" care and there also appears to be general consent among both parents and others, that those in charge of children are "always on the go"

3 4 2 1 7 Incidental

This category refers to physical activity which is non-specific but rather involves a variety of typical everyday usual actions, "it's moving, it's just moving" [FG8, F, 50, SOC 5] The activities described in this category are examples given by the participants of the types of activities that they would undertake in various contexts/situations throughout their day, usually at a light intensity

The main feature of these activities is that they are usually performed at a subconscious level and many of the examples given include day-to-day things such as having a shower, getting breakfast, walking around the office or as one participant suggests, "it's just the little things" [FG4, F, 20, SOC 5] The general consensus among the participants is that most of the activities in this category are "just every day life" [FG2, M, 20, SOC 3] or "everyday activities" [FG 2, F, 23, SOC 5] The variety of activities done everyday, and included below, reoccurred at various stages throughout all of the discussion groups and so are mentioned here as, " Walking up and down the stairs", "Strolling", "Walking around the house", "Moving muscles", "Just toddling along" and "Just walking from room to room and that sort of activity" There is a sense that these are the types of things "that you do automatically, like get out of bed, walk to the bathroom, have a shower and wash your hair" [FG3, M, 51, SOC 5] or "moving your arms or your legs or something, just everyday activities" [FG2, F, 23, SOC 5]

In summary, the properties of activities in this category include doing things that require very little conscious or physical effort on behalf of the participant They involve doing normal everyday activities and mostly involve any type of "movement" which does not apply to any other given category and essentially involves anything that is equal to "once you're moving and doing something, not just sitting propped in front of the telly" [FG5, F, 35, SOC 5] Particular adjectives used to describe "Incidental" activities include

"just every day things" or "general movement" These are activities that are not necessarily done at a moderate or high intensity and so are not relevant to the current research question

3 4 2 2 Properties of Intensity

The type of activity performed and the context in which it occurs (Section 3 4 1), provides some information about one aspect of physical activity A more comprehensive description of physical activity would also provide details of *how* this activity is performed Just as there are a plethora of opportunities available in which one can do physical activity, so too are there various intensities in which one can engage and perform these activities Every activity, regardless of its context, has the capacity of being done at on a continuum of intensities, which range from low to moderate to high

Each of these intensities affect the participant in a variety of ways, both physiologically and psychologically The perceived effect however, was considered a very individual issue by the participants of these groups This section explores their perception of the differences between moderate and high intensities The emergent information from the participant's own experiences, provided a useful insight into how physical activity was interpreted and recalled, depending upon the intensity undertaken

3 4 2 2 1 Moderate Intensity

During the course of the discussions, many similar terms and phrases were used by the participants to describe what moderate intensity meant to them These adjectives described how performing physical activity at a moderate intensity affected them physically and psychologically "Moderate intensity" was also perceived as an experience that is relative to the individual, across all types of activity As participants spoke about moderate intensity, specific examples of activities performed at this intensity were mentioned recurrently across focus groups These examples were occurrences of the types of activity mentioned in Section 3 4 2 1 Representation of these examples and their associated properties, which emerged from the data, are illustrated in Table 3 9

Table 3.9. *Summary of examples of moderate intensity activities.*

Type	Example and Descriptions	Participant
Commuting	"Getting to work cycling"	FG4; F; 28; SOC3
	"Walking to work"	FG4; F; 31; SOC5
Sport & Exercise	"Swimming", "running", "gym-work", "tennis", "judo", "rugby"	All focus groups
Leisure	"Going for a cycle at the weekend"	FG4; M; 30; SOC5
	"Something like dancing"	FG3; F; 42; SOC5
	"Going for walks in the evening"	FG6; F; 30; SOC3
Occupational	"Different for every job"	FG3; F; 53; SOC2
	"Going up and down stairs"	FG6; M; 34; SOC5
	"Helping people to move around"	FG3; F; 53; SOC2
	"Moving machinery and boxes"	FG3; F; 53; SOC2
	"Lifting stock around"	FG7; F; 21; SOC5
	"Lifting and laying blocks"	FG6; M; 42; SOC3
Housework & Gardening	"Hoovering is moderate for some"	FG3; F; 53; SOC2
	"Hoovering against a resistance"	FG3; F; 41; SOC5
	"Washing the window like you're taking out a vengeance"	FG3; F; 30; SOC2
	"Working against a weight, like digging, lifting or raking or moving the mower around"	FG4; F; 41; SOC3
	"Shopping, carrying your groceries home"	FG4; F; 20; SOC5
Family	"Running around"	FG3; F; 30; SOC2
	"Playing with the kids"	FG2; F; 29; SOC5
	"Carrying a child"	FG1; F; 36; SOC3
Incidental	"Not doing things slowly"	FG1; F; 32; SOC4
	"Dashing here and there"	FG3; M; 51; SOC5
	"Instead of taking the stairs, I take the lift"	FG1; M; 31; SOC5

Note:FG=Focus Group; F=Female; M=Male; Number= Age(Years); SOC=Stage of Change

Participants were asked to gather into groups of three to discuss the following question and then report back to the main group, "When performing an activity at a moderate intensity, what changes do you notice about your body in comparison to the resting state, as you are now?" Responses were varied but essentially were concerned with changes that were categorised as either *physiological* or *psychological* in nature. This section explores the individual physiological and psychological changes of moderate intensity activity, as described by the participants.

3 4 2 2 1 1 Moderate Intensity Physical Descriptors (See Table 3 10)

Within this category, physical changes that participants noticed about themselves were separated into two distinct categories Initially, open coding (stage one, see methodology section) identified similar words and phrases that were expressed in the participant's vocabulary These were then grouped (axial coding) and assigned a common node title The distinguishing properties of each of these nodes centred around whether the physical changes were visible or not to an observer This distinction was made simply to assist in the coding and grouping of physical effects

A Visible to observer

All focus groups referred to changes in breathing pattern, facial colour and perspiration For each of these items, the participants described them in very similar terms, at times using the same words

i Breathing pattern

There was an understanding among the participants that doing activity at a moderate intensity alters the regular pattern of breathing, causing the individual to "breathe deeper" (1) There does appear to be a discrepancy between groups as to whether the individual would be classified as "out of breath" or "breathless" While some groups did feel that this was a possibility, it was sometimes offset by the inclusion of the word "slightly" The extent of enhanced breathing was put into context by two participants in FG8 (2, Table 3 10) The inference from this dialogue is that one could not be "out of breath" if also expected to be able to talk, suggested by ,"taking deeper breaths and stuff, but not out of breath" [FG3, 30, SOC 2] In summary, this physical change could be described, from participant's vocabulary, as "slightly out of breath"

ii Facial Colour

This visual physical change is described in very similar ways by all groups Descriptive terms range from "colour change", "high colour", "red face and flushed" In two groups (F G 2 and F G 1) the term "heightened (or high) colour" was synonymous with "flushed face" (3) No other descriptive terms were used in reference to this physical change This physical change cold be referred to as "slightly flushed"

iii Perspiration

Increased perspiration is a property of moderate intensity, although ways of describing its appearance are varied (4) While a noticeable difference is felt, it was commonly agreed that it is only the beginning of a potentially larger occurrence This is evident from the use of words such as "start", "breaking", "slight" and "light" (see further descriptions of "sweat" in Table 3 10, Appendix three) Apparent from these descriptions is the concept that moderate intensity physical activity causes the body to begin the production of sweat In summary, an appropriate phrase from participant's descriptions could be "slight sweat"

Table 3 10 *Full quotes from the raw data that detail properties of moderate intensity physical affect*

No	Quote
1	"Slight breathlessness until you kind of, until the rhythm starts going" "Breathing gets deeper" [FG8]
2	"Like walking and being able to hold a conversation at the same time, without getting breathless, but not walking very slow, walking really quick and keeping it up, that would be moderate" [F, 40, SOC 3] "I think moderate is, you know like she said, being able to do the exercise and being able to have a conversation as well" [M, 29, SOC 5]
3	"Heightened colour, that's flushed" [FG1, F, 29, SOC 3] "Flushed face, that can be high colouring" [FG2, M, 20, SOC 3]
4	"A light sweat" [FG8, M, 25, SOC 5]
5	[FG1] "I think pain is slightly extreme when you're describing it, if it's only moderate" [M, 50, SOC 5] "Well moderate is relative, isn't it though?" [F, 42, SOC 3] "Maybe discomfort " [F, 32, SOC 4] "Yeah, discomfort would be better" [M, 50, SOC 5]
6	"I suppose for me the main thing that I would notice is there would be a certain suppleness in your body in relation to say leg muscles or whatever if it's walking that you're doing" [FG6, M, 52, SOC 5]
7	"But something that strikes me is that while somebody might be physically fit for one type of activity, it does not necessarily follow that they're fit for another I was out for a walk with an intercounty footballer and superfit and after a number of hours walking in the Burren, he sort of said, oh my God! There's a fitness requirement here that isn't in football, and I said, yeah you're probably not being challenged He thought no, it was the contrary There are muscles in my legs that are being tested now that would not normally be" [FG6, M, 52, SOC 5]
8	[FG1] "You'd get warmer Is that flushed probably is it?" [F, 42, SOC 3] "It's different to flushed" [F, 29, SOC 3] "Yeah" [F, 31, SOC 5] "We don't have anything else do we? We've just got hot down as well, it's probably the same" [F, 32, SOC 4] "Yeah warmer" [F, 31, SOC 5]
9	"You reach a comfortable temperature the example I have is if you go out on a cold morning and you go out for a walk and you're cold and you feel like you should put another

	jacket on? After a while you hit a temperature where you feel fine, you know, but you're not bucketing sweat, you know, you're not overheating you're just a nice comfortable temperature" [FG2, M, 31, SOC 4]
10	"Your body would be more drained than when you are just sitting here and now and obviously your body is losing something maybe fluids or something" [FG5, F, 19, SOC 5] "Dryness or thirst" [FG8, F, 31, SOC 3]

Note FG=Focus Group, F=Female, M=Male, Number= Age(Years), SOC=Stage of Change

B Not visible to an observer

Moderate intensity is believed to affect heart rate, muscle tone, body temperature, hormone response and hydration status These are node titles that emerged following the initial open coding of the data that identified similar descriptions of physical affect which are not visible to an observer

i Heart Rate

There is a general consensus that heart rate increases above normal values following initiation of moderate intensity At times it was referred to as "heart rate (or beat)", other times it was referred to as "pulse" In general, it was described in terms of "an increase in heart rate" or "your heart beats faster" and also "a raised heart rate" or "stronger heart rate" [All focus groups] Participants in FG1 spoke of "you feel your heart going", indicating an awareness of increased incidence When the word "pulse" is used [FG's 2, 3, 7], it is in the context of "your pulse increases" and "they'd have an increase in pulse maybe" Typically this event can be described from participant's vocabulary as "an increased heart rate/beat"

ii Muscle Tone

The effect of moderate intensity on the muscular structure of the body is described in different terms For some it was a mere "tightening of the muscles" [FG8, F, 31, SOC 3] while for others it was a process of "exercising the muscles [until] you feel muscle ache" [FG8, F, 45, SOC 5] The interpretation of "ache" is subjective but appeared to mean more of a pressure on the muscle than involving any negative entity (5) There was a perception of tension or stress on muscles when performing physical activity at a moderate intensity However, further discussion signified a difference between the muscular effect in activities that are familiar to the participant and those that are unfamiliar If one is used to performing a particular activity, there is no obvious tension on muscles in the body (6) However, in situations where the activity is new or challenging to the participant but still at a moderate

intensity, the reaction can often be perceived as less favourable to muscles (7). The extent of the effect of moderate intensity on muscular structures is contextual and relative to the person's experience of that activity. As a result, there is no typical description of muscular effect, but could be explained with a reciprocal depiction such as "stretched or tense muscle". This is an image that is hard to communicate and may be open to ambiguities on a self-administered questionnaire. Thus, it has been omitted from this questionnaire.

iii. Body Temperature

Descriptions pertaining to this outcome of moderate intensity all indicated a rise in core body temperature. Descriptive terms were recurrent across all focus groups and included "your body warms up" and "becoming warmer" [FG8]. In FG1, participants teased out this concept a little more (8). These thoughts were expressed in similar fashion in FG5, where terms such as "a higher body temperature", "a slight increase in body temperature" and "we'd be that much warmer" were apparent. The synonymous use of words like "hot" and "warm" by the participants may appear conflicting. This may be due to a contrast in descriptive terminology. In FG3, one participant said of moderate intensity, it "makes you hot initially, makes you feel warm" [41; F; SOC 5], while a participant in FG2 suggested, "you'd get warmer, hotter type of thing" [20; M; SOC 3]. The central thought emerging from these expressions is that there is an increase in body temperature above the norm. An analogy to explain this thought is given in (9).

Moderate intensity activity educes an increase in core body temperature, which is apparent to the participant by a feeling of warmth. Although the word "comfortable" was only used once, there was no apparent discomfort associated with the use of "warm" or "hot". Therefore, a typical description from participants vocabulary could be "comfortable increase in temperature" or "feeling comfortably warmer".

iv. Hormone response

There appeared to be an awareness among some of the participants that moderate intensity activity elicits a hormonal response within the body. The exact nature of that response was unknown but was referred to on occasion as "an adrenaline thing" [FG1; F; 32; SOC 4] or having "some adrenaline" [FG7; F; 46; SOC 3] and also as getting "the adrenaline pumping" [FG2; M; 20; SOC 3]. Reference was also made to other hormones, i.e., "Endorphins…the happy hormones" [FG3; F; 29; SOC 5]. While there is no clear cut

explanation of the effect that an increased hormone release has on the body, there was an implication from the context of the discussion that it is a positive impact and one which assists the participants to perform at this intensity The infrequent mention and lack of detailed discussion about this effect suggested a lack of common knowledge about method of action or terminology associated with "the happy hormones" Their effect may be better described in relation to the psychological benefit of "feel good" (discussed below, section 3 4 2 2 1 2) In summary, the physical effect of enhanced hormone production may be described in conjunction with psychological perceptions as "feel good factor" or singularly by "increasing adrenaline"

v Hydration Status

This effect of moderate intensity on hydration appeared to be either minimal or not considered to be one of the more prominent effects of participation Reference to hydration was made only twice during the course of these discussion groups and even in these instances was mentioned only in passing (10) Dehydration may be an outcome of participation at moderate intensity Its impact however, while not disputed by other participants in the groups, was not discussed in any great detail Perhaps the significance of dehydration was only apparent in certain activities or is dependent upon duration of the event As this point was not discussed, these can be only assumptions In summary, "slight thirst" could be used to describe the effect moderate intensity has on hydration status, although it is a less prominent physical effect than others previously mentioned and is therefore not included in this questionnaire

3 4 2 2 1 2 Moderate Intensity Psychological Descriptors (See Table 3 11)

Within this category, psychological responses that participants notice about themselves can be separated into two distinct categories The distinguishing properties of each of these categories centre around whether the responses have a positive or negative affect on their psychological form These distinctions were made for the purpose of grouping the emergent data into secondary axial nodes, following the open coding stage of unstructured data (see methodology, section 3 3 2)

A Positive affect

Adjectives used to describe the positive feelings associated with moderate intensity were varied and plentiful While some expressions have very similar meanings, the sentiments were articulated using different phrases The principle phrases referred to changes in mental focus, confidence, well-being and mood

i Increased awareness

Many participants made reference to an increased awareness or alertness Another emergent theme from the participants' comments was an increased, though not exclusive, concentration on the activity at hand (1) Although words like "aware", "alert" and "focus" were used within the same discussion, there seemed to be subtle differences between them Aware and alert suggested a heightened mental state They also seemed to have a property of readiness or enhanced sensitivity to other stimuli, akin to "a sense of awakening", [FG8, M, 25, SOC 5] whereas "focus" suggested a narrowed degree of attention, "you could be doing something like warming up for judo and you're focussing on what's right in front of you", [FG4, M, 30, SOC 5] For this same participant in FG4, the degree of focus or concentration "depends on what you are doing [lesser intensity] could also be when you're just doing it on autopilot as well, you're not really, kind of, concentrating hard on what you're doing Whereas more intense stuff you're really watching or physically concentrating on what you're actually doing" Suggestion of a continuum of concentration for a range of intensities is in accord with previous discussion This participant referred to lower intensity activities being done "on autopilot", inferring a kind of subconscious level As the level of intensity increases, it appears that the level of conscious involvement also rises The difference between "focus" at moderate intensity and high intensity is outlined by two participants (2) (See also Appendix three)

ii Mood enhancer

Descriptions in this category explored an association of sentiments For some participants moderate intensity meant enhanced self-confidence that also appeared to effect levels of motivation While many of the words and terms used here appear varied, many of them share a common property of positive alteration in mood, which is noticeable from the tone of the discussion

A recurrent term in this category was "feel good" [FG8, M, 29, SOC 5] Two other participants from different focus groups reinforce this affective quality, "I feel so good after it [exercise]" [FG7, F, 36, SOC 4] and, "I feel so much better afterwards" [FG2, F, 23, SOC 5] A reason for this positive feeling may be the sense of enjoyment, which is widely expressed in all focus groups (3) Activities that are "enjoyed" appear to relate more to the leisure or sport and exercise domains rather than the more functional domains like occupational or domestic (4) One participant associated her enjoyment with an activity (walking) that she does because " I wanted to get fit and all I know is that I enjoyed it" [FG7, F, 36, SOC 4] On the other hand, there is no direct mention in this context of any moderate intensity activities that are work or home based In FG4, a female [37, SOC 5] makes reference to "being in a frenzy" when cleaning the house where "sweat is dripping off you" While this may relate more to a high intensity activity, she does at least admit feeling a sense of satisfaction It seems therefore, that activities that are voluntarily undertaken for reasons of health, fitness or sociability tend to elicit a more enjoyable response, as noted by the following discussion among FG3 participants (5) In general, there is no such positive affect mentioned for activities that are done as a matter of necessity

Elevation in mood was also evident elsewhere, "An uplift in your mood, you can feel happier doing something good" [FG4, F, 31, SOC 5] In this case, moderate intensity is responsible for the happiness of the participant Other words used in this context included "pleasurable" [FG7, F, 37, SOC 5] and "nice" [FG6, M, 52, SOC 5] Moderate intensity is also described in FG3 as feeling "invigorating", "stimulating", "motivating" and "positive" A sense of confidence is also associated with doing moderate intensity physical activity For some though, it appears this "confidence" is akin to a level of competency required to perform the activity (6), while for others it appears to be an outcome of performing activities at a moderate intensity (7)

In summary, moderate intensity physical activity appears to be a relatively positive mood enhancer, particularly in regard to activities not undertaken for the purpose of necessity and practicality Positive outcomes of participation include a sense of enjoyment, happiness, improved concentration and focus and for some, there may be an improvement in self-confidence A term that could be used to describe these words collectively might be "feel good"

iii Well-being

Although many of the terms referred to above in "mood enhancer" could equally apply in this section, and vice versa, the terms described in this section were concerned mostly with issues of stress and relaxation, contributing to an overall "sense of well-being" [FG3, F, 41, SOC 5] For participants in FG7, the lack of stress was quite a significant factor (8) It appears from this text that moderate intensity is synonymous with being in control of ones' own body and stress levels The activity is performed at a pace that the participant controls, resulting in a feeling of relaxation and a non-stressful state All of this would seem to contribute to an overall sense of well-being (9) "De-stressing" was a word that is also used by this participant [FG2, M, 29, SOC 3], " [it] just has, a de-stressing effect on your body" Frequently used words by the participants, to collectively describe these sensations include, "uplift mood", "feel good" and "enjoyment"

Table 3 11 *Full quotes from the raw data that detail properties of moderate intensity psychological affect*

No	Quote
1	[F G 8] "You get more alert" [F, 30, SOC 5] "You're more alert and you're aware of your breathing and your movement" [F, 27, SOC 5]
2	(Moderate) "Slightly focussed on activity, not fully focussed on what you were doing, you might be wandering a bit" [FG7, F, 46, SOC 3] (High) "You're focussed directly on one item, not just concentrating but you're focussing only on one item at a time Everything else is filtered out " [FG4, M, 29, SOC 5]
3	[F G 8] " finding a level where they actually enjoy it" [F, 50, SOC 5] "Yeah, I think they need to enjoy the activities" [F, 27, SOC 5]
4	"It's like an effort in itself but once you're there you enjoy it and when you are finished, you say ah yes! I can't wait until the next class " [FG2, M, 20, SOC 3]
5	[F G 3] "It's probably a social thing" [F, 30, SOC 2] "Absolutely, there's a social aspect as well" [F, 37, SOC 3] "Because yes I think it would be very enjoyable, there'd be no sort of tense feelings or anything, just enjoyment" [M, 51, SOC 5]
6	"Well for me, it would be something that you are confident doing, that you can continue to do" [FG6, M, 52, SOC 5],
7	"You'd feel more confident as well, like when you start walking and once you start getting into a rhythm, you would feel like, you would get into focus basically and confident" [FG2, M, 20, SOC 3]
8	[FG7] "[It's] not stress related" [M, 42, SOC 2] "[For] moderate, you'd be just doing it at your own pace and you wouldn't be worried about time" [M, 25, SOC 5]

	"The moderate one would be more relaxing because you are doing it at your own pace", [M, 55, SOC 5]
	"Basically, that's it you know, pacing yourself but maybe a little bit of relaxation So, if it's for a long walk, you know, you'd pace yourself" [F, 42, SOC 1]
9	[F G 3]
	"[It's a] sense of well-being" [F, 41, SOC 5]
	"That'd be the same as pleasantly relaxed, wouldn't it? A sense of well-being then" [M, 51, SOC 5]
	"Which ties in with we said a sort of de-stressing, aware that you're de-stressing" [F, 41, SOC 5]
10	"I think an awful lot of it has to do with your humour as well If you're in good form, things you know I feel it's an awful lot easier to do things Whereas if you get up and you're, Oh God! you know everything is just a drag, well then I think it takes more out of you to do something" [F, 37, SOC 3]
	"Yeah that's true" [M, 51, SOC 5]
12	[FG1]
	"Sorry, the way we described it was more that it depends on the level of fitness, that if they're very unfit " [F, 32, SOC 4]
	"That's what I meant, depends on their weight and their body" [F, 31, SOC, 5]
	"Yeah so moderate is different for different individuals, isn't it?" [F, 42, SOC 3]
13	"Yeah, for what one person might be moderate, might be very high for somebody else" [FG8, F, 31, SOC 5]
	"So the same exercise for different person, it can be moderate or intense or even easier for some" [FG3, F, 41, SOC 5]
14	[FG 7]
	"I would think moderation is inside the individual What is moderate, whether you're running or walking in a moderate way, moderation is from within so it's equal" [FG7, M, 47, SOC 3]
	"It's equal for each person" [FG7, F, 46, SOC 3]

Note FG=Focus Group, F=Female, M=Male, Number= Age(Years), SOC=Stage of Change

B Negative affect

Perceptions of physical activity intensity appeared to be dependent on mood and circumstance For some, the physical ability to complete an activity was linked to the relative psychological well being of the individual at that time (See Appendix three (10)) The perception among this group was that the ability to physically perform moderate intensity (or maybe any level of) physical activity is preceded by the psychological status of the individual Specifically, psychological ability affects physical ability but this relationship is not reciprocal This can be seen in one participants' analogy of the classroom setting, where he has done the same amount of "physical exercise", but whether or not he feels tired will depend on how mentally tough the day was Similarly with a female participant in the same FG, who states that if she is not enjoying the days work, she'll go home "feeling shattered"

It seems therefore that the perception of physical activity has a lot to do with "your form", as stated above Participants in FG3 tend to agree (10) This suggests that the perceived physical exertion of the activity is predetermined by the psychological status of the individual at the time Therefore, physical activity can be perceived as a negative experience if the individual is feeling tired or in "bad form" prior to engaging in the activity The underlying factor may be the perceived amount of "effort" required (11 (Appendix three)) One participant suggested that perceived ability to do the activity is a transient state, "somedays" you can do it Other participants also suggested that it is not a permanent state, but is dependent upon whether or not you're "mentally into it" If a person is not in a good mood or humour before they commence the activity, it seems that physical activity of moderate (maybe higher) intensity assumes a negative affect

In summary, although some activities may be eliciting physiological responses that are associated with performance of a moderate intensity, the perceived intensity may not be in concordance A negative perception of the activity may follow if the participant is tired or is "in bad form", "bad mood" or not "in humour" The perceived effort of the activity is affected by these factors and also by the relative enjoyment factor and novelty of the activity

C Interpretations of Moderate Intensity

A description given of "moderate intensity activity" was that it is "different for different individuals" [FG1, F, 42, SOC 3] The type and amount of activity will "depend on the individual as well You know, what's intense for one person won't be for another" [FG6, F, 30, SOC 3] In this way, it "depends on their level of fitness" [FG1, F 32, SOC 4] and "on their weight and their body" [FG1, F, 31, SOC 5] Participants also discussed the influence of fitness and weight on performing moderate intensity (12) The amount of activity appropriate to moderate intensity is therefore considered "individual" by these participants The type of activity performed however, is not specific to the moderate intensity category but can instead vary across intensities (13)

Relative familiarity of the activity will determine whether it's moderate for the individual while there was also a suggestion that relative fitness in some activities doesn't necessarily transfer to other activities, " something [that] strikes me is that while somebody might be physically fit for one type of activity, it does not necessarily follow that they're fit for another", [FG8, M, 41, SOC 5]

It was suggested that the effects of moderate intensity on the individual are perceived as similar for everyone One participant compared two people of varying fitness doing the same activity, " it's the same stresses on your body but you'll get it done in half the time" [FG7, F, 46, SOC 3] This suggested that people experience the same effects, but one persons superior fitness is advantageous in reducing the time to complete the same activity, at the same relative intensity as the less fit person

In summary, length of time engaged in completing the activity appears to depend upon fitness levels, familiarity of event to the individual and body weight, but the effects are usually the same for each individual Recurrent adjectives for describing moderate intensity, throughout all focus groups, included, "relative", "individual", "own fitness level", "everyone is different" Each of these terms referred to a basic understanding that the performance of moderate intensity physical activity is unique to the individual doing it, as described by participants in FG7 (14)

3 4 2 2 2 High Intensity

Participants were asked about specific examples of activities they performed at high intensity Similar examples were mentioned recurrently across focus groups and were, at times, identical to examples of activities also identified for moderate intensity Differentiation of these activities by intensity is apparent from adjectives used to describe them, e g , "lifting very heavy things" Representation of these examples and their descriptions which emerged from the data, are illustrated in Table 3 12

Table 3.12. *Summary of examples of high intensity activities.*

Type	Example and Description	Participant
Commuting	"Walking to work when you are late"	FG4; F; 23; SOC3
Sport and Exercise	"Aerobics", "Swimming", "Running", "Gym", "Weights", "Cycling", "Power walking"	All focus groups
Leisure	"Walking, like hill walking" "Clubbing and dancing"	FG6; M; 52; SOC5 FG3; F; 41; SOC5
Occupational	"Lifting very heavy things, like with builders and farmers"	FG6; M; 52; SOC5
Housework and Gardening	"Cleaning the window like you're taking out some sort of vengeance" "Doing hard gardening"	FG1; F; 42; SOC5 FG4; M; 31; SOC5
Family	"Entertaining a child is high intensity"	FG2; M; 20; SOC3
Incidental	"In a general hurry" "Running up the stairs instead of at a normal pace"	FG3; M; 51; SOC5 FG6; F; 54; SOC3

Note:FG=Focus Group; F=Female; M=Male; Number= Age(Years); SOC=Stage of Change

During the course of the discussions, the participants used many similar terms and phrases in order to describe what high intensity meant to them. These adjectives helped to describe how performing physical activity at a high intensity affected them physically and psychologically. This section explores these perceptions as described by the participants. Questions relating to high intensity were asked in the context of how it was different to moderate intensity.

3.4.2.2.2.1. High Intensity Physical Descriptors (See Table 3.13)

Physiological properties of "high intensity" generally referred to a state of extreme tiredness/fatigue. There were also infrequent references to other physical attributes such as heart rate and general fitness. High intensity relates to reaching a personal limit of ability, going outside of that person's comfort zone. Physically, the person is very tired, with a few participants describing the sensation as feeling "wrecked" (1). Other descriptive terms were also used that indicated a performance threshold (2). From these descriptions, there appears to be a perception that the individuals must "push" themselves to reach or maintain a level

of high intensity physical activity This indicates a conscious intent of engagement for many to bring themselves to their "limit", and in one case, to go past his limit Imagery associated with these terms infers a level of discomfort and maximal effort, suitably described by one participant as a state of near collapse The maximal effort required by the individual is depicted as, "you put everything in, 100%, until you can't move anymore, but that's it", [FG4, F, 20, SOC 5] Other descriptors of physical affect reflect a level of pressure or strain associated with this level of intensity and affirm the previous references to going beyond a comfort zone (3)

The excessive demands of high intensity activity was reflected by the term "extreme" to describe it [FG4, M, 29, SOC 5] He suggested that "extreme exercise is shockingly demanding" For many, the ability to cope with the stresses and pressures placed on the body during high intensity, as referred to above, is a reflection of the fitness of the individual There was a common perception that an individual must be fit to perform prolonged high intensity and also have a certain level of fitness to do any small part of it (4) The improved fitness of the individual may occur because of the properties attributed to high intensity activities, i e they were described as "faster and harder work" [FG8, F, 19, SOC 5] and "you'd be speeding up" [FG7, F, 46, SOC 3] Other physical properties of high intensity indicated that the types of activities done at this intensity are "very, very short sports" [FG8, M, 25, SOC 5], because "you couldn't keep it up for long" [FG3, F, 41, SOC 5]

In summary, adjectives used to describe the physical properties of high intensity physical activity suggest demands from the individual that are maximal and tiring, pushing the individual up to or over their comfort zone and their limit of ability One female participant referred to it as "when you reach the burnout stage" [FG6, F, 52, SOC 2] Frequently used words included "wrecked" and "limit"

Table 3 13 *Full quotes from the raw data that detail properties of high intensity physical affect*

No	Quote
1	"They'd have you wrecked high intensity wouldn't that be if your heart is beating very fast, like the most intense that " [FG8, f, 19, SOC 5]
2	"High intensity would surely mean pushing oneself beyond one's comfort zones, feeling under stress and under strain physical strain" [FG8, F, 45, SOC 5]
3	"I suppose for me, it's a case of the first 5 or 10 minutes of that exercise would be the hardest until the adrenaline started kicking in, or the endorphins start to kick in because it's kind of beyond your fitness level" [FG8, M, 29, SOC 5]
4	"It's the fitter you get isn't it some say you have to be fit to do a high intensity workout" [FG8, F, 49, SOC 3]

Note FG=Focus Group, F=Female, M=Male, Number= Age(Years), SOC=Stage of Change

3 4 2 2 2 2 High Intensity Psychological Descriptors (See Table 3 14)

The psychological properties of high intensity refer mostly to a degree of consciousness that is very focussed and specific This may be an indication of the difficulty of the task (1) Terms such as "concentrate", "focus" and "mindset" indicate purposeful conscious intent to engage in this activity The effort involved in doing the activity is apparent when words such as "push" and "haze of intensive" are used The narrowing of focus appears to help the individual to use all available resources to complete the task (2) Other infrequent references to a negative enjoyment affect are also made Exerting maximal effort to the point of leaving one's comfort zone however, seems to compromise the enjoyment factor which was present in moderate intensity activities (3) It would appear from these comments that the purpose of high intensity activity is for something other than enjoyment One participant likes the idea that she's done it, but feels that she cannot do that level of activity at home, because of the amount of discipline it requires (4)

In summary, adjectives used to describe the psychological properties of high intensity physical activity suggest specificity in concentration and intentional focus The demanding nature of this intensity removes the enjoyment factor for many participants Key words used for descriptive purposes include "focus" and "concentrate"

Table 3 14 *Full quotes from the raw data that detail properties of high intensity psychological affect*

No	Quote
1	"Whereas high intensity would be to do that exercise, you would have to concentrate totally on that one exercise, you know, no interruptions and you have a mindset just to push yourself to get through it" [FG8, M, 29, SOC 5]
2	"You'd be more focussed on what you want to do, you'd be more focussed on achieving a certain aim Getting from A to B within a certain amount of time or whatever" [FG7, M, 25, SOC 5]
3	"[It's] not finding a level where they actually enjoy it" [FG8, F, 50, SOC 5]
4	"Well, I have to do that yeah, but I don't have the discipline at home I have to do exercise that's of a high, whether I want to do it or I don't" [FG3, F, 37, SOC 3]

Note FG=Focus Group, F=Female, M=Male, Number= Age(Years), SOC=Stage of Change

3 4 2 2 3 Conscious Intent in Undertaking Activity

The term "physical activity" has been described by the participants of these focus groups in terms of the type and intensity at which the activity is done During this process however, the participants also made indirect reference to the degree of intent or consciousness engaged in, while doing moderate and high intensity activity Descriptive phrases illustrate that activity at both intensities is the product of purposeful intent, the extent of which is superior for high intensity In undertaking moderate intensity leisure activity, one participant stated that "something like dancing might get me to moderate and then that could be sustained" [FG3, F, 42, SOC5] The phrase "get me to moderate" infers that a conscious decision has been made to do an activity that is something different or above normal A degree of effort being applied to certain domestic activities, by "trying to" do something is suggestive of some awareness that the activity has been consciously undertaken "As soon as you're [hoovering] on a carpet and especially if you are trying to lift something that's in the carpet, rather than just on the carpet, then you start to feel like you're working against it", [FG3, F, 41, SOC5] In this case, effort has to be exerted, indicating a degree of consciousness about the activity This notion is substantiated by comments from three participants from different groups, talking about participation in various activities

1 "Shopping you know a 10 minute walk and carrying your groceries It's just moderate, you've made an effort to carry them but it's not like championship weightlifting or anything", [FG4, F, 20, SOC5]

2 "I remember going for walks in the evening, which I would class as moderate because I was pushing myself just that little bit extra, where I felt it physically", [FG6, F, 20, SOC3]

3 "If you're talking about walking, it's moderate intensity I don't think you're just walking without thinking about it I think you're making a conscious effort to you know, put some effort into it", [FG2, M, 31, SOC4]

The degree of intent or consciousness at this level is noticeably less than that for high intensity activity Quote number one above infers there's a level above carrying shopping that's comparable to weightlifting In other words, when performing "moderate exercise, you wouldn't have realised your full potential, because you haven't really been tested", [FG4, F, 29, SOC2] Other participants use words like, "fairly", "quite" and "somewhat" when describing moderate intensity In contrast, high intensity activity is perceived as "extreme", "maximal" and "tiring" Specific words like "burnout" and "wrecked" suggest an endpoint in terms of possible effort from the participant The psychological aspects of participation in high intensity PA, which have previously been discussed (Section 3 4 2 2 2), also suggest that a decidedly focussed and intentional effort has been undertaken to complete a high intensity task Table 3 15 provides a summary of the properties of categories which emerged from the focus groups

Table 3 15 *Summary of properties of PA categories*

Type	Intensity	Conscious Intent	Adjectives
Commuting	Varied, usually moderate	Moderate-High	"Going to " "To get to "
Sport & Exercise	Varied, usually moderate-high	Moderate-High	"Sport" "Fitness" "Exercise" "Gym"
Leisure	Varied, usually moderate	Varied	"For leisure" "at your own pace"
Housework	Varied, depending on activity	Varied	"Housework" "Gardening" "Shopping"
Occupational	Varied, depending on job	Varied	"My work" "My job" "My occupation"
Family	Usually moderate-high	Moderate-high	"Looking after the kids" "Constant"
Incidental	Usually low	Usually low	"Just everyday moving" "The little things"

Note FG=Focus Group, F=Female, M=Male, Number= Age(Years), SOC=Stage of Change

3 5 Discussion

Accurate retrieval of autobiographical memories is dependent upon how the event was encoded at the time of occurrence as well as the type and number of cues that are offered at the time of retrieval (Tourangeau et al , 2000, Wagenaar, 1986, Linton, 1982) This study was concerned specifically with the way participants understood (encoded) and described their participation in "physical activity" at "moderate" and "high intensity" Recurrent terms and phrases were used across all focus groups by participants to express the type of PA that they do, when and where they do it, as well as how they feel when they do it These answers provided the basis for identifying the specific recall cues that should assist them in retrieving memories of performing PA The following discussion describes the description of PA and the assimilation of recall cues to produce a draft version of Physical Activity Recall Questionnaire (PARQ) (Appendix five) For clarity, this is undertaken in three sections, i e , the findings that relate to the cues associated with 1) the type of PA, 2) the intensity of PA and 3) the general layout of PARQ

3 5 1 *Types and Examples of Physical Activity*

Results from this study revealed that participants readily described and recalled PA in terms of "type" of activity Answers to the question, "What does the term "physical activity" mean to you?", freely elicited information about incidences of performing activity in various contexts Exploration of the recall of PA in young adults revealed a similar natural tendency, without being specifically probed, to refer to the type of activity that was undertaken (McKenna, Page & Foster, 2004) Of the seven types of PA that emerged from the focus group data in this current study, only six were deemed applicable for inclusion in this questionnaire This was because "Incidental activity" which emerged from the discussions tended to be associated with light intensity or "subconscious intent" and therefore was not relevant to the main research question of this dissertation, i e , physical activity of at least moderate intensity (Pate et al , 1995) The accuracy of recalled activities that are performed at an intensity level that the participant is not even aware of (i e , less than moderate intensity), raises issues about the inclusion of this intensity of activity and requires further investigation The types of activity identified in this study were

1 Housework and Gardening
2 Sport and Exercise
3 Family
4 Leisure
5 Occupational
6 Commuting physical activity (moderate intensity)
7 Other physical activity (high intensity)

Categories and titles listed in this study are similar, though more extensive, than those depicted in other recently developed physical activity recall questionnaires, e g , International Physical Activity Questionnaire (IPAQ), (Craig et al , 2003), Lifetime Physical Activity Questionnaire (Friedenreich et al , 1998), or the Scottish Physical Activity Questionnaire (SPAQ), (Lowther et al , 1997) The Lifetime Questionnaire is limited to collecting data on occupational, sport and exercise and household activity, while the IPAQ seven-day (long version) is concerned with housework, commuting, occupational and recreation/sport The Kaiser Physical Activity Survey for women (Ainsworth et al , 2000) reflects similar categories to PARQ, in that as well as being inclusive of housework, sport and occupational activities, it also seeks information about caregiving activities In this current study, "family activity" was identified as offering another opportunity to be physically active during the day, and is comparable to Ainsworth and colleagues "caregiving" category The term "other physical activity" was used as the title of the last category in the high intensity question and referred to activities that were not relevant to other categories but were specific to high intensity only, such as "running up stairs" and "running when late"

The method of item generation employed for each of the above questionnaires is poorly represented in literature, therefore it is often not clear how the questionnaire designers arrived at these categories, or why some categories were included while others were excluded Implications from research that has explored participant's understanding of terms and phrases, indicates that researchers need to talk to their target population and, in the case of questionnaire design, include them in the development process to ensure population specificity (Tudor-Locke et al , 2003, Warnecke et al , 1997) Ainsworth (2000) has also indicated the importance of taking a comprehensive approach to including categories and types of physical activities that are relevant to women and other groups, such as ethnic

minorities The outcome of this study (a draft version of PARQ) has evolved following consultation with members of the target population, on topics relevant to the research question, i e , lifestyle physical activities of moderate and high intensity

Examples of activities for each category were included in the questionnaire These were provided by the participants and were included as a recall cue to assist in remembering "what" activity was performed This information was freely given by participants without need for specific prompting, indicating that information pertaining to "type" of activity has an effective encoding procedure Barsalou (1988) also found that the provision of a "what [went on]" cue, outperformed other types of cues in the retrieval of accurate information In some cases, the same examples of activities were mentioned for both moderate and high intensity questions In these cases, participants described different levels of effort employed to complete the task, to distinguish between the intensities This is reflected in the "lifting" example given in the "housework and gardening" section, where the high intensity example includes "lifting furniture" and the moderate intensity example includes "lifting or carrying shopping", or in the "occupational" category where "heavy lifting" is contrasted with "lifting and carrying small things" Where the same example or type of activity was identified for both moderate and high intensity, distinctions between the way both intensities were encoded, was made apparent by the associated descriptions of physiological and psychological responses These descriptions and other properties of intensity are discussed below

3 5 2 Intensity of activity

Findings from this current study showed that when probed specifically, participants can offer information that has been stored or encoded about PA at different intensities This provided information about how participants understand and define "moderate and high intensity" and thus what cues may trigger the recall of activity at these intensities Descriptions of intensity centred specifically around physiological and psychological responses, e g , heart rate, breathing, colour, concentration and enjoyment Discrimination between intensities was expressed through words like "very" for high and "slight" for moderate, (e g , very fast heart rate, slight increase in heart rate) Similar wording ("really" and "not really" or "a lot" and "not a lot"), was also used to differentiate between high and moderate intensity among young adults (McKenna, Page & Foster, 2004) These findings highlight the importance of communicating with the target population, as the difference

between how people understand and encode moderate and high intensity physical activity, is distinguished by the addition or omission of only one word

Differences between moderate and high intensity also extended to the degree of conscious engagement in undertaking the activity High intensity activity involves a more deliberate intent to perform a task or activity, i e , there is usually a reason for undertaking this level of intensity (e g , running to catch a bus, playing a match) This distinguishing factor has the advantage of a more elaborate initial encoding process and greater rehearsal after the event, which contributes to a better retrieval route (Tourangeau et al , 2000) Distinctive events were more easily recalled for Linton (1982) and Wagenaar (1986) in their diary studies of autobiographical memory These were usually events that had occurred infrequently or were unusual In terms of PA, moderate intensity is perhaps less distinctive than participation in high intensity activity because of its comparatively less conscious intent and more frequent occurrence Durante and Ainsworth (1996) have identified this as an area for further investigation In this current study, much of moderate intensity activity, other than sport and exercise, was reported as occurring in unstructured settings (e g , walking to/from shops, around the house) Activity patterns that are typically unstructured and independent of participation in sports activities are reflective of many women's lifestyles (Ainsworth, 2000, Ainsworth et al , 2000) The challenge in identifying ways of assisting the respondent to accurately recall activities from this intensity has been further compounded by the acknowledgement that participation in moderate intensity activities can be difficult to isolate This can occur when two moderate intensity activities are performed simultaneously (e g , housework and child-minding) or if it is performed simultaneously with high intensity activity (e g , gym session, soccer training) PARQ has been designed to help the respondent to recall time spent separately in moderate and high intensity, by identifying characteristic movement patterns and individual types of activity (as above) The provision of separate physiological and psychological responses to each activity was intended to act as an additional recall cue

3 5 3 General layout of PARQ

Questionnaires are more likely to obtain accurate information about lifestyle PA when the retrieval conditions offered to the respondent are similar to the original encoding conditions (Tourangeau et al , 2000) and accurate memories are more likely to be reconstructed when the retrieval conditions reflect the structure of autobiographical

memory, where the information is stored (Conway, 1996) Provision of multiple cues, e g , what, when, where, better reflect the conditions in which the original activity took place, than the provision of singular cues (Wagenaar, 1986, Brewer, 1988), so long as they are distinctive cues that can help isolate the activity in memory The layout of PARQ addresses the temporal and thematic pathways that reflect the storage structure and retrieval processes of Conway's autobiographical memory, through the provision of cues relating to what (example of activity), when (personal calendar) and where (type or context of activity) Cues that reflect the emotional impact of the activity, increasing its saliency in memory were also provided, in terms of the physiological and psychological responses to activity The layout and cues are discussed below

3 5 3 1 Definitions and Instructions

The properties of "physical activity" that emerged from this study provided a detailed description of what these participants understood by the term A definition was articulated that incorporated the emergent thoughts and expressions of the target population about "lifestyle physical activity" and was placed on page two of the questionnaire, prior to the first question This stated, "Lifestyle physical activity is any body movement that is done during any part of a person's usual day For example, lifestyle physical activity can occur at home, at work, during sport and exercise, looking after children or while travelling" Investigations of participant understanding of this term are rare (Tudor-Locke et al , 2003) and this study afforded the opportunity to state the context of the questionnaire in the words of the target population Cultural-specific vocabulary has been shown to be an important consideration for obtaining accurate information from a questionnaire (Warnecke et al , 1997)

Terminology used by participants to describe the concepts of "high intensity" and "moderate intensity" were also formulated into a descriptive statement High intensity was described as, "This is an activity at a maximum level It is very tiring and cannot usually be continued for long periods It requires near to 100% effort" Moderate intensity was described as, "This is activity at a steady, constant level You're aware that you are active, but it's not too tiring You are making an effort (50-70%), working against a resistance" These descriptions were placed immediately prior to the respective intensity questions This was in order to contextualise the question for the respondent and to clarify the requirements of each question A table of significant physiological and psychological changes associated

with each intensity was also placed in between each "type of physical activity" question These were to act as recall cues to help the respondent remember related activities for that intensity For example, high intensity included "Breathe very deeply, fast heart rate, sweat a lot, concentrating hard and feeling wrecked" Please see the draft version of PARQ (Appendix five)

3 5 3 2 Questions

The participants' natural decomposition of physical activity into various categories provided an instinctive design structure of questions that specifically probed time spent doing activity in these six categories The salient (physical and psychological) intensity descriptors identified for both high and moderate intensity activities also provided an additional recall cue, to help narrow the focus of the retrieval enquiry in autobiographical memory, to a specific activity The questionnaire was formatted to enquire about each of the six types of previously identified activities, firstly at high intensity and then at moderate intensity, during the previous seven days This decomposition of total PA into smaller categories offers three advantages over a single all-inclusive question in that it (a) clarifies what activity is included in the question, (b) provides an organised framework for retrieving information and responding and (c) reduces the computational burden on the respondent (Blair & Burton, 1991, Means & Loftus, 1991, Tourangeau et al , 2000)

The wording of each question was deliberately phrased to encourage the respondent to be as accurate and effortful as possible in calculating their answers This was achieved by the use of words such as "count minutes", "each", "all questions" and "every box" These words were also placed in bold colour to emphasise their importance In this way, it was expected that the respondent would engage in a deliberate enumerative strategy to calculate individual incidences of activity for each category and avoid a rate based estimation, which has been shown to be a less accurate frequency enumerator (Blair & Burton, 1991, Burton & Blair, 1989) Each question had the same wording and was also only one sentence long This was to avoid taxing the working memory with long, conditional sentences, to reinforce the enumerative strategy and also to motivate the respondent to apply effort in accurately retrieving appropriate memories

3 5 3 3 Calendar feature

This questionnaire asks the participant to recall PA from the previous seven days To enhance accuracy and to reduce cognitive burden, it is important to assist the participant to limit their scope of memory retrieval to the specific timeframe in question Issues of telescoping (including activities that occurred outside of the question reference period) and omission (forgetting of relevant activities) have previously been sources of error in other types of questionnaires (Sudman, 1980, Tourangeau et al 2000) The completion of a personally relevant reference calendar for events that extend across one or many lifetime periods have previously been used in other research (Belli, 1998, Means & Loftus, 1987) This is thought to contextualise the time frame for the respondent and to provide some personally salient recall cues that may help to trigger other information that may lead to the retrieval of a relevant memory

Although it appears that the application of a "personal calendar" (PC) or a "timeline calendar" has been restricted to questionnaires detailing with lifetime topics and also in interview settings (Belli, 1998), it has been included in the design of this questionnaire in an effort to reduce the effects of telescoping and omission It was also hoped that the personal calendar would provide contextual cues to the respondent about their whereabouts and their own activities over the previous seven days The PC provides a means of exploring the thematic and temporal aspects of activities over the previous seven days, which reflects the structure of the retrieval processes in Conway's (1996) model of autobiographical memory The seven-day calendar effect was also applied to the space where answers to each of the questions would be placed By decomposing the week into seven individual days instead of one block unit of time (a week), it was expected to encourage the respondent to go through each individual day and count any relevant incidences The name of each day was written into each individual box

3 6 Implications for Questionnaire Design

In summary, the data from these focus groups has provided an invaluable insight into the understanding of "physical activity" as it applies to the day-to-day lives of the participants of this study It has also enabled access to the everyday jargon and understanding of phrases associated with the operation of physical activity, including its varied physiological and psychological effects In this way, the layout and content of PARQ should appropriately reflect how participation in physical activity impacts on the

daily lives of these 18-55 year old Irish adults. The emergence of many properties from the qualitative data, such as the various types or categories of activities, the specific examples within those categories and the different physiological and psychological effects for each intensity, has provided detailed information about the encoding properties of different aspects of PA. Models of autobiographical memory (ABM), such as Conway (1996), suggest that people will more likely retrieve a specific event if they have cues that are distinctive to the event and that match its original encoding. The added inclusion of the personal calendar layout completes the possibilities of using multiple memory retrieval cues relating to the "when", "where", "with who" and "how it felt" aspects of participation in PA. This should enhance the retrieval of more accurate memories (Wagenaar, 1986).

Provision of these cues with the use of the participants own language to describe the context, encourages a question format which should in some way reflect the original encoding conditions and therefore facilitate an appropriate retrieval strategy. By decomposing the questions by intensity, category of activity and by day, it is hoped that the number of events to be recalled for each question will be low. This should encourage the respondent to access the autobiographical memory at General Event (GE) level, before proceeding to the retrieval of sensory rich and detail specific information at Event Specific Knowledge (ESK) level (Conway, 1996). This is the level at which information about recently performed specific activities, that last minutes and hours, is held and is thus most relevant to this questionnaire. The application of two models of cognitive psychology to the questionnaire design and item generation stage, provides a theoretical basis for the structure and content of the outcome questionnaire.

3.7. Limitations and Future Research

The focus groups in this study enabled an understanding of key physical activity related terms, among the target population for whom the final version of PARQ is intended. This method of investigation provides the researcher with a means of gauging people's interpretations, descriptions and opinions on a specific topic (Barbour & Kitzinger, 1999). There are however, acknowledged limitations to using focus groups, namely that the data analysis procedures can be complex and results may have restricted generalisability (Flick, 2002).

The results of this study showed that type of activity was readily remembered and discussed without probing. Detailed descriptions of intensity of activity were also readily

recalled with varying degrees of probing Other aspects of the FITT principles (i e , duration and frequency) were not specifically enquired about in this study, as the retrieval strategies used to calculate them were a specific objective of the cognitive interviews in study two However, the depth of information about type and intensity obtained in this study suggests that duration and frequency should have also been specifically probed Future research should undertake this exploration as it may provide valuable information about the encoding properties and method of storage of duration and frequency within ABM This information could then be incorporated into the design of a new questionnaire

CHAPTER 4 - PRE-TESTING THE PARQ A COGNITIVE INTERVIEW STUDY

4 1 Introduction

The process of cognitive interviewing aims to structure the pre-testing situation by employing a model of cognitive theory to construct the interview It focuses explicitly on the cognitive processes that participant's use as they go through the questionnaire, attempting to identify sources of response error and to evaluate the quality of information received from the participant (Gerber, 2002) One pre-test method uses Tourangeau's (1984) and colleagues (Tourangeau et al , 2000) model of the question-answer response process to track the thought development, by assessing the participants' answers and observing their behaviours According to this model, there are four "cognitive processes" that the participant may use to answer a question These processes are (i) Encoding (comprehension), (ii) Retrieval, (iii) Judgement and (iv) Response and have been previously described in detail in Section 2 5 2

The term "cognitive interview" (CI) may be slightly misleading, in that it is not solely the participants' cognitive processes that are taken into account during the interview Not everything the participant writes on the questionnaire or states to the interviewer will clearly explain their thought processes In many instances, misreading a question or skipping some important instruction is often done at a subconscious level, i e , away from the participants' awareness Such information is acquired by the interviewer, through careful observation of the participants as they complete the questionnaire (Willis, 1999) The interview setting also provides many non-verbal indications of problem areas that an attentive interviewer may detect These are suggestive of implicit comprehension (encoding) problems that are not always identified through conscious verbalisation of thoughts (Beatty, 2004) Potential problems can be identified by watching the participant and noting things like hesitations, sighing, looking back, noting what parts they attend to and what parts they are inclined to ignore, whether they follow the questionnaire sequentially or whether they tend to skip or jump across questions/sections (Gerber, 2002) These observations were concurrently recorded in a notebook as the interview progressed In this way, cognitive interviewing acts as an effective pre-testing tool providing invaluable and informative cues to the interviewer about the practicalities of completing the questionnaire (Beatty, 2004)

Cognitive interviewing has been noted as one potential method to identify and correct problems with questionnaires. Numerous academic survey centres, government agencies and commercial firms in the U.S.A. have incorporated cognitive interviewing into their usual procedures for questionnaire development (e.g., U.S. National Institute of Health, U.S. National Cancer Institute, National Centre for Health Statistics and the U.S. Government Census Bureau). Cognitive interviewing has also been ingrained into procedures for developing large-scale U.S. federal surveys, e.g., National Health Interview Survey and U.S. National Census (Willis, 1999). Despite its use in high profile government and academic institutes, records of methodologies used and empirical evidence supporting the use of cognitive interviews is relatively scarce (Beatty, 2004). Converse and Presser (1986) similarly report that "...we rarely leave records for each other. How a pre-test was conducted, what investigators learned from it, how they redesigned their questionnaires on the basis of it – these matters are reported only sketchily in research reports, if at all", (p.52). A review of studies that evaluated questionnaire pre-testing interviews found that the criteria for measuring the effectiveness has typically been poorly defined and that the data used to evaluate these studies has neither been systematically collected or analysed (Foddy, 1998).

In the context of physical activity, different cognitive interviewing techniques have been used to (i) explore the recall of PA in young people (McKenna et al., 2004) and (ii) to pre-test the Lifetime Physical Activity Questionnaire (Friedenreich et al., 1998). McKenna and colleagues applied the four CI mnemonics advocated by Allison (1996) (detailed in Chapter Two). These mnemonics were used to elicit information from the participants about how they recall physical activity. This form of CI has traditionally been used by police officers and criminal psychologists to enhance the memories of eye-witnesses (Allison, 1996; Milne et al., 2002). The mnemonics are intended to specifically probe for details of a particular event from memory and do not appear to have been used for the purpose of questionnaires pre-testing (Allison, 1996). Questionnaire development and pre-testing has traditionally used a cognitive model of the question-answer process (Tourangeau, 1984; Tourangeau et al., 2000) to facilitate the exploration of the process involved in understanding a question and retrieving an answer from memory. In their CI's, Friedenreich and colleagues (1998) pre-tested the Lifetime Physical Activity Questionnaire by using Tourangeau's model to assess the cognitive processes used when answering questions. Although this study did not evaluate the effectiveness of the CI technique, results

provided the researchers with information about what the respondent understood by the question and what retrieval strategies were used

4 2 Aim and Objectives of this study

The aim of this study was to qualitatively pre-test the draft version of the new physical activity recall questionnaire, by means of a cognitive interviewing technique (Gerber, 2002, Willis, 2002) The four cognitive processes of the question-answer process model (Tourangeau, 1984, Tourangeau et al , 2000) were assessed during the pre-testing of the questionnaire The specific objectives of the study were to (i) identify which cognitive processes are affected by the problems, ambiguities and misunderstandings present in the draft version of the physical activity recall questionnaire (PARQ), (ii) identify what retrieval processes participants use to calculate the frequency and duration of activities at moderate and high intensity, and (iii) amend problematic areas identified within the questionnaire The outcome of this study was to produce a draft version of the questionnaire that could then be assessed for reliability and validity

4 3 Methods

The sequence of events involved in conducting a CI study are outlined overleaf in Figure 4 1 These include (i) consulting an expert review panel, (ii) creating an interview script, (iii) planning recruitment, (iv) conducting the series of interviews and (v) analysing the data

Face Validity

Prior to assessment by the expert review panel, PARQ was administered to nine friends and family (Male=33%, age range 20-55) This form of casual examination of the questionnaire by members of the public provided a form of face validity for PARQ Each participant competed the questionnaire and offered their opinions on its design and structure Time to complete PARQ ranged from 25-40 minutes Results indicated that there were difficulties in understanding the context of the questionnaire, because of the lengthy instructions given on pages two and three There was also difficulty in seeing the word "minutes" in the answer boxes, while the use of the variety of different colours caused confusion and were considered a form of distraction These problems were noted and alterations made prior to the expert review panel

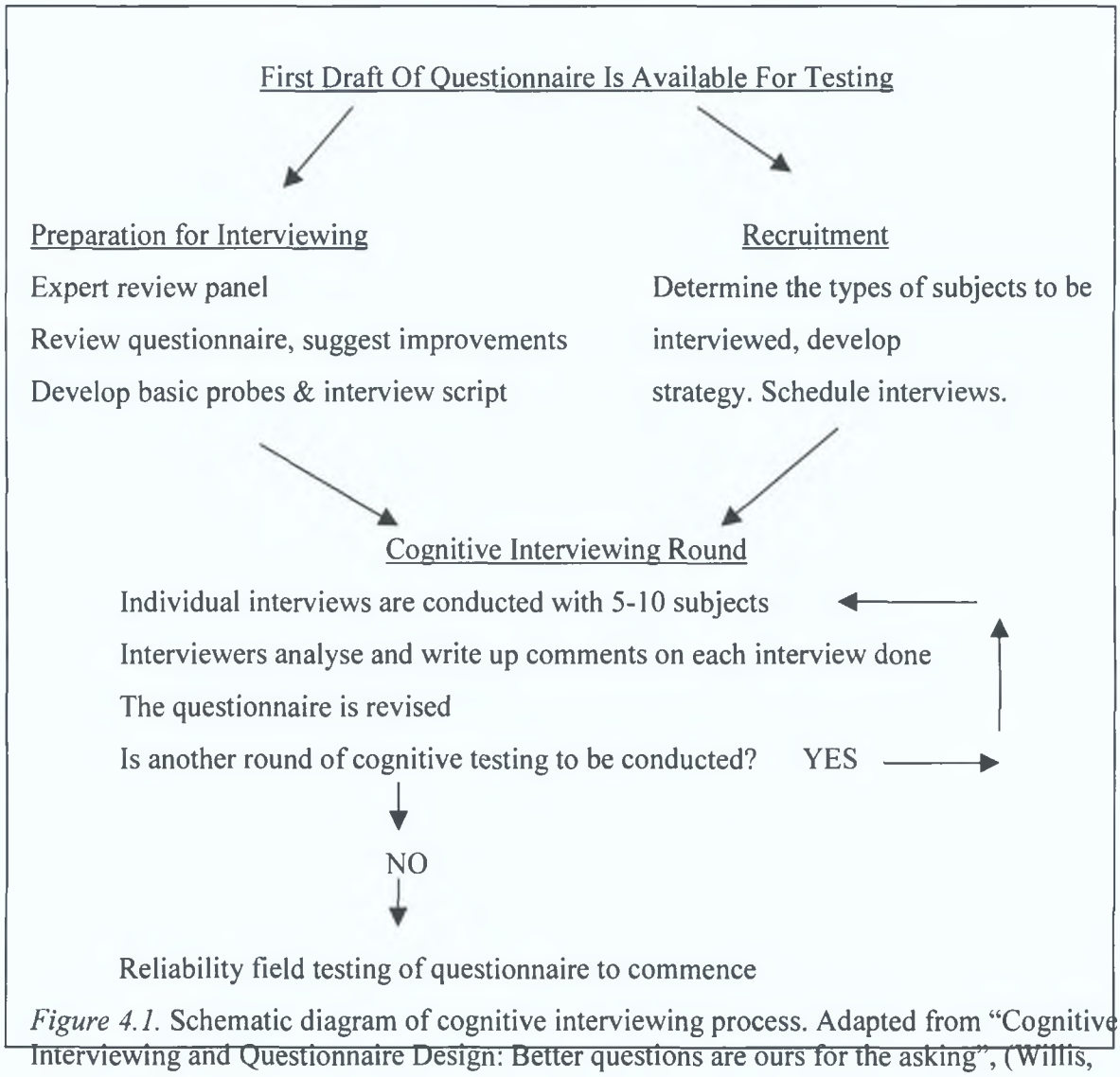

Figure 4.1. Schematic diagram of cognitive interviewing process. Adapted from "Cognitive Interviewing and Questionnaire Design: Better questions are ours for the asking", (Willis, 2002, p. 24).

4.3.1. *Expert Review Panel*

The purpose of establishing an expert review panel in this study was to review the questionnaire prior to testing on the general public, so that the following objectives could be achieved, (i) detect obvious problems, (ii) suggest improvements (iii) modify prior to commencing interviewing, and (iv) create probes for an interview script (Willis, 2002).

4.3.1.1. Recruitment of expert review panel members

Review panel members (N=6, Female=83%) were recruited to represent experience in both the subject area (physical activity) and general questionnaire use. These included two sport and exercise science lecturers and two sport and exercise science postgraduates, who each had knowledge of the physical activity domain and had also used questionnaires

within this field The group also included two psychologists who use a variety of questionnaires within their profession

4 3 1 2 Procedure of expert review panel meeting

Prior to meeting, each member of the review group was given a copy of the questionnaire and asked to fill it out in their own time Having completed the questionnaire, all review members were then asked to consider the questionnaire under specific headings that were given to them by the researcher (Table 4 1) They were also asked to bring these observations and comments to the group meeting for open discussion The review group met for approximately 90 minutes The group was given a brief introduction to Tourangeau's (1984) cognitive model of the question-answer response process, as this was the method that would be used to evaluate participants' responses in the forthcoming cognitive interviews All participants were then invited to freely contribute their impressions and comments about the questionnaire under the various headings outlined in Table 4 1

Table 4 1 *Expert review headings for analysis of the draft PARQ*

Heading		Details
Questionnaire Level	1	The instrument as a whole, rather than particular features of individual questions
	2	Assessments of the conceptual and logical bases of the questionnaire
	3	Issues of formatting, clarity of purpose of questionnaire and instructions etc
	4	Other administrative issues
Section Level	1	Findings that relate to the series or range of questions that are grouped together
	2	Question ordering etc
Individual Question Level	1	Findings that relate to - the phrasing - recall demands - inherent bias - logic - any other features of each question in isolation

Note Adapted from "Cognitive Interviewing and Questionnaire design A Training Manual", (Willis, 1999)

4 3 1 3 Analysis of data from expert review panel

Data was collected via written comments submitted by the reviewers (after they had completed the questionnaire) and note-taking by the researcher of verbal comments made by the reviewers during the meeting Analysis of data was conducted simultaneously with data collection Potential problems were identified by the group and discussed until suggested improvements were agreed upon within the group Participant verification was facilitated by reiterating suggested changes and encouraging the panel to comment Discussion and analysis were structured according to the headings outlined in Table 4 1 above, i e , (i) questionnaire level, (ii) section level and (iii) individual question level

4 3 1 4 Results from the expert review panel

1 Questionnaire Level

The questionnaire was suggested to be initially off-putting because of its bulk The quantity of text and the variety of colours were thought to contribute to a sense of confusion and a lack of clarity about what was required The review group suggested that the use of colours did provide a visual aid in helping to identify different parts of the questionnaire, but that the use of alternating colours for different categories within the same section was a source of confusion

The logical and conceptual basis of the questionnaire was challenged Firstly, the layout of the questionnaire was questioned as it did not assist the respondent to know in advance what the context of the questions within the questionnaire would be It was queried as to whether the current layout would allow respondents to understand the concept of answering seperate questions about the same activities they performed at two different intensities, in a consecutive format It was recommended that this should be briefly explained at the beginning of the questionnaire In this way, the purpose of the questionnaire would become more apparent to the participant

The length of the questionnaire was deemed to be off-putting and may contribute to de-motivating the participant It was suggested that the general layout of the questionnaire could be altered to maximise space A landscape format was proposed in favour of a portrait set-up, for both high and moderate intensity questions It was suggested that this would reduce the bulk of the questionnaire and provide more space to contextualise the questions, thus enhancing the participants' understanding of the task at hand

ii Section Level

The inconsistency in category titles within the two sections (i e , (i) high and (ii) moderate intensity sections) was queried It was felt that a more consistent formatting and labelling of categories should be applied to both intensity sections, as this would reduce any potential ambiguity or misunderstanding For example, the high intensity section did not have a "commuting" category and it was felt that minutes spent "walking to work" may therefore be included in another category such as the "leisure" category However as "commuting" does appear in the ensuing moderate intensity section, the possibility exists for participants to enter duplicate minutes spent "walking to work" in both the "leisure" and "commuting" categories, for this moderate intensity section Equally, the "other" category in the high intensity section that encouraged participants to include minutes spent doing things like "climbing stairs" and "rushing around", was not present as a category option in the moderate intensity section It was recommended that both sections should be formatted in a consistent manner to diminish opportunity for error

The repetition of the yellow boxes, detailing physiological and psychological responses to participation in activity at a specific intensity, was suggested to be needlessly using space as well as unnecessarily adding to the cognitive demands of the participant The rationale for having repeated reminders of how the participant should feel was to help them to focus their answers on activity at a specific level of intensity It was suggested that reducing the frequency of reminders and improving the quality of presentation of the responses in the yellow box, could better serve this utility

iii Individual Question Level

The nature of the information requested by the question was considered to be very burdensome on the participant's memory Two reviewers felt that it was very difficult to be exact in answering the questions, with one suggesting that an average would be more amenable to being user-friendly It was explained that there is a requirement for the information received from the participant to be as accurate as possible, in order to assess whether the participant is performing the minimum amount of physical activity recommended for health purposes

The general instructions for completing the questionnaire and the provision of an example, located on Pages 3 and 4 of the questionnaire, was deemed to be very confusing by three of the reviewers As the participant could not visually see any questions and could

not relate to the demands of the question, it was felt that this depth of explanation was not required at this stage. There was also a general feeling of "needing to do something" at the end of Page 3, having read two pages of text. A suggested amendment was to reduce the amount of text and to rearrange the positioning of the instructions and example to a location that was of contextual significance to the participant, i.e., above the actual question. It was felt by some reviewers that it was unclear from the instructions whether the "past-seven-days" instruction included or excluded "today". It was suggested that a clarification sentence was required to reduce any potential misunderstandings.

4.3.1.5. Discussion of findings from the expert review panel

The expert review appraisal of the draft version of the questionnaire proved to be informative and insightful. It provided an opportunity to test the questionnaire on participants who are informed about the topics of physical activity and/or questionnaire use and design. These reviewers had not been involved with the development of this questionnaire (PARQ) to this point and were therefore able to provide fresh and objective constructive criticism. The review panel provided an insight into potential problem areas that were not overtly apparent to the designer of the questionnaire. The importance and benefit of submitting the draft version to a group of informed experts was reflected in the variety and volume of suggested amendments that were discussed during the course of the review.

The discussion facilitated the achievement of the objectives of the expert review panel. Firstly, the most obvious problems within the questionnaire were detected, as illustrated in the results section above. All levels of PARQ were found to be affected by problems, i.e., at the questionnaire, section and individual question levels. Secondly, having identified these problems, appropriate amendments were suggested and agreed upon by the expert review group. These amendments pertained to issues of layout, wording, conceptual logic and recall demands. For example, to address a problem identified by the group as potentially affecting comprehension and recall, it was suggested that the content of both intensity questions should be consistent and contain the same titles. This suggestion was implemented in order to reduce the opportunity for ambiguity and misunderstanding among respondents. Coherent and uniform vocabulary has been shown to reduce respondent burden and improve comprehension, thus facilitating better quality answers (Warnecke et al., 1997). Another issue identified as a potential comprehension problem was

the inconsistent use of multiple colours to represent different questions It was suggested by the group that the use of so many colours was unnecessary and might distract and confuse the respondent It was recommended that they should either be removed or made uniform, i e , one colour per question This amendment was also made as there did not appear to be any theoretical basis for using colours, therefore the absence of them may assist in reducing the distraction, suggested by the group, and may improve the conceptual logic of the questionnaire An example of a suggestion that was not accepted, as it was contrary to evidence-based opinion, was the recommendation to reduce the burden on the respondents memory by phrasing the question to encourage the respondent to provide "usual" or "average" estimates, rather than having to think back on every day over the past week As the purpose of developing PARQ is to produce an instrument that accurately reflects actual participation rates, that can be compared against recommended minimum participation guidelines for health benefits, this suggestion was rejected Encouraging respondents to think in terms of "average", results in retrieval strategies that are more likely to be estimated, producing information that is inaccurate (Burton et al , 1991, Tourangeau et al , 2000)

The third objective of the expert review panel was to modify the draft questionnaire prior to commencing the first round of cognitive interviews This was carried out, once the suggested revisions were agreed upon by the review group (Appendix seven) Finally, the problems identified and suggested amendments made by the review panel, provided an insight into which types of errors and problems associated with the model of the question-answer process (Tourangeau, 1984) were evident This enabled the construction of an interview script for use in the subsequent cognitive interview study

4 3 2 *Interview script*

The advantages of an interview script are (i) that it provides the interviewer with a plan, (ii) permits consistency across interviews and (iii) allows the researcher to identify in advance, which types of probes are best suited to evaluate the cognitive demands of the questionnaire (Beatty, 2004, Willis, 2004) A variety of verbal probing techniques were employed in this study to ensure that all aspects of Tourangeau's (1984) question-answering process model were assessed A brief description of each of the different types of probes is given overleaf in Table 4 2 The interview script should include a combined verbal probing and observation strategy, due to the self-navigation of the participant

through the questionnaire (Gerber, 2002) Observation strategies used in this study included (i) watching what the participant had looked at and ignored, (ii) following their navigation through the questionnaire, and (iii) observing if they did what was asked of them A summary of the types of verbal probes used in cognitive interviews is given in Table 2 8 The verbal probes and observations included in the interview script used for this study is included in Appendix six The questions and probes used were designed to prompt the participant to reveal some information about which types of cognitive processes were used to answer the questions While the script provides a structure for the interview, interviewers are also encouraged to be flexible and spontaneously explorative, in order to obtain situation-specific information As successive rounds of interviewing brought improvements to the questionnaire, the content of the interview scripts changed slightly with each round This reflected a reduction in the number of errors observed and problems identified, as the rounds progressed

4 4 Recruitment for cognitive interviews

Volunteers for this study were asked to complete a screening form (Appendix one) which included the following selection criteria a) age, b) gender, c) place of residence (Dublin vs outside Dublin), d) occupation / unemployment status and e) physical activity stage of change (SOC) SOC is a construct of the Transtheoretical Model (TTM), applied to physical activity (Prochaska et al , 1994) Recruitment of potential participants was conducted via e-mail, word of mouth and face-to-face contact at a hospital, university, school staff room, office buildings and a bank Participants were selected for the cognitive interviews by stratified sampling (Thomas et al , 2001) This occurred by firstly dividing the volunteers by one specific criteria and then randomly selectmg them for participation In this study, the dividing criterion was physical activity SOC and participants were divided into group one (stages 1-3) or group two (Stage 4-5) Where possible, 50% of participants for each round of interviews represented each of these two categories This was intended to represent the thoughts of those that are not yet regularly physically active (stages 1-3), in contrast to those that are (stages 4 & 5)

4 5 Researcher training for cognitive interviews

Cognitive interviewing is a specific form of qualitative inquiry, which requires the interviewer to have received specialised training Prior to conducting this study, the

researcher (author) attended a course on cognitive interviewing This was done in order to obtain the knowledge, skills and experience required for conducting cognitive interviews to pre-test PARQ This course was a one-day workshop that formed part of the Questionnaire Development, Evaluation and Testing (QDET) conference, which was held in South Carolina, U S A , in November, 2002 The conference was sponsored by the American Statistical Association, the American Association of for Public Opinion Research, the Council of American Survey Research Organisations and the International Association of Survey Statisticians During this workshop, which was run by Eleanor Gerber of the U S Census Bureau, the researcher gained proficiency in writing interview scripts, conducting pilot interviews and applying the cognitive model of the response process to the analysis of the data

4 6 Procedure for cognitive interviews

All cognitive interviews were conducted at locations convenient to the participants Before commencing, each participant was asked to read and sign an informed consent, which detailed the purpose and intended outline of the interview, notified the participant of the presence of audio recording equipment and also allowed the opportunity to ask any questions (Appendix four) All participants were given €15 remuneration for travel and each interview lasted approximately 40-60 minutes The same interviewer (author) conducted all of the interviews and each interview followed the same format, i e , the participant was asked to complete the questionnaire as though they were at home, while the interviewer observed them and interrupted occasionally with questions from the scripted protocol

The questionnaire and interview script were modified after each round of interviews The system of conducting "rounds" allows for amendments to questions identified as problematic by the preceding group The improved questionnaire is then tested on the following group Willis (2002) has identified that " cognitive interviewing appears to function best with multiple small rounds of interviewing, as opposed to a much bigger single round, i e , three rounds of six rather than one round of 18" (p 25) The first draft of the questionnaire was administered to six participants Amendments were then made to the questionnaire following analysis of interview notes and the second draft was administered to a further six participants A final round of five participants were administered the third draft of the questionnaire The number of interviews in each round was decided by the

severity and number of problems that emerged The benefit to be achieved in pursuing interviews with a questionnaire that has obvious problems is minimal, particularly in the early stages of CI and at this point 4-6 interviews may be sufficient for a round (Gerber, 2002, Willis, 2002) In this study, data saturation appeared after 17 interviews At this point, no apparent significant misunderstandings of terms or mappropriate retrieval or judgement strategies were obvious Willis (2002) suggests that the process of CI is such that theoretically, one could interview on an infinite basis and still find minor errors because of individual differences His experience as a cognitive interviewer and survey methodologist for the National Cancer Institute in the U S A has suggested that, " it is seldom necessary to conduct more than 12-15 interviews", (p 39)

4 7 Data Analysis

Analysis of the data from these interviews involved a combined assessment of the non-verbal expressions observed by the interviewer as well as the concerns and comments made directly by participants during the interviews Analysis of the output from this study was structured according to guidelines proffered by Willis (2002, 1999) and at a short course on "Cognitive Interviewing" by Gerber (2002), which the researcher (author) attended These guidelines are

(i) Following each interview, observations made by the researcher and comments made by the participant, were summarised on a question-by-question basis The problem(s) identified within each question was classified into the stage (or stages) of the cognitive model of the question-answering process that was affected (Tourangeau, 1984, Tourangeau et al , 2000)

(ii) Comments and problems identified were accumulated over a series of interviews (usually 5-6), to provide a more complete picture of suggested problematic areas

(iii) Having identified which stage of the question-answering process was affected, the question structure or layout was modified Guidelines for amending problem questions that relate to specific stages of the question-answer process model, were provided by Gerber (2002) and Willis (2002, 1999) and included rewording phrases, removing words or relocating instructions

(iv) A new round of interviewing was commenced with the amended questionnaire, following the same format of analysis

(v) Rounds of interviewing ceased when it was deemed that no further significant or recurrent problems were emerging and that further interviewing would not reveal any new discoveries, i e , when data saturation had occurred

4 8 Validity issues

The validity of the cognitive interview process was facilitated by systematic planning of the interviews This provided a visible, structured and accountable approach to undertaking the study Records of expert review panel findings, sample recruitment, interview scripts, questionnaire developments and analysis procedures were maintained to provide a transparent, thorough and rigorous method of data collection and analysis Validity of the data was considered by purposely including participant verification At the end of each interview, the interviewer provided the participants with the opportunity to clarify any problems they had encountered while completing the questionnaire At the end of each interview, the interviewer held a five minute debriefing interview with the participant, to again voice their overall thoughts and opinions about the questionnaire, particularly those that they did not voice during the interview Evidence of data validity was also provided in the reduced number of errors and problems reported as the rounds of interviewing progressed

4 9 Results

The participants were divided into groups in order to conduct separate "rounds" of interviewing (N=17, mean age = 34 82, S D =+/- 9 32 years, Female=70 6%) Table 4 2 below shows the demographic characteristics of the 17 participants

Table 4 2 Demographic details of cognitive interview participants

Round	N	% Female	% SOC (stage 4 & 5)	Age Range (yrs)	% Employed
One	6	66	75	29-52	100
Two	6	66	50	22-55	83
Three	5	80	40	27-31	100

Results from this study are presented in a question-by-question breakdown The layout reflects the method of data analysis described above, i e , (i) identification of the cognitive process affected, (ii) identification of the associated questionnaire problem, and (iii) the suggested amendment to the questionnaire A detailed account of the results for

round one are presented below The number and gravity of problems diminished as each series of rounds progressed Therefore, only a summary of results from rounds two and three are presented in this section, with a full description of problems and amendments for these rounds located in Appendix eight

4 9 1 *Round One*

Page One (Front Cover)

Problems associated with the Encoding process

(i) When first given the questionnaire, participants were asked to complete it as if they were alone or at home Observation of their immediate behaviour revealed that one participant totally ignored the instructions and information on page one, two appeared to "speed-read" it while the other three participants seemed to approach the first page more attentively Reasons given for the lack of time invested in reading page one properly, indicated that there may be too much information to take in as well as a desire of "just wanting to get started", on behalf of the participant [Female, age 29, SOC 5] For those that did take the time to read it, no obvious encoding problems emerged

Amendments to addressing Encoding issue

Too much detailed information on page one placed unnecessary cognitive demands on the participants' working memory The amount of information was reduced and decomposed into smaller sections to reduce the load on working memory

Page Two (Personal Calendar)

Problems associated with the Encoding process

(i) Problems in understanding exactly what was required for filling out the personal calendar were evident for some participants This was emphasised by hesitations when following the instructions in the box, as well as repeatedly going back and forward between the personal calendar boxes and the instruction box, as if rechecking what was to be filled in Two participants in particular appeared visibly frustrated with completing this task, with one eventually verbalising her concerns, "I don't know what you want here, is it like anything I did or just anything I did that was exercise?" [Female, age 35, SOC5] This indicates an *encoding* problem of misunderstanding of what is required of the participant

(ii) Participants were probed to find out which seven days they had included in order to see whether they had mistakenly understood the instructions as to include "today"

This indicated a variation in the *encoding* processes of all participants For some, it was very clear that they should not include today, " because it says place a circle around yesterday and **starting**[1] with that day the **previous** seven days" [Female, age 34, SOC 3] For others, the phrasing of the instruction was the source of their confusion For two of the participants, this was attributed to the fact that the word "day" appeared three times within the same sentence This was deemed to be too difficult to read and to workout exactly which day the participant was expected to include This sentence was also a source of misunderstanding for one participant in that she was unsure how far to go back, " as yesterday was Wednesday, do I need to go back to the previous Thursday?" [Female, age 35, SOC 5]

Amendments to addressing Encoding issue

There are too many instructions and the wording is too complex and too verbose The colour of the box is off putting (because there are too many colours on the page already) Therefore, the box of instructions and colours were removed and one clear instruction was placed above the personal calendar, stating, "Please write **one thing** in each box that stands out in your memory about each of the past 7 days Start with yesterday "

Problems associated with the Recall process

(i) Participants took varying amounts of time to complete the personal calendar This may be a result of the encoding problems outlined above, but may also indicate the degree of difficulty in retrieving such information This is a *recall issue* As the instructions did not give specific guidance to the participants about which direction to recall the information (i e , a forward, backward or free recall direction), each person was observed to see what their own recall strategy was They were then probed as to why they answered in that direction All participants but one completed the personal calendar using a backward recall method (i e working methodically back through the past 7 days, starting with yesterday) Only one of the participants used a free-recall method (jumping between days until all days are completed) Neither recall strategy appeared to be quicker than the other at completmg the personal calendar

[1] Bold indicates emphasis placed on these words by participant

Those that used backward recall stated that they used this method because for some, it was what they thought they had been asked to do, " that's the way I was thinking back it's just the way I thought I was instructed and so I thought I'd follow the commands", [Male, age 34, SOC 5] For others that used backward recall, they suggested that they did it this way because " it felt most natural" [Female, age 29, SOC 5} and because " I'm a systematic thinker, so I would probably have gone that way anyway I felt that way avoided confusion" [Male, age 40, SOC 2] The free-recall method was engaged in by one participant because "some things were easier to remember further back than the things closer yesterday was fine and Tuesday was fine because I did something a little out of the ordinary on both of those days Saturday and Sunday, I tend to have a very structured weekend, so it can be hard for me to remember I found Friday difficult but then last Thursday something remarkable happened which helped me to remember Friday", [Female, age 34, SOC 3]

Amendment to addressing Recall issues

As each participant appears to naturally select the most efficient direction of recall for them, no specific amendment appeared to be required

Problems associated with the Judgement process

The problems raised on this page indicated a level of misunderstanding However, it also indicated that having recalled particular memories, participant's weren't sure whether they'd retrieved the right information, i e , did the memory match the question? A problem was identified as participants seemed to hesitate and re-read the examples given in the boxes This suggested that participants were having problems deciding whether the information they had retrieved was relevant to the question asked One participant stated that she wasn't sure what type of memories were to be filled in When asked to read the instructions out loud, she indicated that she now understood it was for general memories " oh yeah! I get it now", [Female, age 34, SOC 3] This is a *judgement issue*

Amendment to addressing Judgement issues

The two participant's that had difficulty with what the question asked of them, were asked to read the instructions out loud and verbalise their thought process in deciding what to do For both participants, this revealed that the problem occurred only because they had

not read the instructions fully the first time, and not because of any misunderstanding in the reading of the question. Therefore, no amendment was required.

Page Three (Definitions)

Encoding

All participants appeared to read through all of page three without undue difficulty. There was no obvious pausing or re-reading of sentences. Participants were asked to paraphrase the difference between moderate and high intensities. There were no apparent misunderstandings noticed, although one participant did state that she wasn't sure she'd agree with all of the associated feelings for moderate intensity.

Problems associated with the Judgement process:

(i) Two of the six participants verbalised their uncertainty about not being sure whether they should "do something" at the end of the page. This is a *judgement* issue. Having spent some minutes reading about how they should answer the coming questions and then being provided an example of a completed question, the participants felt that they were expected to write something in somewhere before turning the page. The location of the example box on this page, made participants question whether they should write something.

Amendment to addressing Judgement issues:

The example was removed and placed on the next page, at the beginning of the questions. This placed the instructions in the context of the question being asked. The detailed descriptions of moderate and high intensities were removed as these were also present in the next two pages on the questionnaire. This was intended to reduce the cognitive burden on the participant by reducing the amount that was required to read. Additionally, it was anticipated to enhance understanding and to reduce the uncertainty of whether to "do something".

Pages Four and Five (High Intensity Question)

Problems associated with the Encoding process:

Two participants did not fill in zero minutes into boxes where they had done no activity (contrary to instructions given). These two participants were asked to explain why

they'd left those blank. Both participants stated that it was because those questions didn't apply to them. "I didn't do any housework", [Male; age 40; SOC 2] and, "I didn't do any sport this week", [Female; age 34; SOC 3].This is an *encoding issue* as they clearly did not understand the instruction. However, concurrent probing as the participant was completing that part of the questionnaire revealed that both participants had not seen the instruction on the previous page requesting them to fill in zero minutes. This now becomes a *layout* issue.

A further layout issue was identified when two participants went back to a previous category and erased the answers they'd given. This was because they realised that they had given answers that were more appropriate for a category later on. In general, many of the participants sighed a lot and rechecked the instructions on the previous page on numerous occasions. The poor layout design affected the encoding process as it prevented the participants from fully understanding what was expected of them.

Participants had difficulties interpreting the meaning of some phrases. One participant [Male; age 34; SOC 5] who works as a cleaner did not differentiate between the "Housework and gardening" and "Occupational" categories. This caused him to have to erase the minutes he had given for high intensity cleaning in the housework category because he realised later on that they were more applicable to the description of occupational physical activity. He states, " I see that type of work as what I do on the job. Maybe you should write "in the home" instead? This is an *encoding issue* as it indicates a misinterpretation of the word.

Participants indicated discrepancies between their own understanding of various category titles and the intended understanding (depending on which category had been the subject of the probe). For different categories, participants had been asked, "What does the word X (e.g. "Leisure") mean to you in this context?". In this study, the word leisure was understood to mean another form of "...structured exercise activities like gym and aerobics while sport means team activities like football" [Female; age 34; SOC 3]. Another participant suggests that it is "...something you do for enjoyment, away from work" [Male; age 40; SOC 2]. Both of these explanations are quite diverse and indicate varied interpretations of the same word. This is an *encoding issue*.

The phrase "physical activity" at the end of each category name caused confusion for two participants [Male; age 34; SOC 5; Female; age 34; SOC 3]. For example, in the category title "Occupational physical activity", both of these participants understood it to mean "Occupational activity" AND "physical activity". "Physical activity" was interpreted

to mean "exercise". This led to inaccurate answers because, for "occupational physical activity", the participants were then including minutes spent doing exercise in addition to those spent doing occupational-related activity. Four participants restricted their answers to minutes spent doing only the specific examples given within each category. This indicated an *encoding problem* and a discrepancy in the meaning of the word "example". Its intended meaning of "something along the lines of…", is in contrast to "only these activities", as understood by the participants.

Amendments to addressing Encoding issues:

The "please write zero minutes" instruction was moved to the top of page four (from the previous page) and written in red bold for emphasis. This placed the instruction in context with the question and was intended to reduce encoding (comprehension) errors.

The layout of the questionnaire was changed to a landscape format. Category titles were now listed underneath each other (instead of across two pages). Participants could see all of the categories to be answered at a glance and could see what questions were to come.

As the above amendments were expected to improve comprehension, it was decided not to alter the wording of the category titles. Instead it was decided to wait and see whether the comprehension of participants in the next round were also adversely affected.

Problems associated with the Recall process:

Two participants [Male; age 34; SOC 5; Female; age 34; SOC 3] paused while completing the questionnaire in the high intensity section because they began to realise that the next section was going to request similar information for moderate intensity activity. Both participants felt that they may have added those minutes into the high intensity section. While this represents an *encoding issue* (i.e., the misinterpretation of the question, as referred to above), it also suggests an unsuitable and inefficient *recall strategy*. One of the participants (the female) had verbally referred to using the recall cues that denoted physiological and psychological aspects of high intensity when calculating her answers, yet did not limit her answers to high intensity activities.

Amendment to address Recall issues:

In order to encourage the appropriate recall of information for each question, the following instruction was placed on page three, i.e., before the participant begins answering

any questions, "The first questions will ask about activities you did at a high intensity. Then you will be asked about the same activities at a moderate intensity". The purpose was to provide further context to the questions.

Problems associated with the Judgement process:

Having recalled information about minutes spent in high intensity, participants showed a degree of hesitancy about whether they had recalled accurate data or spent enough time/effort in arriving at an answer. One participant suggested that as he was going through the questions, "...you realise you haven't done any of it [activity] and you feel like moving on to the next one [question]", [Male; age 40; SOC 2]. This is a *motivational issue* associated with *judgement,* in that it affects the answer if the participant decides at this stage that he does not wish to engage in accurate, effortful answering. This may be for reasons of self-interest if recalling the correct information is perceived as taking too much time.

Two participants in particular [Male; age 34; SOC 5; Female; age 34; SOC 3], highlighted specific reporting processes involved in giving an answer to the question. Having checked back to see what the instructions were given on the previous page and looked forward to see if the next question was about moderate intensity, both decided they had answered incorrectly and needed to edit their responses. The intimation of "just wanting to move on to the next one" [Male; age 40; SOC 2] was because he realised how little activity he does. This participant wanted get the questionnaire finished quickly, possibly because he perceived that the answer was not socially desirable. This indication of poor motivation to provide accurate answers is suggestive of a failure at the *judgement* process stage.

Amendments to address Judgement issues:

A short "motivating" question was placed at the end of each intensity question. This allows the participant to report whether they usually do more activities at this intensity. This is in the form of a tick box (yes or no) format and is solely for the purpose of motivating the participant to provide accurate answers by allowing them to justify any unusually low answers they may have given.

Layout issue:

Two participants did not see the "What activity did you do most of?" question until they had completed three categories in the high intensity section. They then went back and filled in the previous questions. This is a *layout* issue.

Amendment to address layout issue:

Put an extra space between the answer grid box and the "what activity did you do most of" question. Also place the "Q" (denoting "question") in red bold for emphasis. This should enhance the visibility of the question.

Pages Six and Seven (Moderate Intensity Question)

Problems associated with the Encoding process:

One participant [Male; age 34; SOC 5] re-emphasised his misunderstanding of the context of the questions to date. He stated that he was unsure from the outset what was expected of him. It was only when he began the moderate intensity question that the purpose of the questionnaire and the context of the previous question became obvious to him, "Oh I see now. Its how long I spent doing these…how many minutes. You see I didn't feel like the other description last week, heavy breathing and sweating, so I didn't know what to write". This is an *encoding issue*.

Two participants were asked to paraphrase the sentence, " only count activities that made you feel like most of these" (followed by the list of physiological and psychological responses). Both gave a good understanding of the intended meaning "…that means don't count anything that didn't make you feel like this and you don't have to have them all, you could afford to leave out one or two of them, so long as you had the rest of them", [Female; age 35; SOC 5]. No other *encoding issues* were apparent.

Amendment to address Encoding issues:

Categories were changed to a landscape format (similar to the high intensity questions) so that the participant could see the context of the whole question at a glance. The examples within each category were placed on a single line (rather than over two lines). This was intended to enhance clarity and give the participant maximum assistance in understanding the question.

Problems associated with the Recall process:

In general, participants tended to recall in a forward direction starting on Monday or else in a free recall movement. The personal calendar was noticeably used, one participant reporting that he used it to "help place himself during the week", (Male, age 34; SOC 5). The participants appeared to take longer to arrive at answers for this question compared to the high intensity question, perhaps due to the increased amount of information to process. This was evident from the increased amount of time reported in the boxes for this intensity in comparison to high intensity. This is a *recall issue*.

Participants were asked to verbalise how they were arriving at their "minutes per day" answers, for various categories. This gave an indication of which recall strategy they were using. Recall strategies also vary depending upon the specific activity being recalled, i.e., as frequency increased, participants were less likely to engage in effortful enumerative strategies. Comparative strategies were used by one female for different activities she does [age 35; SOC 5]. For housework, she stated the day "Tuesday" and then began to individually list the things that she was doing in the house on Tuesday when she felt like the physical and psychological descriptors in the yellow box. Then she counted how many minutes she spent doing each of those activities, giving a final figure of 15 minutes. She continued to go through the week methodically, using her personal calendar to help her remember what she was doing each day.

However for the stair climbing activity at work, the same participant stated that she would find it "...hard to guess how many minutes I'd be going up and down the stairs at work...because I'd be at my desk and getting up and down and I might go up and down the stairs a few times". Despite the obvious difficulties that she expressed in trying to remember, she continued to try and remember as best she can "...that'd be hard to try and add up because I'd have no way of recalling how many times I'd walk up and down the stairs at work...some days I might just walk up and down four times and then I might be up and down constantly, depending...". Although she tried to enumerate incidences of stair climbing, eventually she decided the best way to calculate an answer was to "...try and figure out if I walked up the full staircase which I would usually, then how long that would take me and then try and make a guess at how many times I would have done that in a day. But it would be just a guess really".

Despite the difficulties encountered in arriving at an answer, this participant appeared to have invested a lot of effort in trying to remember exactly what she had done.

When asked how confident she would be in the accuracy of the answers she had given, she states that she would be very confident (rating 8/10).

This description indicated a difference between the strategies used for calculating the frequency of the event, in comparison to calculating the duration of each event. For four of the participants this was obvious from the individual incidences of activities that were mentioned in calculating the frequency, compared to the language used by the participant when calculating the duration. In this instance, it was more obvious that participants estimate the duration of an activity through use of language such as "about" and "usually". One participant [Male; age 34; SOC 5] stated that he has "rounded up or averaged" his answers, while at the same time felt he was not far off the true time spent, offering an acceptable 6/10 in confidence about the accuracy of his answers for this section.

The discrepancy between enumerative precision for frequency and rate-based estimates for duration of activities is highlighted by another participant [Female; age 34; SOC 3], "...Sunday is the day I do housework and I did vacuuming and scrubbed the floors and rearranged the furniture. Saturday I would have carried the shopping and they were heavy. So between [all that] I'd say I'd have spent about 2 ½ hours doing that. It was moderate, I wasn't out of breath when I'd finished, so that's about 150 minutes on the Sunday. Carrying the shopping? That would have been about 15 minutes". Here this participant has used an enumerative strategy to calculate the incidences of activity but has rate based the durations. Theses examples provide an insight into the *retrieval issues* associated with this questionnaire.

Amendments to address Recall issues:

An increased amount of information to process and retrieve was evident in this question. Given the increased cognitive workload and recall demands, the link between the moderate intensity questions and the physiological and psychological recall cues was highlighted. Descriptors within the yellow box needed to be clarified. This was achieved by removing one of the yellow boxes and making the other one bigger, thus emphasising the physiological and psychological descriptions within. The word "slightly" was placed in front of "increased heart rate". This was to help differentiate it further from the high intensity description of heart rate and to make it more consistent with other moderate intensity descriptions. These amendments were intended to assist participants in focussing only on activities of moderate intensity.

Problems associated with the Judgement process:

Participants verbalised, by thinking-aloud, that they were sometimes unsure whether the information they had recalled was relevant to the moderate intensity section. One participant had difficulty deciding whether to include some information she had retrieved, "...going up and down the stairs, it didn't take anything out of me" [Female; age 34; SOC 3]. She also wondered whether this was still relevant to the question asked.

Other *judgement issues* concerned the accuracy of the data recalled and whether they had thought of everything. " I would attend 2 or 3 meetings a day, so you'd be looking at a couple of minutes getting in and out of meetings...so it's very much an estimate" [Male; age 40; SOC 2]. Confidence in accurate answers was generally high for all participants. When asked "how confident are you that your answers are accurate, on a scale of 1 to 10, where ten is extremely confident?". High confidence was indicated by scores of seven and eight out of ten, with only one reporting a six out of ten.

Amendments to address Judgement issues:

A "motivational" question, similar to that identified for the high intensity question, was also added to the end of the moderate intensity question, i.e., asking the participants whether they would usually have engaged in more or less activity in a week. This was done to maintain consistency in the format of both questions and to also motivate the participant to continue answering as accurately as possible.

Page Eight (Walking Question)

Problems associated with the Encoding process:

All participants showed signs of a poor *encoding* process for this question. All appeared to hesitate for long periods while reading. They also re-read the instructions a couple of times. Some verbalised their confusion over the list of categories while others were confused with the list of affective descriptors in the yellow box. For example, one participant [Male; age 40; SOC 2] was asked to paraphrase the "walking" question. He indicated an excellent understanding of what type of information should be retrieved, "...looking at walking that is putting a stress on you, so it's not just going out for a stroll. [It's] walking when you're seriously walking, walking when you know you're doing it, so you've got a purpose". Although he understood the type of walking to be included, he did not understand the relevance of the list of "categories of activity" in the question.

Other *encoding issues* concerned the content of the descriptors in the yellow box. These were a source of uncertainty for many participants and were said to be confusing and misleading. It was also suggested that there was too much information to take in, i.e., they did not understand what the instructions meant, when they were asked to count all minutes they had reported walking, in all types of activity, for both intensities.

Amendments to address Encoding issues:

The problems identified occurred because there was too much information to take in and read at once and because the participants missed the instruction to, "only count walking already included in high and moderate intensity sections". The question was decomposed by placing it on its own, away from the instructions. This was intended to reduce the burden on working memory. The instruction "only count minutes already included" was underlined, to encourage the participant to refer back to previous answers. A sentence to explain the purpose of the question was placed at the beginning of the question.

Problems associated with Recall process:

Because of the encoding problems identified above, the *perceived recall* demands of the question among the participants was high. However, having spent time re-reading instructions and working through the question, participants appeared to adopt an appropriate strategy by referring back to previous answers for cues. Until the participant realised what information the question requires however, all participants appeared to be quite frustrated and unsure about how to recall the information.

Problems associated with Judgement process:

Many participants appeared very unmotivated and frustrated when answering this question. This may have affected their decision to invest sufficient effort to answer the question correctly. Words such as "unsure" and "confusing" used by all participants infer a lack of confidence in the validity of their answers. This is a *judgement issue* related to motivation.

Amendments to address Recall and Judgement issues:

The yellow box of descriptions were removed completely as participants should be familiar with them, having completed the previous two questions. This was done to reduce

the volume of text on the page. It was expected that this amendment would improve clarity, improve motivation and also reduce the burden on the participants working memory.

4.9.2. *Round Two*

and

4.9.3. *Round Three*

Analysis of the data from these rounds was undertaken exactly as described above for Round One. To avoid repetition of detailing the exact same procedure, a summary of the results from Rounds Two and Three are presented below in Tables 4.3. and 4.4., while a more detailed representation of the results of these rounds is included in Appendix eight. The summary Tables below identify (i) the cognitive process affected, (ii) the questionnaire problem and (iii) the amendment made, for each round of interviews. As each round progressed, the number and gravity of problems identified diminished. A discussion of the results of each round and the effect of the cognitive interview process as a whole is also presented below.

Table 4.3. *Summary of the cognitive processes affected, problems identified and amendments made during round two*

Page No.	Cognitive Process	Problem Identified	Amendment
1	None	None	None required
2 & 3 (Personal Calendar)	Encoding	(i) One P did not write in his own examples. Did not understand that this was required.	(i) First time (out of 12 interviews) this has occurred. No amendment required at this stage.
		(ii) One P filled in only two days.	(ii) Further probing revealed that P had not read the instructions but did understand. No amendment.
	Recall	(i) Two P's expressed difficulty remembering each of past-seven-days. (ii) One P found difficulty in recalling days early in week.	(i) and (ii) Requires assistance to recall. Instructions decomposed into five sentences and stated alone on page two, to provide step-by-step help for P.
	Judgement	The difficulty expressed by P's above in recalling data affects motivation to provide accurate information.	Encouragement needed. Better instructions (as above) and explanation of purpose of calendar to provide context of importance.
4 & 5 (High Intensity)	Encoding	(i) Two P's didn't understand that words in yellow box were recall cues to help them isolate high intensity activity in memory. One P mistakenly ticked the words in the yellow box he agreed with.	(i) The words "that make you feel like the description above" were added to the end of each question to help the P to remember to use the recall cues provided.
		(ii) One P asked if similar questions would be asked about moderate intensity, despite a previous instruction.	(ii) Further clarification of the context of the questionnaire will be provided on page three.
		(iii) Two P's misunderstood the word "example", taking it to mean " these activities only".	(iii) The words "and other" will be added to the end of the list of examples, to encourage P's to think of their own activities.

Note. P=Participant

Table 4.3. (continued). *Summary of the cognitive processes affected, problems identified and amendments made during round two*

Page No.	Cognitive Process	Problem Identified	Amendment
4 & 5 (contd.)	Encoding	(iv) Two P's had problems understanding category titles, particularly the use of words "physical activity" in every title. These words were understood as "exercise". The term " leisure" was also confusing, explained as "exercise that's not sport", instead of including less structured activities.	(iv) The word "physical" will be removed from the category titles. Title now reads "Occupational activity that made you feel like the description above". The "sport and leisure" category title now reads "sport and free-time", to emphasise inclusion of more leisureable and less structured activities.
	Layout	Two P's missed the "what activity did you do most of?" question.	A space was placed between each question and a different colour was used to highlight this question.
6 & 7 (Moderate Intensity)	Encoding	(i) The word "example" is misunderstood by two P's not sure where to include "recreational walking". They did not put it in "sport and leisure" as it was not listed as an example.	(i) The words "and other" were added, similar to the amendment outlined for the high intensity question.
		(ii) One P stated that she did not use the recall cues in the yellow box because she didn't understand what they were for.	(ii) The yellow box was put into landscape format. This matched the layout of the questions and made the box bigger and the words clearer. This was also done for the high intensity question.
	Layout	Two P's missed the "what activity did you do most of?" question.	Amended similar to the high intensity section, i.e., with the use of space between questions and colours.
	Judgement	One P stated that she wasn't sure whether to write zero minutes where she had done no activity. This was even though she had read and understood the instructions.	No amendment as no other P has voiced this issue.

Note. P=Participant

Table 4.3. (continued). *Summary of the cognitive processes affected, problems identified and amendments made during round two*

Page No.	Cognitive Process	Problem Identified	Amendment
8 (Walking Question)	Encoding	(i) Two P's did not understand that answers should only include moderate and high intensity walking. (ii) One P stated that he was confused with the layout and listing of categories provided. Not sure what he was asked to do.	(i) and (ii) The volume of text was reduced and the question was decomposed into smaller sections. A specific question about level of intensity was added to remove ambiguity about what level of activity was to be included and the list of categories was removed completely.

Note. P=Participant

Table 4.4. Summary of the cognitive processes affected, problems identified and amendments made during round three

Page No.	Cognitive Process	Problem Identified	Amendment
1	None	None	None
2 (Personal Calendar)	Judgement	Two P's took extended time to decide what type of memory to include in the personal calendar. Finally decided to include general information and not just something related to physical activity.	No amendment required.
3 (Definitions)	None	None	None
4 & 5 (High Intensity)	Encoding	(i) One P was not sure whether to put zero in boxes where she did no activity. (ii) One P did not know whether to put minutes walking to work in occupational or commuting	(i) With further probing, P stated she had not read the instructions about placing zero in the box. No amendment required. (ii) P stated that if she had read through the titles of the categories first she would have known that this walking was more appropriate to commuting.

Note. P=Participant

Table 4.4. (continued). *Summary of the cognitive processes affected, problems identified and amendments made during round three*

Page No.	Cognitive Process	Problem Identified	Amendment
4 & 5 (contd)	Encoding	(iii) Two P's answered the "what activity did you do most of?" when they had not recorded any minutes of activity for that category. This indicated a misunderstanding of this question.	As this had not occurred previously, no amendments were undertaken in order to avoid introducing the risk of unnecessary ambiguities to a question that up till now had not been problematic.
6 & 7 (Moderate Intensity)	None	None	None
8 (Walking Question)	Encoding	(i) Four P's had difficulty understanding the question and did not know what information was being sought. (ii) One P was confused by the use of the word "you" and did not understand the purpose of the question.	The number and variety of problems associated with this question in all three rounds suggest that it is not a question that is ready for use within this questionnaire. As it is the only question consistently causing problems, it has been removed.

Note. P=Participant

4.10. Discussion

Over the course of three rounds of interviews, the Physical Activity Recall Questionnaire (PARQ) was profoundly modified. Each round of interviews uncovered various problems and misunderstandings that were identified by the observation and probing of respondents as they completed the questionnaire. Each incongruity noted was categorised according to Tourangeau (1984) and colleague's (Tourangeau et al., 2000) cognitive model of the question-answer response process and amendments were made in an attempt to correct the flawed feature. Findings are discussed in terms of how the CI method improved the overall functionality of the PARQ, with reference to the individual stages of the question-answer process.

4.10.1. *Stages of the question-answer process (Tourangeau, 1984; Tourangeau et al., 2000)*

Initially, the CI process revealed that the encoding and judgement processes were heavily affected. These processes represent two important parts of answering a question, in that they deal with the initial understanding of what information is being sought (encoding) and whether the memory constructed is appropriate to the question asked (judgement). Problems associated with the method of information retrieval were not as obvious, except for instances of calculating the duration of events. Examination of the problems associated with each of these processes, and their implications are discussed in more detail below.

(i) Encoding (Comprehension)

The type and regularity of encoding problems that emerged in this current study were consistent across interview rounds one and two (although they had slightly reduced in round two). In general, these problems related to the layout of specific instructions and the position of these instructions relative to the question being asked. The amount of text that the participant was expected to read and take in before beginning the questionnaire appeared to place heavy cognitive demands upon the working memory of the participant. The lack of clarity about the context of the questionnaire, and the questions within it, also contributed to the encoding problems experienced by the participants. Problems associated with the meaning of words and sentence structure were also observed by Conrad and Blair (2004). In their study, comprehension type problems occurred more often, compared to questions relating to calculations or timing of

events. In PARQ, encoding problems appeared to occur more frequently for encoding than for any of the other cognitive processes, particularly in round one. Evidence of misunderstanding dissipated by round three, indicating that the amendments made to the questions and layout had improved the functionality of PARQ.

(ii) Retrieval

Issues of appropriate retrieval strategies were encountered throughout all three rounds of interviews. For questions that required calculations (i.e., frequency and duration of activities), participants would ideally enumerate individual incidences of performing activity and then calculate specifically how long they had engaged in that activity. This would provide the most accurate information but would unfortunately not be economical in terms of the amount of effort expended (Tourangeau et al., 2000; Blair et al., 1987; Burton et al., 1991). In this current study, participants did articulate their difficulties in remembering the duration of some frequently performed activities in terms of minutes, e.g., time spent climbing stairs. The recall of the total duration of an activity like stair climbing is difficult because of its discontinuous nature, i.e., it consists of numerous interrupted incidences of activity that occur in fragments over the day. One retrieval strategy may be to recall partial descriptions of events and use other recall cues (e.g., places, people) to aid the parallel retrieval of additional information from memory. This method of retrieval coincides with the description of access to information from the autobiographical memory (ABM) as outlined by Conway (1996). Conway's model advocates a thematic parallel cueing system that assists in retrieving memories. For example, thinking of a shopping trip reminds the participant that they used the stairs and not the lift because it was broken.

Sequential cueing is also a feature of Conway's (1996) model of ABM structure. In this type of cueing, duration may be calculated by recalling another event that corresponds to the beginning and the end of the activity, e.g., the end-time of a TV programme and the time the participant went to bed. The time difference between these two events would provide an estimate of the duration of an activity. McKenna, Foster and Page (2004), found that recall of the duration of an activity was refined among young adults, when the period had a recognisable start and end time. In the context of this current study, findings indicate that while this may be a practical retrieval strategy for prolonged and continuous activities, such as a football game or washing floors, it appears impractical for numerous, discrete events such as climbing stairs or carrying

items at work, at various times during the day as they have no recognisable start-end time. The implication of this finding for questionnaire design is that many different types of recall cues need to be incorporated, so that the structure and the pathways of ABM are represented in the design of the questionnaire. The presence of cues relating to when, where, what, may initiate the parallel or sequential retrieval of information that helps the respondent to recall and sum individual and discontinuous events. The presence of multiple cues provides a more accurate presentation of information into memory, than when only one cue is presented (Wagenaar, 1986). Parallel and sequential access to information about frequency and duration was encouraged in PARQ through the provision of multiple cues. These included providing the respondent with information about (i) *what* activity they performed (by decomposing physical activity into different types and providing examples of activities within the context of that type), (ii) *when* they did the activity (by decomposing the week into individual days), (iii) the *affect* that was experienced while doing the activity (by decomposing the activity into different levels of intensity and through the provision of a list of physiological and psychological responses to activity), and (iv) other information such as *who* was there and whether it was preceded by or succeeded some *other personally salient event* (cued by personal information contained within the personal calendar).

(iii) Judgement and Response

Indications from the interviews in this study are that the amount of participant effort and motivation invested is highly variable. The motivation of the respondent appeared to be linked to the incidence of encoding problems and the perceived cognitive workload involved in understanding the context of the questionnaire. As the encoding issues were resolved, judgement issues were noted to also decrease. Problems associated with judgement and response selection were identified as being second highest in occurrence, after encoding (DeMaio et al., 2004). Problems with the judgement process in this current study appeared to be reduced following the introduction of the "Would you usually do more of "X" intensity activity?" question at the end of each intensity question. This appeared to improve morale among participants that may have been de-motivated by the low answers they had written in the high intensity section. The problem of social desirability bias is a significant one in questionnaire research, particularly in relation to perceived threatening questions (Sudman, 1980). Sudman's research indicated that physical

activity is not perceived as a threatening topic, however people will still want to present themselves in as favourable a light as possible. Equally, people are motivated to be "good respondents". Implications from this study highlight the responsibility of the designer to facilitate the respondent's willingness to give the most appropriate information while not placing them in a social desirability dilemma. The question that asked respondents whether this amount of activity was typical for them, permitted an opportunity to justify their answers, particularly if they had reported unusually low participation rates because of illness or injury.

4.10.2. *Effect of the CI process on pre-testing PARQ*

Comments from all 17 participants in this study, but most specifically from the five in the last round, indicate that the process of CI has contributed to producing a questionnaire that is much improved from the original draft (Appendix nine). The overall reduction in encoding and judgement problems is a reflection on the advancement of the layout and the design of the questionnaire. The reduction or absence of problems identified in a questionnaire has previously been used as the method of verifying the positive impact of the CI process in producing an improved questionnaire (Presser et al., 2004; Subar et al., 1995; Willis, 2004; Friedenreich et al., 1998). Confirmation of the improvements made by the CI process was provided in the debriefing interviews of round three. Participants were also asked to rate the user-friendliness of the questionnaire. "User-friendliness" was explained to the participant as ease in following the instructions, visual impact of the questionnaire (colours, fonts used etc) and understanding what was expected. All participants rated it as 10 out of 10 and described it as "very clear", "easy to understand" and "unambiguous". Participants also stated that they were "very happy" with their answers which were "the most accurate I could be for this type of question".

The layout and design of the questionnaire was adapted to enhance the memory retrieval strategies of the respondent, through the provision of different types of recall cues. The provision of cues and memory triggers such as the personal calendar, activity examples, descriptions of physical and psychological changes and the decomposition of activities into categories, appears to have assisted many participants. This seems to be particularly true of the personal calendar which helped the participant to place themselves in the past week, provide landmarks to the participant and thus contextualise the question. The use of landmarks was identified by McKenna and colleagues (2004) as a method of enhancing the recall of PA in young adults. The specific

use of personal landmarks have been suggested as preferential for recalling personal memories by Belli and colleagues (2001) and Shum (2001). The personal calendar provides this facility. Recall calendars have been found to be an effective recall cue for assisting the respondent to retrieve information about when an activity took place and where it occurred (Friedenreich et al., 1998; Means et al., 1991). Friedenreich and colleagues (1998) used a cognitive interviewing protocol and recall calendar to pre-test an interview-led lifetime recall physical activity. They found that the recall calendars enabled respondents who had difficulty remembering past activities, to recall them easily.

4. 11. Limitations and future research

The issue of combining enumerative and estimation based strategies for calculating duration times was highlighted as not theoretically ideal. The recall of duration and frequency information was also found by McKenna and colleagues (2004) to be most problematic among younger children. Despite the provision of multiple cues, it may simply be that estimation-based strategies are inherently the most practical method of providing an answer for duration questions that involve discontinuous events. While this is noted as a possible limitation of this type of recall questionnaire, further research into the cognitive processes of recalling discontinuous events/activities may reveal a cognitively efficient and economical method of recalling accurate information about duration.

Cognitive interviewing provides a format for assessing the functionality of a newly developed questionnaire. However, more research is required to assess both the validity and reliability of the technique. To date, much of the research into the efficacy of pre-testing procedures is qualitative and impressionistic in nature (Foddy, 1998). Evaluations of CI techniques tend to rely upon data reduction methods that are too dependent upon the researcher's subjective judgements. This may affect reproducibility and generalisability to the field (Willis et al., 1997). Finally, more research needs to be conducted to document methods of reliably assessing which types of probes are efficient in identifying problems associated with particular cognitive processes.

The comparatively lower detection of problems associated with the recall process may be a function of two factors. Firstly, the subjective nature of identifying and categorising specific problem areas has been identified as a limitation of CI's and may have resulted in a recall-related

problem being labelled under another process. Secondly, the high incidence of encoding-related problems may have masked the true frequency of problems related to recalling information. For example, if a participant was initially observed to have had, or verbalised that they had, a problem understanding the meaning of a question, this was noted as a comprehension problem. The question may have also been problematic for other processes, but they were not pursued once it was clear that the participant did not understand the question.

In conclusion, cognitive interviewing is a qualitative method of pre-testing a questionnaire, specifically its questions, layout and format. This method provides an ideal opportunity for the researcher to identify problem areas and to make appropriate amendments, based upon the guidelines of a proposed cognitive model of the question-answering process (Tourangeau, 1984; Tourangeau et al., 2000). However, it is not a definitive gold-standard answer to questionnaire pre-testing and issues pertaining to interviewer competency, complex analysis and relevance to wider populations are valid (Beatty, 2002). Although results from this study indicate that the cognitive interview process appears to have been a very effective tool for pre-testing this specific questionnaire, the lack of objective measurement of problems is a limitation of this study.

The value of CI in teasing out problem areas and producing a more user-friendly questionnaire is evident from the significant reduction in difficulties associated with the encoding and judgement components of the question-answering process. The CI process has enabled an insight into the methods that participants understood and retrieved information relating to questions about each of the FITT principles, in particular duration and frequency. Through its improved functionality and design, the PARQ should now provide more accurate information relating to the frequency, intensity and duration of different types of PA. In this way, PARQ can provide information that is relevant for the purpose of assessing whether respondents are meeting the minimum participation guidelines for health benefits, as outlined by the ACSM/CDC (Pate et al., 1995). Whether the CI process has produced an accurate questionnaire that can provide this information will be indicated in the reliability and validity studies which are outlined in Chapters five and six.

CHAPTER 5 - ASSESSING THE RELIABILITY OF PARQ.

5.1 Introduction

The previous chapters outlined the methods used to generate and pre-test items for the Physical Activity Recall Questionnaire (PARQ). This questionnaire was produced to acquire information about physical activity (PA) participation in an Irish adult population, aged 18-55 years. Prior to being made available for experimental or field-based studies however, a new questionnaire must affirm acceptable levels of validity and reliability (Litwin, 1995). *Validity* is a determination of the extent to which an instrument measures what it is supposed to measure and this is dealt with in Chapter Six. Reliability of a questionnaire is "a measure of the consistency of the data, usually determined by the test-retest method, where the first measure is compared to the second measure on the same subjects, under the same condition" (Vincent, 1999).

The reliability of PA questionnaires has usually been determined by a Pearson's or Spearman's interclass correlation coefficient (Sallis et al., 2000), see also Table 2.5. Some authors have suggested that an interclass correlation statistic is an inappropriate method of assessing *test-retest reliability* because Pearson's r is unable to detect the change in the means (Patterson, 2000; Vincent, 1999). The reliability of the Kaiser PA Survey for women was assessed with an intraclass correlation because "...it gives a measure of the relative homogeneity of the test-retest survey scores within subjects in relation to the total variation between subjects" (Ainsworth et al., 2000b, p.1330). In other words, unlike the interclass correlation, the intraclass correlation is sensitive to changes in the magnitude and order of scores between each individual's own test and retest. The intraclass correlation's ability to analyse repeated measures on the same variable means it can more appropriately analyse the mean differences of a test-retest questionnaire reliability study (Vincent, 1999).

The use of Pearson's correlation coefficient has also been questioned by Bland and Altman (1986, 1995), as it does not signify the extent of agreement between the test and retest scores. These authors suggest that the coefficient of repeatability (R) should be used for assessing the test-retest reliability of a questionnaire. In addition to indicating the level of agreement between the two scores, this method provides information about the heteroscedasticity of the questionnaire, i.e., whether the measurement error relates to the magnitude of the

measured variable. In the case of a PA questionnaire, this would mean assessing if error increases as the amount of PA reported increases (Atkinson & Nevill, 1998). This statistical method also facilitates calculation of a limits of agreement (LOA), which represents the test-retest differences for 95% of the population (Bland & Altman, 1986, 1995). The use of this statistical analysis procedure has recently been advocated as an appropriate method of assessment for reliability studies in the sport and exercise domain (Ainslie et al., 2003; Atkinson & Nevill, 2000; Atkinson et al., 1998) and was used to assess the reliability of SPAQ (Lowther et al., 1999).

Questionnaires that seek information about both moderate and high intensity activity have indicated that moderate intensity activities are less reliably recalled than high intensity (Craig et al., 2003; Sallis et al., 1985). This is possibly explained by the less salient nature and higher frequency of this intensity of activity. Wagenaar (1986) and Linton (1982) found, in their separate six-year diary studies of autobiographical memory, that information about less frequently occurring events were easier and more accurately remembered. Other factors known to affect reliability include (i) the presence of a personal calendar recall cue, (ii) the characteristics of the population, (iii) the interval between test and retest, and (iv) question order effects.

(i) Recall cue - Personal Calendar

When recalling events that are restricted to a particular timeframe, the accuracy of recall can be affected by factors such as telescoping (i.e., the inclusion of data from outside the relevant timeframe) and omission of data (Means et al., 1991; Brown et al., 1985). This is particularly evident for activities that are frequent and usual (Conway, 1996). Memories for specific, usual events are known to lose their distinctiveness over time and can often blend with similar memories, making it more difficult to retrieve episodic information and increasing the likelihood that what is remembered is inaccurate (Linton, 1982; Wagenaar, 1986). Recall has been shown to be improved by constructing a personal time line that permits respondents to date activities more accurately relative to landmark (important) events that they can remember with high degrees of accuracy (Belli et al., 2001; Means et al., 1991; Brown et al., 1985).

The inclusion of a personal timeline in a questionnaire, as an aid to recalling activities or events, has two advantages that are associated with the structure of Conway's (1993, 1996)

model of autobiographical memory (ABM). Firstly, according to Conway, memories in ABM are essentially reconstructions of the perceived self. As this personal reference system is the integral organising system in ABM, it implies that personal landmarks are of more assistance than non-personal landmarks in cueing the accurate recall of ABM's (Rubin & Kozin, 1984). In a study that asked freshman undergraduates (N=75) to recall three flashbulb memories (flashbulb is defined as a major public event) and to rate them for national and personal importance. Rubin and Kozin found that over 70% of the events recalled were rated as above average in personal importance and less than 5% were rated as nationally important. This indicates that salient personal events are the usual types of experiences that stand out in ABM. Research using personal time lines and calendars has rarely directly asked individuals for personal landmarks in their life (Shum, 2001). Interview-led questionnaires (Belli, 1998; Means et al., 1991) and a *lifetime* physical activity questionnaire (Friedenreich et al., 1998) have used specific topics such as "place of work" and "health care events" as landmark events, to help the respondent place themselves at a given point in time.

Secondly, timelines or calendars with personal landmarks are believed to improve recall because they emulate the thematic and temporal arrangement of Conway's (1993, 1996) structure of ABM (Belli, 1998). The personal landmarks provide multiple cues to the respondent about where they were, what they were doing and who they were with, during a specific timeframe. This cueing device thus provides access to the retrieval of information across multiple temporal and thematic indices in ABM. Respondents of a health survey (N=40, Male = 50%, age=21-65 years) were asked to recall health visits over the previous 12 months. Participants were randomly assigned into one of two groups, i.e. to answer the survey (i) with or (ii) without any cueing assistance. Accuracy was assessed by comparing the respondent's information about when and why the visit occurred, with their medical records. Results indicated that the use of the personal timeline provided more accurate dating and detailing of events. The accurate reporting of individual events that were recurrent doubled in comparison to when a personal timeline was not available as a recall cue (Means & Loftus, 1991). Implications for questionnaire design, from the results of this and Rubin and Kozins (1984) studies, suggest that the inclusion of a personal timeline calendar would provide an effective recall cue as it reflects the storage and retrieval process of ABM, as detailed by Conway (1996).

Event history (EHC) and personal calendars have, to date, been confined to use in interview-led questionnaires that are concerned with time periods of months and years. Further research is required to investigate the use of a personal time line in a short-term recall questionnaire and more specifically a PA recall questionnaire. Investigation of the effectiveness of the personal timeline as a recall cue is limited (Belli et al., 2001). In Belli and colleagues study, an interview-led questionnaire that used an EHC was compared against the same questionnaire without an EHC. This study was concerned with a retrospective questionnaire for social and economic behaviours for the previous two years. Respondents and 20 interviewers were randomly assigned to the EHC interview (N=309) or question-list interview (N=307). The EHC group demonstrated significantly higher correlations and levels of agreement with previously recorded data for the same period, compared to the question-list interview condition. An investigation of whether the inclusion of a personal timeline will enhance the recall of information in a PA recall questionnaire is required and is an objective of this study.

(ii) The characteristics of the sample population

Characteristics such as age, educational status, physical activity stage of change and gender should reflect the general population for which the final questionnaire is intended. One feature of some reliability studies is a homogeneous element within the sample, e.g., educational status, " ... [participants were taken] from the local university community with college or graduate degrees and employed in administrative or professional positions" (Jacobs, Ainsworth, Hartman, & Leon, 1993, p.83). Another study recruited all subjects from an aerobics class or an exercise promotion project (Lowther et al., 1999). Homogeneity within a sample does not provide a true indication of the functionality of the questionnaire in the general population and reliability will also be adversely affected if the context of the questionnaire is inappropriate for the sample participants (Ainsworth, 2000). For questionnaires that are intended for population-based studies, a sample that is more representative of the general population's average age, education level etc., is preferential.

(iii) The timing of test-retest administrations

Test-retest administrations should be conducted across time intervals that reflect the timeframe of the questionnaire (Patterson, 2000). Thus, a past-seven-day recall PA should be

retested within seven days, so that the data of the overlapping days can be compared. If the retest time is outside of the past seven days, the results are more likely to indicate actual differences in behaviour (e.g., physical activity), than be indicative of the reliability properties of the questionnaire (Presser et al., 2004; Sudman et al., 1982). Questionnaires asking for the recall of specific events but not covering an overlapping time period reported only moderate correlations (Kriska et al., 1997; Baranowski, 1988), also see Table 2.5. Test-retest data for a past-two-week questionnaire, when collected for the same two weeks (N=115) was compared against data collected for two different weeks (N=116). Results (Female=45%, age range=18-79 years) indicated superior reliability for the same two week period (r=0.86) over for the different two week time period (r=0.56) (Booth et al., 1996). The optimum test-retest interval has not yet been determined, but times between tests should be long enough to reduce the opportunity for remembering what was written and short enough in order to remove the opportunity for true variability in activity levels (Sallis et al., 2000; Patterson, 2000).

(iv) Question order effects

Testing of question order effects dominates in attitude and opinion surveys (Tourangeau et al., 2000). They can occur when questions have a close substantive relationship to one another, so that answers to one question may have logical implications for others (Strack et al., 1985; Sudman et al., 1982). The order effect of two questions posed to members of the American public (N=140, mean age=34.6 years) was assessed (Schuman et al., 1983). Questions related to 1) access of American journalists into a communist country to freely report events back to their own country, and 2) the access of communist journalists into America to freely report events back to their own country. Findings showed that participants' agreement with allowing communist journalists to freely report from America increased from 55% to 75% when the question about allowing access to American reporters in a communist country came first. An examination of whether the reliability of PA questionnaires is susceptible to question order effects is not evident in literature and thus requires further investigation (Friedenreich, 1994).

In summary, this study initially assessed, (i) the test-retest reliability of PARQ without a personal calendar (PC). These results were then compared to the reliability scores of PARQ when a PC was included as a recall cue. Secondly, the test-retest reliability of PARQ was assessed using (ii) a diverse sample and (iii) a four-day overlapping methodology. This study

was also intended to evaluate a specific design issue within the questionnaire by assessing (iv) the effect that changing the order of the questions would have on the reliability of PARQ. Parts (i) and (iv) represent novel approaches to the assessment of reliability in a PA recall questionnaire.

5.1.2. Hypotheses

The following hypotheses were formulated with respect to the cognitive recall of physical activity:

(i) That the questionnaire which includes the personal calendar will have better reliability scores than the questionnaire without the personal calendar.

(ii) That high intensity physical activity will have better reliability scores than moderate intensity physical activity.

(iii) That assessment of the question order effect will show that the group beginning with a high intensity question on both test and retest will have better reliability scores than those beginning with a moderate intensity question on both test and retest.

(iv) That the reliability score will be better for the type of activity that is performed less frequently than the others.

(v) That PARQ will be reliable for participants at all SOC but show a higher reliability score for the not regularly active participants (SOC 1-3), because of lower reported activity levels.

5.2. Methods

5.2.1. Participants

Upon ethical approval from DCU, 210 participants were recruited for this study from various Dublin businesses, hospitals, a school staff room, a university campus and a training programme for the unemployed. All participants were requested to complete a screening form (Appendix one) that sought information relating to age, gender, occupation status, place of residence (Dublin vs. outside Dublin) and physical activity stage of change (SOC). SOC has been previously explained in chapter two. One hundred participants were selected for each part of the study by stratified sampling (Thomas et al., 2001). This form of sampling occurred by firstly dividing the volunteers by one specific criteria and then randomly selecting them for

participation. In this study, the dividing criterion was physical activity SOC and participants were divided into group one (stages 1-3) or group two (Stage 4-5). This was because stages 1-3 represent those that are not yet regularly physically active, in contrast to those in stages 4 and 5.

5.2.2. Equipment

PARQ (Questionnaire)

Part one of study (questionnaire without a personal calendar recall cue)

Questions one and two asked participants to complete questions about (i) high intensity and (ii) moderate intensity physical activity. Within each intensity question, five sub-questions asked about the amount of minutes that was spent doing activity in different categories, i.e., 1) at home, 2) sport and exercise 3) caring for family members, 4) at work and 5) commuting. The same categories were asked for each intensity. Different descriptions of physiological and psychological changes that the respondent would expect to feel for each intensity were given. It was estimated that time for completion of this questionnaire would be approximately 15 minutes.

Part two of study (questionnaire with a personal calendar recall cue)

The questionnaire for this study was a modified version of the PARQ described above. Prior to beginning the PARQ, a past-seven-day personal calendar (PC) was completed. The PC asked participants to recall anything that they remembered happening over each of the past-seven-days. It was estimated that time for completion of this version of PARQ would be approximately 20 minutes.

5.2.3. Procedure

All participants were requested to complete the PARQ on day one and day four of a one week period. As identified in the literature, this design reduced the possibility of variation caused by actual change in physical activity patterns. This provided a four day overlap of activity information which could be used to assess the reliability properties of the questionnaire. The four days included information on two weekend days and two weekdays for all participants. The design of the test-retest administration time-scale is provided in Table 5.1 overleaf.

Table 5.1. *Design of the test-retest administration of questionnaires*

	Mon	Tues	Wed	Thurs	Fri	Sat	Sun	Mon	Tues	Wed	Thurs
Time1	Days of Activity recalled during completion of PARQ Time 1 (past seven days)							Complete PARQ			
Time2				Days of Activity recalled during completion of PARQ Time 2 (past seven days)							Complete PARQ
				Thurs	Fri	Sat	Sun				

Note. Shaded area = Four days of overlapping data which was analysed

Informed consent was obtained on day one. Participants were asked to return on day four in order to receive their travel expenses payment (€15 for each participant) and some personal feedback. No indication of having to complete a second recall questionnaire was given to the participants. All participants used the exact same facility for their test-retest administrations, i.e., either a designated room at the participant's workplace or the Centre for Sport Science and Health at Dublin City University. Before completing the questionnaire, all participants were instructed that participation was absolutely voluntary and that they were free to leave at any time. Issues of confidentiality were reinforced through the use of assigned identification numbers in the final report rather than individual names. This was intended to reduce social desirability effects. Participants were given no specific instructions to complete the questionnaire, other than to complete it as if they were at home. No assistance was given by the researcher and no questions raised by the participants were answered. This was to ensure uniformity and to assess the functionality of the self-administered aspect of PARQ.

In order to (i) reduce the possibility of participants becoming familiar with the layout of the questionnaire (a familiarity effect) and (ii) to assess the presence of a question order effect, the order of questions relating to intensity was altered at random for all participants between test and retest. Participants were thus divided into one of four groups, labelled according to whether the order of the first question referred to either high or moderate intensity at Time 1 (Test) and Time 2 (Retest), respectively. The groups were labelled as; Group 1= High-High, Group 2 = High-Moderate, Group 3 = Moderate-Moderate, Group 4 = Moderate-High. The random assignment of individuals into one of four groups occurred as they presented themselves to the

researcher. Questionnaires were laid out in four bundles on the table according to the order of their questions as described above. Questionnaires were then administered to participants in a consecutive circular method from the bundles (See Figure 5.1. overleaf). This was done for both parts one and two of the study, i.e. (i) questionnaire without the personal calendar recall cue and (ii) questionnaire with the personal calendar.

Figure 5.1. Demonstration of the random allocation of participants to a group.

5.3. Reliability of total, high and moderate intensity physical activity

5.3.1. Part One (PARQ without personal calendar)

The following section outlines the results of the test-retest reliability for part one (PARQ with no personal calendar). This is the first part of testing Hypothesis One, i.e., that the questionnaire which includes the personal calendar will have better reliability scores than the questionnaire without the personal calendar.

5.3.1.1. Data Analysis

All data from both the test and retest administrations of the questionnaire were entered into a statistical package (SPSS, version 10) for analysis. Univariate distributions of PA at Time one (Test) and Time two (Retest) were inspected for normality and screened for outliers. The removal of one outlier identified by SPSS significantly improved skewness values, resulting in a normal distribution that was confirmed by a test of normality, the Kolmogorov-Smirnov, ($Z=1.145$, $p=0.145$). This left a total of 70 (98.5%) complete data sets for analysis. Analysis of total, moderate and high intensity physical activity was undertaken. Pearson correlation coefficients (r), intraclass correlation (ICC) and the coefficient of repeatability (R) were calculated, to provide a comprehensive assessment of reliability.

5.3.1.2. Results

Descriptive statistics

There were 70 complete sets of data were analysed (mean age=19.75 years (SD+/- 2.6), Female=53.5%). In total, 81.4% of all participants were at either the action or maintenance stages of change. Table 5.2. shows the wide range in activity reported for both test and retest conditions. Differences between test and retest for total activity appear to be predominantly influenced by large variations in moderate intensity activity. A paired sample t-test confirmed no significant difference between test and retest questionnaires for total activity (t (69)=1.88, p=0.06), high intensity (t (69)=0.65, p=0.52) and moderate intensity (t (69)=1.58, p=0.12).

Table 5.2. *Descriptive statistics for test and retest questionnaires measured over four overlapping days, for total, high and moderate intensity activity (minutes).*

Measurement and intensity		N	Mean (mins)	Median (mins)	SD (mins)	Min (mins)	Max(mins)
Total	Test	70	605.3	500	475.1	202	3180
	Retest	70	543.3	427.5	387.9	34	2190
High	Test	70	220.3	160	171	0	690
	Retest	70	209.16	175	166.24	0	630
Mod	Test	70	385.04	272.5	417.4	10	3030
	Retest	70	334.19	252.5	297.57	34	1610

Note. Mod= Moderate, SD = Standard Deviation.

Pearson's correlation coefficients

The relationship between test and retest minutes of PA was initially evaluated with a Pearson's correlation coefficient (r). Results indicated a strong correlation (r=0.81) for total activity. Correlations for individual intensities showed that moderate intensity activity (r=0.76) was unexpectedly more reliably recalled than high intensity (r=0.64). All results were significant ($p<0.01$).

Intraclass correlations coefficients

The average measure intraclass correlation coefficient is acceptable for total activity at 0.88 (one-way random effect model). Values above 0.7 are considered acceptable (Vincent, 1999). Acceptable ICC values were also found for high intensity = 0.78 ($F(1, 69) = 0.42$, $p>0.05$) and moderate intensity =0.84 ($F(1, 69) = 2.5$, $p>0.05$). A non-significant F-value in the analysis of variance ($F(1, 69) = 3.54$, $p>0.05$) shows consistent reporting for total physical activity.

Coefficient of Repeatability (R)

The extent of the agreement between the test and retest questionnaires was assessed (Bland & Altman, 1986) (for description of procedure see Appendix ten). The descriptive statistics for the differences between the test and retest are displayed in Table 5.3.

Table 5.3. *Descriptive statistics for the differences between test and retest PA (minutes) for all participants.*

Category	N	Mean Difference (mins)	Standard Deviation (mins)
Total activity	70	61.97	275.66
High Intensity	70	11.11	143.5
Moderate Intensity	70	50.86	269.04

Calculation of the coefficient of repeatability (twice the standard deviation of the differences) and the 95% limits of agreement are detailed in Table 5.4 below.

Table 5.4. *Coefficient of Repeatability and 95% Limits of agreement for PARQ without the personal calendar.*

Category	Coefficient of Repeatability (R) (2 SD's) (minutes)	Limits of Agreement (MD + /-2SD's) (minutes)
Total Activity	551.32	-489.32 to 613.29
High Intensity	287	-275.89 to 298.11
Mod Intensity	538.08	-487.22 to 588.94

Note. MD = Mean Difference, SD=Standard Deviation.

The coefficient of repeatability (R) for total activity is 551 minutes. The results in Table 5.4 above show that 95% of the differences in the questionnaire from one measurement to the next, under similar conditions, would be between –489 minutes and 613 minutes, for this sample. These are known as the limits of agreement (LOA) (Bland et al., 1986).

5.3.2. Part two - (PARQ with a personal calendar)

The following section outlines the results of the test-retest reliability of total, high and moderate intensity physical activity for part two (PARQ with a personal calendar) and compares them to the reliability scores of the questionnaire from part one (PARQ without a personal calendar). This is the second part of testing Hypothesis One, i.e., that the questionnaire which includes the personal calendar will have more reliable scores than the questionnaire without the personal calendar.

5.3.2.1. Data Analysis

All data from both the test and retest administrations of the questionnaire were entered into a statistical package (SPSS, version 10) for analysis. Univariate distributions of PA at Time one (Test) and Time two (Retest) were inspected for normality and screened for outliers. The removal of two outliers identified by SPSS significantly improved skewness values, resulting in normal distribution that was confirmed by a test of normality, the Kolomogorov-Smirnov, ($Z=1.167$, $p=0.131$). This left a total of 73 (97%) of complete data sets for analysis. Analysis of total, moderate and high intensity physical activity was undertaken. Pearson correlation coefficients (r), intraclass correlation (ICC) and the coefficient of repeatability (R) were calculated, to provide a comprehensive assessment of reliability.

5.3.2.2. Results

Seventy three participants completed this study (mean age 31.4 years (+/- 10.1), F=42%). In total, 57.1% of all participants were in full-time employment, 15.2% were long-term unemployed and 27.7% were university students, some of whom had part-time jobs. Fifty two percent of the sample described themselves as being at either the action or maintenance stage of the SOC model.

5.3.2.2.1. Total activity, high intensity and moderate intensity

Descriptive statistics

Table 5.5. displays the descriptive statistics for all participants minutes of total, high and moderate intensity PA. A paired sample t-test confirmed no significant difference between test and retest questionnaires for total activity (t=-0.34, df=72, p=0.78), high intensity (t=0.29, df=72, p=0.78) and moderate intensity (t=-0.52, df=72, p=0.6).

Table 5.5. *Descriptive statistics for test and retest questionnaires measured over four overlapping days, for high and moderate intensity and total physical activity.*

Questionnaire		N	Mean (mins)	Median (mins)	SD (mins)	Min (mins)	Max(mins)
Total	Test	73	314.14	250	264.43	0	1500
	Retest	73	319.89	231	265.49	15	1500
High	Test	73	78.08	60	97.01	0	560
	Retest	73	76.48	40	107.69	0	630
Mod	Test	73	241.81	155	216.68	15	1220
	Retest	73	248	195	209.58	0	1200

Pearson's product moment correlation coefficient

The relationship between test and retest minutes of physical activity was initially evaluated by means of a Pearson correlation coefficient. Both intensity questions and total physical activity displayed significant correlations that were higher than those reported in study one (see Table 5.6. below). In study two, moderate intensity physical activity has a lower correlation compared to high intensity physical activity data.

Table 5.6. *Pearson's product moment correlation coefficient for test-retest data on four overlapping days, for both experimental conditions*

Condition	Total physical activity	High Intensity	Moderate Intensity
No calendar (study one)	0.81**	0.64**	0.76**
With calendar	0.91**	0.9**	0.88**

(study two)

**p<0.01 level (2 tailed)

Across all categories (total, moderate and high intensity), the correlations are lower for study one (PARQ without PC). Though all results are significant, the absence of the personal calendar recall cue appears to have had a notable effect on all questions, particularly high intensity. This result supports Hypothesis One, which states that the questionnaire which includes the personal calendar will have more reliable scores than the questionnaire without the personal calendar.

Intraclass correlations coefficients

The average measure intraclass correlation coefficient is acceptable for total activity at 0.95 (one-way random effect model). Values above 0.7 are considered acceptable (Vincent, 1999). Acceptable ICC values were also found for high intensity = 0.94 ($F(1, 72) = 0.08$, $p>0.05$) and moderate intensity =0.94 ($F(1, 72) = 0.26$, $p>0.05$). A non-significant F-value in the analysis of variance ($F(1, 72) = 0.12$, $p>0.05$) shows that the reported test-retest minutes for total physical activity are very consistent.

Limits of agreement

The extent of the agreement between the test and retest questionnaires was assessed (Bland et al., 1986). A limits of agreement was conducted for total activity and also for each intensity. The descriptive statistics for the differences between the test and retest are displayed in Table 5.7.

Table 5.7. *Descriptive statistics and t-test results for the differences between test and retest physical activity (minutes) for all participants*

Variable	N	MD (mins)	t	p	SD of MD (mins)	R (2 SD's)	LOA (MD + 2SD) (mins)	LOA (MD-2SD)(mins)
Total	73	5.75	0.43	0.66	113.83	227.66	233.41	-221.91
High	73	1.74	0.31	0.75	47.63	95.26	97	-93.52
Moderate	73	-6.19	0.52	0.67	102.45	204.9	198.71	-211.09

Note. MD= Mean Difference, SE= Standard Error, SD= Standard Deviation, R= Coefficient of Repeatability, LOA= Limits of Agreement

As the same measurement method was used on both occasions, it was expected that mean differences between test and retest would not be significantly different than zero. Bland and Altman (1986) recommend the use of one sample t-tests to assess this and results indicated no significant difference (Table 5.7.). For total physical activity, the mean difference is 5.75 minutes and the standard deviation is 113.8 minutes, thus the coefficient of repeatability (R) is 227.6 minutes (i.e., twice the SD). The total activity data appears to be influenced mostly by the moderate intensity data. The coefficients of repeatability are lower for each category in part two (PARQ with personal calendar), compared to part one (PARQ without personal calendar).

The limits of agreement (MD +/- 2SD's) are given in Table 5.8. above. Under a normal distribution, 95% of the differences should be less than two standard deviations from the mean difference (British Standards Institution, 1979). This is graphically represented for total physical activity in Figure 5.2. overleaf.

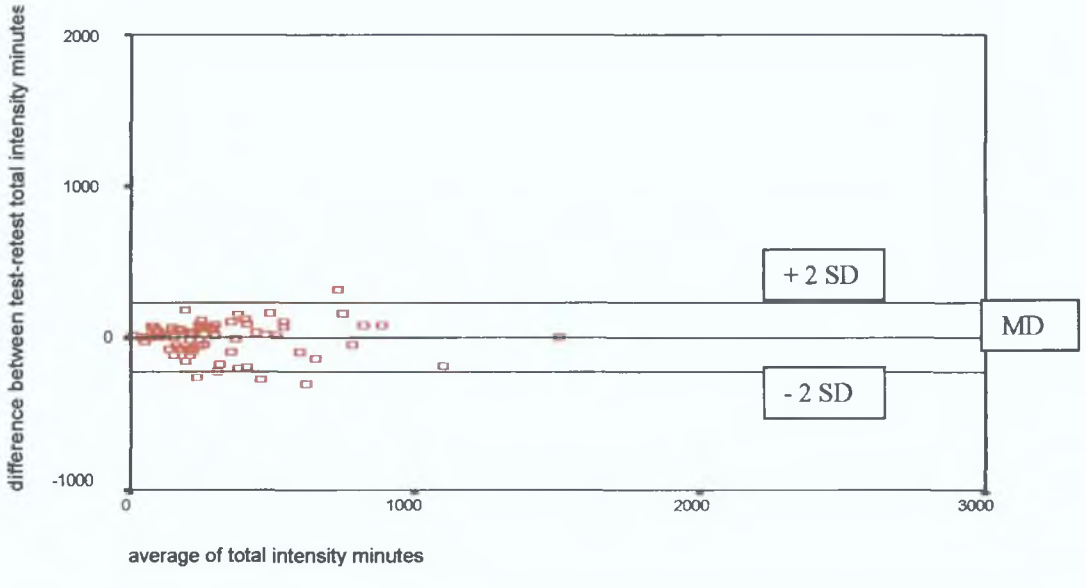

Figure 5.2. The differences in test-retest results (minutes) plotted against the average test-retest results (minutes) for total activity (with personal calendar). Note. MD=Mean Difference, SD = Standard Deviation.

The limits of agreement and 95% confidence intervals are better in all intensities for the questionnaire that has used the personal calendar as a recall cue. The plot of mean differences scores against average scores, for both conditions total activity, illustrates a much wider spread

of data away from zero with increasing magnitude, for the questionnaire without the personal calendar (Figures 5.2. and 5.3.).

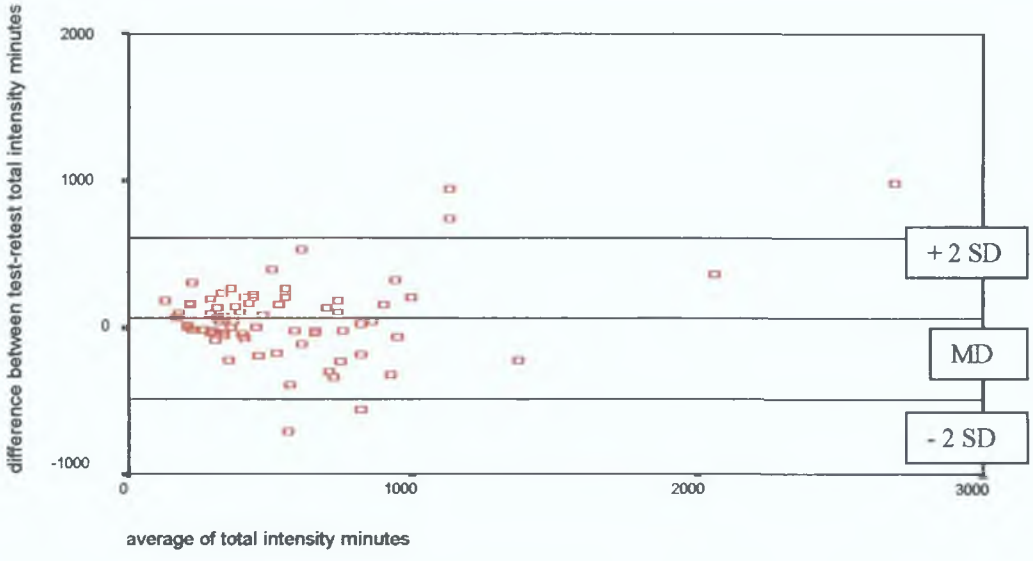

Figure 5.3. The differences in test-retest results (minutes) plotted against the average test-retest results (minutes) for total activity (*without* personal calendar).

Hypothesis One

This stated that the questionnaire which includes the personal calendar (PC) will have better reliability scores than the questionnaire without the personal calendar. A comprehensive analysis of the reliability of both formats of the PARQ has revealed support for this hypothesis. Analysis by (i) Pearson's correlation coefficient, (ii) Intraclass correlation coefficient, and (iii) the coefficient of repeatability, have each displayed better reliability properties for the PARQ, when the PC is included. Although a significant difference between the age profile of both groups was observed, independent t-tests revealed there were no significant differences in the mean total, high and moderate intensity activity between both groups (see Table 5.9 below).

Table 5.8. *Differences in total, moderate and high intensity activity and age between PARQ with and without a personal calendar.*

Variable		M (SD) minutes	t	p	df
Total activity	PARQ with PC	-4.52 (112.9)	-1.9	0.06	141
	PARQ no PC	61.97 (275.6)			
High intensity	PARQ with PC	1.6 (47.9)	-0.6	0.6	141
	PARQ no PC	11.1 (143.5)			

Moderate intensity	PARQ with PC	-6.19 (102.4)	-1.69	0.9	141
	PARQ no PC	50.9 (269.1)			
Age	PARQ with PC	31.3 (10)	9.3	<0.01	141
	PARQ no PC	19.8 (2.5)			

As the age range of both groups was appropriate to the criterion age profile of the questionnaire (18-55 years), the results of the t-test and reliability tests suggest that PARQ with the PC enhances recall and creates a more reliable questionnaire. Therefore, further analysis of (i) the 95% confidence intervals of the LOA, (ii) the question order effect, (iii) reliability of the different types of activity and (iv) reliability of stages of change, will be assessed using only the data from the PARQ with a personal calendar.

Limits of Agreement (LOA) 95% confidence intervals

As the 95% limits of agreement found in this study may not be true for another sample under similar conditions, Bland and Altman (1986, 1995) advise the calculation of 95% confidence intervals for each limit. This will indicate the precision of the limits of agreement in study two.

The equation for calculating the confidence interval = sqrt $(3*SD^2/N)$, with t=1.667 at 72 degrees of freedom. For total physical activity, this means = sqrt $(3*113.83^2/73)$ = 23.07

Thus the 95% confidence interval for total physical activity is:

Upper limit of agreement (233.41 minutes) = 233.41+ (1.667*23.07) to 233.41-(1.667*23.07) = 194.5 minutes to 271.86 minutes.

Lower limit of agreement (-221.91 minutes) = -221.91 + (1.667*23.07) to –221.91 – (1.667*23.07) = -183.46 minutes to –260.3 minutes.

The upper limit of agreement for high intensity is 80.9 minutes to 113.09 minutes, while the lower limit was calculated as –77.43 minutes to -109.61 minutes.

The upper limit of agreement for moderate intensity is 177.95 minutes to 219.47 minutes, while the lower limit of agreement is –190.33 minutes to –231.85 minutes.

These 95% confidence intervals are comparable to the 95% LOA reported in Table 5.7. above. This indicates that the LOA reported for this sample are accurate and transferable to the general population. These results indicate that in similar conditions, 95% of further samples will vary between –221 and 233 minutes when reporting total physical activity. The total activity data appears to be influenced mostly by the moderate intensity data. The favourable comparison of R for high intensity (95.2 minutes) against moderate intensity (204.9 minutes) lends further support to hypothesis two.

5.4. Reliability of the question order effect and different types of physical activity

Hypothesis Two

This stated that high intensity physical activity will have more reliable scores than moderate intensity physical activity. The ICC displayed identical results for both high and moderate intensity (0.94) in study two. The Pearson's correlation coefficient were similar (r=0.9 for high, r=0.88 for moderate). The coefficient of repeatability (R) analysis above, reported larger differences in the measure of repeatability for moderate intensity over high intensity. While this shows support for hypothesis two, an examination of the question order effect and different types of activity, might provide information about whether other factors are responsible for the large R reported above.

5.4.1. Question order effect (Hypothesis Three)

Hypothesis three stated that assessment of the question order effect will show that the group beginning with a high intensity question on both test and retest will have better reliability scores than those beginning with a moderate intensity question on both test and retest. The following section displays the results of the four test conditions (groups) that explored the question-order effect.

Descriptive Statistics

Table 5.9. displays the descriptive statistics for each group's physical activity at test and retest for all participants.

Table 5.9. *Descriptive statistics for test and retest questionnaires for each group's total physical activity (minutes).*

Measurement and group		N	Mean (minutes)	SD (minutes)	Min (minutes)	Max (minutes)
One	Test	20	220.3	215.9	15	1000
	Retest	20	234.2	246.7	0	1190
Two	Test	17	287.59	183.58	95	855
	Retest	17	261.29	206.08	20	775
Three	Test	20	372.7	226.44	99	920
	Retest	20	353.55	211.17	100	840
Four	Test	16	412.69	388.68	78	1500
	Retest	16	420.94	361.92	45	1500

Note. Group one=high-high, group two=high-moderate, group three=moderate-moderate, group four=moderate-high

Repeated Measures ANOVA

A repeated measures ANOVA was undertaken to assess whether there were significant differences between the test-retest scores of each of the four groups. Within-subject differences were not significant ($F(1, 69)=0.12$, $p> 0.05$). This indicated that scores did not significantly differ from time one to time two, for each group. Between-subject differences were also not significant ($F(1, 69)=2.03$, $p>0.05$). This means that the effect of a familiarity bias on respondents (in answering questions) between the four groups was not significant. It also means that no order effect was observed, indicating that reliability would was not improved by placing the high intensity question first, thus rejecting Hypothesis three.

5.4.2. Type of activity (Hypothesis Four)

Hypothesis four stated that the reliability score will be better for the type of activity that is performed less frequently than the others. The following section displays the results of each of the five types of activity. They have been separated here to assess their relative contribution to the overall total activity data.

Descriptive statistics

Table 5.10. displays the descriptive statistics for each type of physical activity at test and retest for all participants.

Table 5.10. *Descriptive statistics for test and retest questionnaires for each type of physical activity (minutes).*

Type and Measurement		N	Mean (minutes)	SD (minutes)	Min (minutes)	Max (minutes)
Housework	Test	73	83.92	112.3	0	560
	Retest	73	70.62	96.77	0	420
Sport & Exercise	Test	73	116.75	138.77	0	800
	Retest	73	127.34	157.71	0	920
Family	Test	73	27.97	71.45	0	480
	Retest	73	23.08	69.72	0	480
Occupational	Test	73	30.77	42.75	0	270
	Retest	73	33.73	63.58	0	375
Commuting	Test	73	62.05	95.95	0	620
	Retest	73	69.71	88.58	0	560

Pearson's correlation coefficient

Assessment of the relationship between test and retest minutes of each type of physical activity (Table 5.11. below) revealed strong correlations. Occupational activity was moderately correlated at r=0.7, indicating relative difficulty in reporting this information. Family activity has the highest correlation.

Table 5.11. *Pearson correlation coefficients by type for test-retest results of total activity.*

Type	Correlation coefficient
Housework and Gardening	0.9**
Sport and Exercise	0.9**
Family	0.95**
Occupational	0.7**
Commuting	0.82**

**p<0.01 (2-tailed).

Intraclass correlations coefficients

A non-significant F-value in the analysis of variance for each of the types of physical activity shows that the reported test-retest minutes are consistent. The average measure intraclass correlation coefficient is acceptable for each type (one-way random effect model). Family activity has the highest ICC (Table 5. 12.).

Table 5.12. *ICC correlations by type for test-retest results of physical activity.*

Type	F-value	P	Degrees of freedom	Correlation
Housework and Gardening	5.46	0.22	72	0.94
Sport and Exercise	1.62	0.2	72	0.94
Family	3.67	0.6	72	0.97
Occupational	0.3	0.58	72	0.79
Commuting	1.36	0.25	72	0.84

Limits of Agreement

Descriptive statistics for the differences between the test and retest and limits for agreement for each category are displayed in Table 5.13.

Table 5.13. *Descriptive statistics for the differences between test and retest results, coefficients of repeatability (R) and limits of agreement (LOA) for all types (minutes).*

Type	MD (mins)	SD (mins)	R (2SD's)	LOA (MD+2SD) (mins)	LOA (MD-2SD) (mins)
House	13.3	48.65	97.3	110.6	-84
Sport	-10.59	70.44	140.88	130.29	-151.47
Family	4.9	21.8	43.6	48.5	-38.7
Occupational	-2.6	45.7	91.4	88.8	-94
Commuting	-7.66	56.08	112.16	104.5	-120.32

Note. LOA= Limits of Agreement, MD= Mean Difference, SD=Standard Deviation

One sample t-tests confirmed that mean differences between test and retest questionnaires were not significantly different than zero, for each of the types of PA. The coefficient of repeatability (R) for sport and exercise shows the highest variability. The LOA for this type of activity means that 95% of the differences in the questionnaire would be between –151 (mean difference minus R) and 130 minutes (mean difference plus R), from test to retest, under similar conditions. The sport and exercise result here is in contrast to the Pearson correlation and ICC results, where it showed one of the highest test-retest correlations. Similarly, occupational activity displayed the lowest correlations but reports the second lowest variability in its LOA (Table 5.13. above). The large limits of agreement for sport and exercise may be a function of the comparatively large amount of data that was recalled (Table 5.10.). The proportionate contribution of minutes performed at each intensity may also reveal some important information, as moderate intensity has been observed to be less reliable than the recall of high intensity activity (Tables 5.6. and 5.7.) (Hypothesis two). To examine this possibility, the high and moderate intensity limits of agreement for the sport and exercise category are compared against the family category (Table 5.14. overleaf). The family category was chosen for comparison because it has shown the highest correlations (Pearson's and ICC) and the best LOA.

Table 5.14. *Descriptive statistics for the differences between high and moderate intensity test and retest results and one sample t-test results for sport/exercise and family categories (minutes).*

Category	MD (mins)	t	p	SD (mins)	LOA (MD+2SD) (mins)	LOA (MD-2SD) (mins)
High intensity						
Sport	1.33	0.28	0.8	40.53	82.39	-79.73
Family	2.97	2.4	0.2	10.7	24.37	-18.43
Mod intensity						
Sport	-11.9	1.8	0.1	56.2	100.5	-124.3
Family	1.9	0.9	0.3	16.94	35.78	-31.98

Note. Mod = Moderate, MD= Mean Difference, SD= Standard Deviation, LOA = Limits of Agreement

One sample t-tests confirmed that mean differences between test and retest questionnaires were not significantly different than zero, for both types of PA. For the mean and SD of the differences to be meaningful estimates, it must be assumed that they are reasonably constant throughout the range of the measurement (Bland & Altman, 1995). Therefore, to see whether the increase in variability is related to an increase in magnitude of activity, the difference against the average of the two measurements was plotted. The variation of data around the mean (SD) is greater for both categories at moderate intensity (N=73). Scatter plots for this intensity confirm a greater increase in the scatter away from the mean as the magnitude increases for the sport and exercise category (Figures 5.5. and 5.6. overleaf). For family activity, only two data sets are >100 minutes and both were within 20 minutes of the mean difference. For sport and exercise, approximately 30 data sets report over 100 minutes and are scattered as much as 180 minutes from the mean difference. The greater number of people recalling larger amounts of sport and exercise data in moderate intensity appears to adversely affect recall. Similar plots of occupational, household and commuting activity reveal a similar trend, though numbers of participants recalling over 100 minutes are fewer (between four and ten) and scatters from the mean are less extreme (except for commuting, where one participant differed by over 200 minutes).

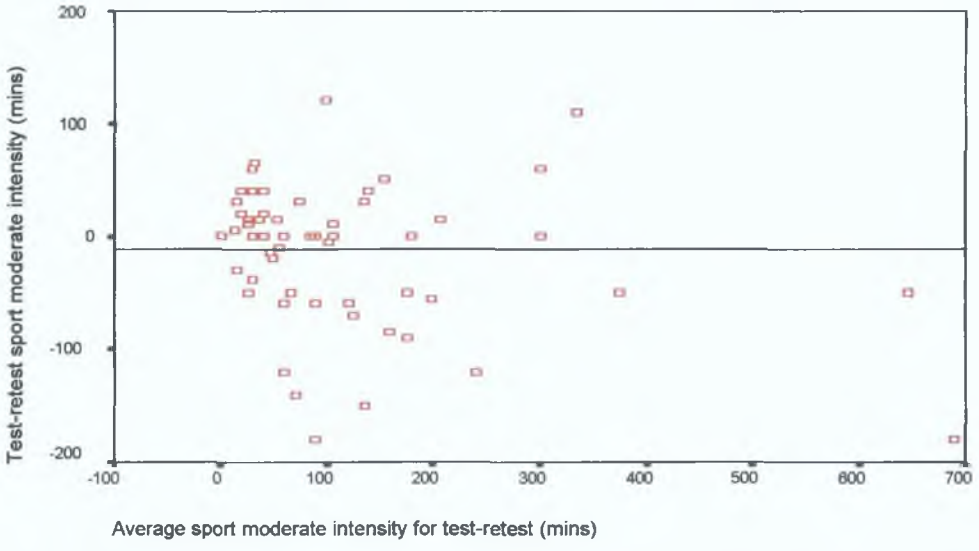

Figure 5.5. The differences in test and retest results (minutes) plotted against the average test and retest results for each participant for moderate intensity sport and exercise activity.

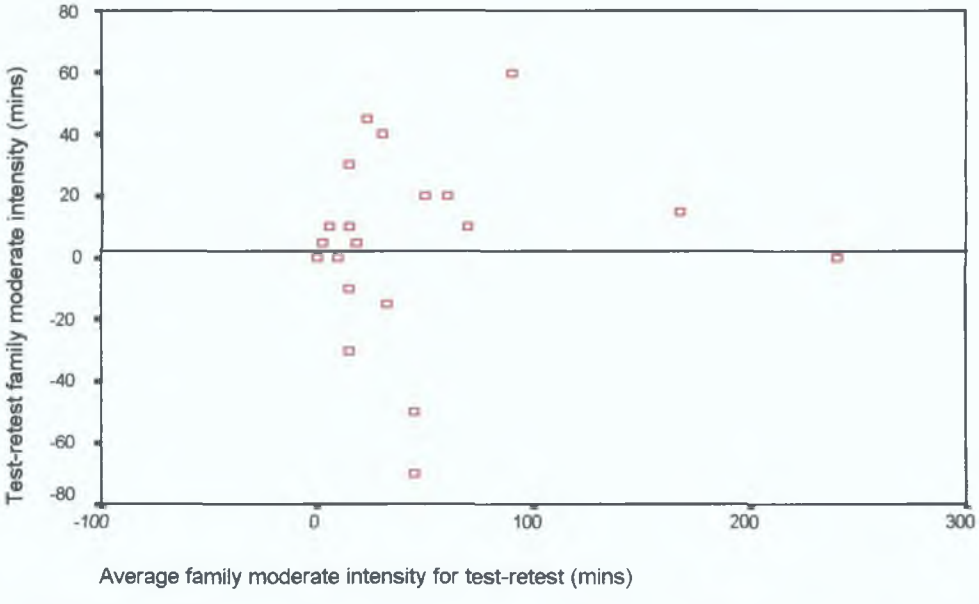

Figure 5.6. The differences in test and retest results (minutes) plotted against the average test and retest results for each participant for moderate intensity family activity.

Hypothesis Four

This hypothesis stated that the reliability score will be better for the type of activity that is performed less frequently than the others. Table 5. 10 shows that family activity has the lowest mean reported activity. Family activity also displayed the highest correlations (Pearson's and ICC) and the least variable LOA, therefore supporting hypothesis four.

5.4.3. Reliability by Stage of Change (SOC)

Hypothesis five stated that PARQ will be reliable for participants at all SOC but show a higher reliability score for the not regularly active SOC group (SOC 1-3), because of lower reported activity levels. The following section displays the descriptive statistics and reliability correlation analysis from (i) Pearson's and (ii) Intraclass coefficients, as well as the coefficient of Repeatability (R), for total, high and moderate intensity physical activity.

Descriptive Statistics

Tables 5.15-5.17 displays the descriptive statistics for each SOC group's physical activity at test and retest for all participants.

Table 5.15. *Descriptive statistics for test and retest questionnaires for each SOC group's total physical activity (minutes).*

Group/Measurement		N	Mean (mins)	SD (mins)	Min (mins)	Max (mins)
SOC 1-3	Test	35	289	236.96	15	920
	Retest	35	278.83	230.77	0	840
SOC 4-5	Test	38	348.34	289.54	79	1500
	Retest	38	361.29	206.08	64	1500

Note. SOC= Stage of Change, 1-3=not regularly active, 4-5=regularly active

Mean total activity recalled is lower, as expected, for SOC 1-3 group. Breakdown of this activity into high and moderate intensity is detailed below in Tables 5.16 and 5.17.

Table 5.16. *Descriptive statistics for test and retest questionnaires for each SOC group's high intensity physical activity (minutes).*

Group/measurement		N	Mean (mins)	SD (mins)	Min (mins)	Max (mins)
SOC 1-3	Test	35	53.77	69.61	0	385
	Retest	35	48.26	71.9	0	390
SOC 4-5	Test	38	100.47	113.07	0	560
	Retest	38	102.47	127.95	0	630

Note. SOC= Stage of Change, 1-3=not regularly active, 4-5=regularly active

Table 5.17. *Descriptive statistics for test and retest questionnaires for each SOC group's moderate intensity physical activity (minutes).*

Group/measurement		N	Mean (mins)	SD (mins)	Min (mins)	Max (mins)
SOC 1-3	Test	35	235.23	202.44	15	730
	Retest	35	230.71	195.29	0	790
SOC 4-5	Test	38	247.87	231.58	65	1220
	Retest	38	263.92	223.35	50	1200

Note. SOC= Stage of Change, 1-3=not regularly active, 4-5=regularly active

The tables above indicate that similar amounts of moderate intensity activity were recalled by both groups but that the regularly active participants (SOC 4 and 5) recalled double the amount of high intensity activity than the non-regularly active participants. Neither SOC group varied much in their reported means and standard deviations from test to retest, indicating consistency in the amount of data recalled. A one-sample t-test also indicated that differences between test and retest were not significant for (i) the not regularly active group ($t=0.74$, $df=34$, $p=0.6$) and (ii) the regularly active group ($t=0.74$, $df=37$, $p=0.7$).

Pearson's correlation coefficient

The relationship between test and retest minutes of activity was initially evaluated for each SOC group (regularly active and not regularly active) at each intensity by a Pearson's correlation coefficient. Both SOC groups indicated similarly high correlations for all intensities.

Table 5.18. *Pearson correlation coefficients by type for test-retest results of total activity.*

Intensity and SOC group		Correlation coefficient
Total	SOC 1, 2 & 3	0.92**
	SOC 4 & 5	0.9**
High	SOC 1, 2 & 3	0.87**
	SOC 4 & 5	0.87**
Moderate	SOC 1, 2 & 3	0.91**
	SOC 4 & 5	0.9**

**$p<0.01$ (2-tailed).

Intraclass correlations coefficients

A non-significant F-value in the analysis of variance for each of the intensities of physical activity shows that the reported test-retest minutes are very consistent. The average measure intraclass correlation coefficient is acceptable for each SOC group (one-way random effect model). Acceptable coefficients are above 0.7 (Vincent, 1999). Each group has demonstrated similarly high reliability scores for each intensity. These results are presented in Table 5. 19. below.

Table 5.19. *ICC correlations for different SOC by intensity for test-retest results.*

Intensity and SOC group		F-value	P	Degrees of freedom	Correlation
Total	SOC 1, 2 & 3	0.41	0.52	34	0.96
	SOC 4 & 5	0.76	0.34	37	0.95
High	SOC 1, 2 & 3	0.8	0.4	34	0.93
	SOC 4 & 5	0.06	0.82	37	0.94
Moderate	SOC 1, 2 & 3	0.1	0.7	34	0.95
	SOC 4 & 5	0.73	0.4	37	0.93

Note. SOC= Stage of Change

Hypothesis Five

This stated that PARQ will be reliable for participants at all SOC but show a higher reliability score for the non-regularly active SOC group (SOC 1-3), because of lower reported activity levels. Results from this study indicate that PARQ is equally reliable for respondents that are not regularly active as for those that are regularly active. Pearson's and intraclass correlation coefficients were similar for both groups for all intensities and the coefficient of repeatability (R), which is twice the standard deviation of the mean difference, was slightly higher for the regularly active group (255 minutes) than the not regularly active group (187 minutes).

5.5. Discussion

This study provided a means of comprehensively assessing the reliability properties of the newly designed PARQ questionnaire. While its main aim was to assess the overall test-retest reliability, this study also provided novel information with regard to particular cognitive aspects of designing a questionnaire. Specifically, the unique approach of using a personal calendar in a

self-administered recall physical activity questionnaire and also the impact of the question order effect were explored.

5.5.1. Personal calendar

This study reports on novel research that compared the reliability of PARQ when a personal calendar (PC) was included as a recall aid, against when it was not. Results indicate that the questionnaire with a PC produces more reliable retrospective reports in comparison to the questionnaire without a PC. The superior performance of the PC questionnaire was evident in all three reliability coefficients (Pearson's, ICC and Repeatability) and also across all intensities of physical activity (moderate intensity, high intensity and total activity). A limitation of this study was the significant difference in the age profile of the participants. However, an independent t-test indicated that there was no significant difference between the reported activity between both groups for total, high or moderate intensity activity and the age profile for both studies was relevant to the criteria for the study,(i.e., 18-55 years). The novel aspect of using a PC in this study, means that comparison of these results to other PA questionnaires is not possible. These results do however compare to Belli and colleagues (2001) social and economic behaviour questionnaire study, which found that the presence of an Event History Calendar in the questionnaire helped respondents to recall more accurate answers, compared to when the calendar was not part of the questionnaire.

An explanation for why the presence of the PC would improve recall is provided by Belli (1998), Shum (1998) and Conway (2001). These explanations seem applicable to the layout of PARQ. Firstly, Conway's (1996) model suggests that personal goals play a major role in the formation, access and construction of specific memories, i.e., autobiographical memory (ABM) is a sense or experience of the self in the past and is induced by images, feelings and other memory details that have been stored as important to the self at some time in the past (Conway, 2001). The personal landmark events recalled by the participants are effective because of their personal nature, which reflects the personal knowledge of ABM. The personal importance attached to events was found by Rubin & Kozin (1984) to be important in making information stand out in memory, compared to events that were not personally salient. The use of landmarks appears to also assist in event-dating by contextualising events in terms of what happened and when it happened. By helping to date events inside or outside the reference timeframe (in the

case of this study, the past seven days), the PC has reduced the effect of telescoping and enhanced the accuracy of the information recalled.

Secondly, landmarks help the PC to make use of both thematic and temporal information which is theorized to be an integral aspect of the structure of ABM (Conway, 1996). The hierarchical structure of ABM permits retrieval through many interconnected pathways that can be activated in a top-down direction (i.e., from Lifetime themes to General Events to Specific Knowledge), or in sequential and parallel directions across themes of events (i.e., while thinking about picking kids up from school, remembered the car had a puncture so had to walk to and from the school). In the case of this study, the recall timeframe was short (seven days) and so was not relevant to Lifetime Periods. However, the multiple personally relevant cues that the PC provides to the respondent, (in terms of details about when, what, who), can help the respondent access information about general events (GE). This level of memory constitutes a personal summary of general actions and make up the most frequent type of AB knowledge (Conway & Rubin, 1993). The cues that are used to index associated GE, are in turn used to provide cues that index the more detailed information at the event specific knowledge (ESK) level (Conway, 1996). Thus, these cues reflected the structure of the ABM by accessing information in many directions, using both thematic and temporal indices.

The results of this current study suggest that the presence of the PC has facilitated the recall of better quality retrospective information because the information it offers, reflects the structure of the storage of information in ABM. More specifically, it suggests that the information and cues provided to the respondent, are personally relevant and are varied in terms of themes and time-related indices. These personally salient, multiple cues are used to access the pathways of ABM. The influence of multiple cues enhancing recall are supported by findings from Wagenaar's (1986) diary study, which stated that "the probability of recall depended upon the number of retrieval cues, as well as on the nature and the particular combination of these cues", (p.249). Implications for further PA recall questionnaire's are that a recall aid, such as the personal calendar included in this study, should be incorporated into the design of the questionnaire, in order to facilitate the retrieval of better quality information. The remaining sections discuss the results of part two of the study, i.e., the PARQ that included the PC.

5.5.2. Total physical activity

The significant Pearson's correlation coefficient for total PA (r=0.9), provides a very strong indicator of acceptable reliability and is favorably comparable to other contemporary PA recall questionnaires that are considered reliable. Included among these are the International PA Questionnaire (IPAQ) with an r=0.8 (Craig et al., 2003), the Lifetime Total PA Questionnaire with r=0.74 (Friedenreich, Courneya and Bryant, 1998) and the Stanford PA Recall Questionnaire with r=0.84 (Blair, Haskell and Ho, 1985). Direct comparison of the results of this present study with other questionnaires is however difficult, due to differences in methodology as outlined earlier. These differences include diversity in sample, variations in recall period of interest, categories and intensity of physical activity being assessed, as well as the disparities in times between test and retest.

Acceptable reliability was also demonstrated in the intraclass (ICC) correlation result (0.95). This form of correlation has been argued to be a more accurate reflection of test-retest reliability than interclass (Pearson's) correlation, because it more appropriately analyses the mean differences of repeated measures on the same variable (Vincent, 1999). Pearson's correlation coefficient is comparatively less suitable for determining reliability because it is a bivariate statistic, whereas test-retest reliability involves a series of univariate statistics (Thomas & Nelson, 2001). Also, unlike the interclass coefficient, the ICC is sensitive to changes in the magnitude and order of scores between each individual's own test and retest. In this way, the ICC provides a thorough examination of different sources of variability, such as changes in individual respondent's means and standard deviations. This form of reliability analysis has not been commonly applied to PA questionnaires (Patterson, 2000), therefore comparison of the results of this study to other PA studies, is limited. Ainsworth and colleagues (2000) assessed the reliability of the Kaiser PA Survey for women with the ICC, with a total activity score of 0.83. Results from PARQ (0.95) are however not directly comparable to the Kaiser Survey, because of differences in recall period (past seven days for PARQ and past year for Kaiser) and test-retest timeframe (overlapping four days for PARQ and one month for Kaiser). The suitability of ICC to assess repeated measures, suggests that future reliability assessment of new PA questionnaires should adapt this form of analysis instead of interclass correlations.

An alternative measure of reliability of the same data, i.e., the coefficient of repeatability (R), revealed a possible variation between two administrations of the same questionnaire of up to

227 minutes for total physical activity. This indicates that while the two measurements are highly related (from the ICC and Pearson coefficient results), they are not necessarily in total agreement. Plots of the difference against the average test-retest results, revealed that agreement between the two questionnaires decreased as the magnitude of data reported increased. This was also a finding in the Scottish Physical Activity Questionnaire (SPAQ), the only other physical activity questionnaire to report a coefficient of repeatability in minutes of activity (Lowther et al., 1999). However, the different design and question content of the SPAQ to the PARQ means that their respective R's are not directly comparable.

In the context of this current study, the potential variation of 227 minutes is not necessarily an indication that PARQ is not reliable. Given the high interclass and ICC correlation coefficients reported in this study, PARQ may in fact be considered more reliable than other PA questionnaires (referred to above), that reported lower correlation coefficients. An interesting observation would be the comparison of the coefficient of reliability (R) of this study against those questionnaires that reported lower interclass and ICC correlations than PARQ. The design of the test-retest methodology engaged in this study (four day overlap) and the relative stability of PA, suggests that variations identified during analysis are caused by PARQ and not actual variations in levels of PA. The results obtained are therefore considered to be a reflection of the reliability of the new questionnaire. However, Conway (1996) suggests that ABM is inherently unstable across repeated retrievals of the same memory. Therefore results obtained may in fact be a reflection of the efficiency of memory retrieval processes and not solely an indication of the psychometric properties of the questionnaire.

Conway (1996) suggests that instability across repeated retrievals occurs for two reasons. Firstly, memory is subject to error because of the natural decay associated with time (Ebbinghaus, 1884). In this current study, time two (retest) had a longer retention interval between the time the activities occurred and were recalled, in comparison to time one (test). The total activity result for time two may therefore have been affected by a form of recall bias, because the four overlapping days included in the analysis were immediately previous to the first data collection date, but were between three and seven days away from the second data collection date. Distortion of memory with time is a finding from other studies, e.g., Subar et al.'s (1995) dietary study showed that participants did not recall specific episodes but resorted to general information as the retention interval increased. Wagenaar's (1986) and Linton's (1982)

diary studies also showed that memories lost their distinctiveness over time and can often blend with similar memories, making it more difficult to retrieve episodic information and increasing the likelihood that what is remembered is inaccurate.

Secondly, the constraints of the task or situation (at time two) may have changed since the information was first retrieved (at time one). The retrieval of the exact same memory at time two may be affected by changes to the self, the respondent developing a changed understanding of the task as well as other factors such as motivation, fatigue, time of day and environmental distractions (Conway, 1996). In terms of test-retest reliability studies, Conway's theory of ABM structure and access to information within, suggests that researchers need to pay particular attention to methodologies employed. Particular consideration should be given to the environment, instructions and timeframe of the test-retest. Conway's model also highlights the need to interpret reliability scores carefully and within the context of the sample and conditions employed. The results of this study indicate that PARQ is a reliable questionnaire for measuring total PA participation. The results are strengthened because of the particular methodology used, i.e., (i) a four-day overlap to remove the possibility of actual variation in PA, (ii) a diverse sample in terms of age, gender, physical activity SOC and employment status, and (iii) consistency in completing the retest in the same place as the test, with the same instructions.

Total physical activity is composed of (i) high intensity and (ii) moderate intensity questions. The relative contribution of each of these intensity questions to the overall reliability of total PA is now discussed.

5.5.3. High and moderate intensity

In this study, the difference between the Pearson's correlation coefficients for high intensity (r=0.9) and moderate intensity (r=0.88), is minimal compared to results from other PA recall questionnaires. The Seven-day Physical Activity Recall Questionnaire (Sallis et al., 1985) reported correlations of r=0.08 for moderate, r=0.31 for hard and r=0.61 for very hard activity, while the Godin Leisure-time Exercise Questionnaire (Godin & Shephard, 1985) reported r=0.46 for moderate and r=0.94 for strenuous activity. The International Physical Activity Questionnaire (IPAQ) does not specify correlations, but states "...the repeated recall of vigorous physical activities were generally better than moderate physical activities", (Craig et al., 2003, p. 1385). The accurate measurement of moderate intensity activity has been identified by researchers as a

challenge (Ainsworth, 2000; Patterson, 2000). Poor reliability coefficients for moderate intensity, are thought to have occurred in questionnaires to date because of (i) the difficulty in capturing the wide variety of activities in which it can be undertaken, (ii) the high frequency of its occurrence during the day (and thus the associated burden of recalling it), and (iii) because questionnaire items have, prior to the 1995 ACSM/CDC health related minimum participation guidelines (Pate et al., 1995), borne little resemblance to the moderate intensity activities that are part of the respondent's life (Patterson, 2000; Ainsworth, 2000).

The Pearson correlation coefficients from this study suggest that the design of PARQ has assisted the participants to overcome some of the difficulties in accurately recalling moderate intensity activity. This theory is further supported by the equal ICC obtained for both high and moderate intensity activity (0.94). Indications from these results are that PARQ has demonstrated strong reliability for measuring high and moderate intensity physical activity, as results above 0.7 are deemed to be acceptable for behavioural sciences (Vincent, 1999).

The coefficient of repeatability (R) of this present study has however revealed a different trend. The extent of agreement between time one and time two is less for moderate intensity than high intensity. The R for moderate intensity shows a possible variation of up to 205 minutes, compared to 95 minutes for high intensity, in further tests under the same conditions. It is evident from these results that the variations in total PA, i.e., R=227 minutes as described above, are mainly caused by variations in moderate intensity activity. Better agreement in the high intensity data may have occurred, because the nature of high intensity activity lends itself to exhibiting many of the characteristics of accurate memories, such as (i) infrequency, (ii) saliency and (iii) distinctiveness (Tourangeau et al., 2000; Jobe, Tourangeau & Smith, 1993).

(i) Infrequency - it is possible that high intensity activity was better remembered than its more moderate counterpart, because of its comparatively lower total reported minutes (Table 5.2). An explanation for the comparative ease of retrieval for high intensity activities over moderate intensity can be explained by Conway's (1993, 1996) structure of ABM. The provision of multiple cues and a lack of interference from numerous experiences of the same type, facilitated easier access to specific information at the Event Specific Knowledge (ESK) level through the General Event (GE) level (Conway and Rubin, 1993). As the number of similar activities increased for the more frequently performed moderate intensity activity, respondents

supplemented fragments of memories for specific events (ESK) with general information about the type of event from the GE level (Conway, 2001).

Findings from study two (cognitive interviews, Chapter Four) indicated that moderate intensity tended to evoke a rate based retrieval strategy instead of a more accurate individual enumeration technique. The accuracy of moderate intensity in this study was possibly negatively affected by the volume of activity retrieved which was greater than high intensity. This compares with results from other researchers (Blair & Burton, 1987; Burton & Blair, 1991; Linton, 1982; Wagenaar, 1986; Means & Loftus, 1991) that showed that the more events that participants had to report, the less likely they were to enumerate them, with enumeration techniques being almost completely abandoned when more than ten events were reported. In this study, it appears to have resulted in an accuracy-effort trade off, where the volume of information to be recalled for moderate intensity became too burdensome for the participant, possibly encouraging them to rate base their answers instead of enumerating.

(ii) Saliency and (iii) Distinctiveness – Conway's (1993, 1996) model of ABM suggests that personal importance and affect are central components of memory formation, so that when an event is judged to be highly self-relevant, then encoding processes co-ordinate to rapidly integrate an event and its memory details into long-term memory. In this current study, the distinctiveness and saliency of high intensity appears to have been further enhanced by the more extreme physical and psychological responses (affect) it elicited from the participant (as detailed by focus group participants in Chapter Three). The extreme emotionality associated with this intensity provided a more elaborate initial encoding at the time of its occurrence, making it more distinctive and also a better likelihood for accurate recall. This is similar to Wagenaar's (1986) diary study which found that more intense and pleasant memories were easier to recall.

Findings from this study do suggest that PARQ is a reliable instrument for reporting participation in high and moderate intensity physical activity. The high variability of reported moderate intensity activity from test to retest appears to be related to the higher frequency of involvement in activity at this level. Examination of the individual types of activity has indicated that some types are performed more frequently at a moderate intensity level and are therefore influential in affecting the overall level of agreement. These results are discussed below.

5.5.4. Types of activity

A closer examination of the composite categories of moderate and high intensity (and therefore total activity), revealed that particular categories provided less reliable data than others. Occupational activity had the lowest Pearson's (r=0.7) and ICC (0.79) correlations, though still demonstrated high reliability (Vincent, 1999). The ICC for family activity was very high at 0.97. These ICC results are in contrast to results from the Kaiser Physical Activity Survey for women (Ainsworth et al., 2000). This questionnaire had similar types of activity to PARQ, e.g., occupational, caregiving, sport and exercise and housework. Results demonstrated an ICC of 0.01 for family caregiving activity and occupational activity had the highest ICC of 0.85. Direct comparison of results is not possible because of the differences in methodologies between the two reliability studies, as described earlier. However, all of the correlation results in this current study demonstrates that each type of activity is reliable.

In assessing hypothesis four, i.e. that the reliability score will be better for the type of activity that is performed less frequently than the others, sport and exercise activity was compared against family activity. Table 5.10 illustrated that the volume of information recalled was highest for sport and exercise and lowest for family, for both test and retest. A further breakdown of these two categories by intensity revealed that both totals were predominantly composed of moderate intensity activity. As demonstrated in section 5.5.3., retrieval of moderate intensity activity represents a relative cognitive burden on the autobiographical memory structure, compared to high intensity. The spread of data away from the mean as the magnitude of recalled activity increased, showed that the volume of moderate intensity activity impacted more extensively on the total reliability (level of agreement) of sport and exercise than family activity (Figures 5.3. and 5.4.). An absence of similar analysis procedures in other questionnaires means it is not possible to extensively compare this trend for specific categories of physical activity. The authors of SPAQ (Lowther et al., 1999) did find a difference in repeatability between occupational walking and other types of occupational activity which was based on magnitude of reported data. The SPAQ does not however discriminate between intensities and therefore the contribution by moderate intensity specifically can not be evaluated. This current study shows that certain categories, by their nature, are more likely than others to elicit higher volumes of information related to moderate intensity. Consequently, these categories will be

more susceptible to the effects of the cognitive burden associated with moderate intensity and may be prone to producing less accurate and reliable information.

The coefficient of repeatability is advocated by Bland and Altman (1985) as an alternative to the Pearson correlation coefficient, for assessing the reliability of an instrument. They argue that items which are highly related to each other are not necessarily in total agreement. The results of this study are in contrast to this point as occupational activity, which had displayed the lowest correlation coefficients, demonstrated the second best repeatability coefficient (R=71 minutes). Only family activity showed consistency by displaying the best level of agreement between questionnaires (R=43 minutes) and the highest correlations. Bland and Altman propose the use of the Coefficient of Repeatability because Pearson's does not account for changes in individual means and standard deviations, from test to retest. The ICC is sensitive to these changes and accounts for differences in magnitude and order. Therefore in this study, the ICC may be the best indicator of reliability in repeated measures of this questionnaire, as it addresses the concerns of Bland and Altman about Pearson's correlation by accounting for magnitude and order, but has not demonstrated the inconsistency of R, as detailed above with occupational activity.

5.5.4. Question-order effect

The results of this novel study rejected hypothesis three, which stated that assessment of the question order effect will show that the group beginning with a high intensity question on both test and retest will have better reliability scores than those beginning with a moderate intensity question on both test and retest. The results of the repeated measures ANOVA demonstrated that there was no significant difference between the four groups (i.e., four different order of questions at test and retest). This indicated that the possibility of a bias caused by familiarity with the questions did not occur. It also meant that there was no advantage to placing the high intensity question before the moderate intensity question, thus rejecting hypothesis three. Using Conway's (1993, 1996) model of ABM as a framework to explain the basis for this hypothesis, access to high intensity activities should be easier, quicker and more accurate because of its more distinctive properties and infrequent occurrence. The thematic and temporal indices that are used to search for high intensity activity first (group one), should then act as a parallel cueing mechanism for the retrieval of moderate intensity activities. In contrast, when the

moderate intensity question came first (group three), the absence of initial cueing from high intensity was expected to have negatively affected the result for that group. Because of the larger amount of less salient information to be recalled for moderate intensity, it was expected that activities would be less distinctive and thus harder to isolate in memory. This would result in participants supplementing their memories for specific events with general information about that event, leading to the retrieval of less accurate information.

In this study, there was no significant advantage observed in placing the high intensity question before the moderate intensity question. This is in contrast to the theoretical explanations of ABM structure, as detailed above by Conway (1993, 1996). The high correlations achieved for moderate intensity activity in this study, may have negated the advantage that high intensity was expected to have. The reasons for moderate intensity's high reliability coefficients are perhaps related to the presence of intensity specific recall cues. Anderson and Conway (1993) state that distinctive memories are associated with sensory information, affect and vivid images. The multiple cues present in PARQ provide this type of information to the respondent, through descriptions of intensity-specific physiological and psychological responses and examples of activity. All of these cues and descriptors were obtained from discussions with the target population (studies one and two), thus making them relevant to the respondent's lifestyle. The additional presence of the personal calendar has also been shown to enhance reliability (Hypothesis One), with the provision of multiple self-relevant cues. The combination of all of these cues may have provided more extensive and accurate probing of the pathways into ABM, resulting in access to the event specific knowledge (ESK) level of ABM and the retrieval of more accurate information about moderate intensity.

5.5.5. Stages of exercise behaviour

Hypothesis five was concerned with whether the reliability of the PARQ would be affected by the participant's stage of change, in particular that those in the action and maintenance stages would demonstrate less reliable data. The basis for this hypothesis was that those participants that are not regularly active (NRA) (Stages 1-3) would have less information to recall and would thus be more consistently accurate in retrieving the comparatively lower amount of data. Results indicated however that PARQ is equally reliable for regularly active (RA) and non-regularly active individuals. Pearson and intraclass correlation coefficients for

total, high and moderate intensity activity were comparable for both sets of individuals. Examination of the descriptive statistics (Table 5.15) suggests that this equality in results is related to the fact that both groups reported similar amounts of moderate intensity activity. This is theorized to be the more difficult of the two intensities to recall data because of its higher frequency and less distinctive characteristics (Durante & Ainsworth, 1996). Therefore, as both groups recalled the same amount of this type of activity, the reliability scores appear to reflect the equality of effort and success in recalling this information. It was expected that the regularly active group would have more high intensity activity than the non-regularly active participants. The regularly active group did recall nearly double the amount of high intensity activity than the non-regularly actives. This type of information is however theorized to be easier to recall because of its more salient and distinctive features, and therefore the increased volume did not necessarily burden the cognition of the individuals. A limitation of this analysis was the collapsing of the five groups into two categories (NRA and RA) instead of analyzing the five stages of change individually. This was not possible in this study because of insufficient pre-contemplator individuals. The collapsed categories also represented the outcome of interest in this study (i.e. whether individuals were reaching the minimum participation guidelines for regular activity). Further research could more closely examine the reliability of each individual stage.

5.5.5. Limitations of this study and future research

The random assignment of participants to one of four groups was novel and was employed to minimize the possibility of a (question) familiarity bias being a factor in the overall reliability result and to assess the question-order effect. This study did not include an analysis of the order-effect of each individual type of activity and further studies could investigate this for activities within each intensity. This would be a novel research topic and may provide valuable insight into other methods of enhancing recall of physical activity.

The use of models of the structure of ABM and of the question-answering process provided a novel approach to the study of the recall of physical activity and also a means of interpreting the results of these studies. However, the lack of comparable results from the physical activity domain means that it was not possible to substantiate findings or conclusions from this research. The employment of three forms of statistical analysis also represents an

unusual approach and enabled a more comprehensive understanding of the reliability of PARQ, although the relative scarcity of comparable data is a limitation.

A further limitation of this study was the design of the study which compared the use of the questionnaire with the personal calendar against without the personal calendar. An improved design would be to randomly assign participants to one of four groups (in a similar design to that used for the question-order effect study). In this design, participants would be in one of the following groups:

Group	Time one	Time Two
1	Personal calendar	Personal calendar
2	No personal calendar	No personal calendar
3	Personal calendar	No personal calendar
4	No personal calendar	Personal calendar

A repeated measures ANOVA could then be used to assess whether the personal calendar had a significant effect on the recall and reliability of information.

The design of PARQ, using multiple cues and a purpose-built calendar layout, was intended to optimally assist the participant to access specific memories, particularly the more difficult moderate intensity activities. Results in this study showed that for specific types of physical activity, e.g., sport and exercise, the recall of moderate intensity activity is particularly affected. This may be because of the higher amount of participation time, in comparison to other types of activities, or perhaps there are other factors that inhibit the recall of information for this intensity of activity. For this reason, further research should develop a model of the individual features of moderate intensity. Given that the central tenet of ABM is the perception of the self when storing information, particular attention should be focused upon the relationship of the properties of moderate intensity and the association with the perceived self and the attainment of personal goals. Finally, the diversity in the sample populations across many variables (physical activity SOC, age, employment status and sex) means that the results from these studies could be considered generally transferable to wider populations.

CHAPTER 6 - ASSESSING THE VALIDITY OF PARQ.

6.1 Introduction

The previous chapter investigated and established one aspect of the psychometric properties of Physical Activity Recall Questionnaire (PARQ), i.e., reliability. Following construction of a questionnaire, the validity or "...the correctness of a test instrument in measuring what it is designed to measure" (Vincent, 1999, p.2), also needs to be determined. Several types of validity can assess the performance of a questionnaire instrument. (i) The *face validity* of PARQ was assessed, following the design of the draft version that resulted from the focus groups in study one. This is an informal assessment of the questionnaire by individuals that have no formal knowledge of the questionnaire topic or design process. This type of validity check is a casual approach that looks at whether members of the target population can read and understand the questions. This was undertaken prior to the cognitive interview study.

(ii) The *content validity* of PARQ was assessed as part of the cognitive interview process in study two (Chapter Three). Content validity was assessed by an expert review panel who have some professional knowledge of the subject matter and questionnaire use. It is a subjective form of validity checking but offers a more formal method than face validity, of assessing how appropriate the items on the questionnaire are. This process is described in more detail in Chapter Three.

(iii) To assess whether the questionnaire is accurate in obtaining information relevant to the construct of interest, it needs to be validated against another piece of equipment that is acknowledged to be a relative "gold standard" for assessing the same variable. This form of validity is known as *criterion concurrent validity* and is assessed by correlating the results of the questionnaire with the scores of the criterion equipment. The direct outcome variable of the questionnaire (PARQ) in this study is "minutes of activity per day over the past seven days", at both moderate and high intensity. This information provides a reference point as to whether the minimum participation levels for health benefits outlined in the ACSM/CDC recommendations and guided by the FITT principles, are being reached (Pate et al., 1995). A limitation to this type of validity is the absence of an acknowledged "gold standard" criterion measurement that can accurately measure different aspects of lifestyle Physical Activity Energy Expenditure (PAEE), i.e., frequency, intensity, type and time (FITT principles).

Criterion concurrent validity has been the method of choice for other physical activity questionnaire studies, e.g., Scottish Physical Activity Questionnaire (SPAQ) (Lowther et al., 1999), International Physical Activity Questionnaire (IPAQ) (Craig et al., 2003), Kaiser Physical Activity Survey (Ainsworth et al., 2000b) and see also Table 2.5, Chapter Two. A fundamental requirement of a validity check is that the criterion equipment has already been validated and is an accepted measure of the variable of interest (Litwin, 1995). PARQ provides some information about each of the FITT principles, therefore the validity of PARQ will be limited to the appropriateness and validity of the criterion measure in also measuring these variables. Methods previously employed to validate PA questionnaires include doubly labelled water, motion sensors such as accelerometers and metabolic indicators such as VO_2 (this is a measurement of the participants rate of oxygen consumption).

Doubly Labelled Water

The doubly labelled water (DLW) technique has been considered a near gold standard for assessing the energy expended in physical activity under both laboratory and free-living conditions (Lamonte et al., 2001). DLW is based on the rate of metabolic carbon dioxide (VCO_2) production, which is tracked in the urinary output of the participant by mass spectrometry, following oral administration of a liquid containing stable isotopes. The VCO_2 is tracked over a few days and total energy expenditure (TEE) is extrapolated from the VCO_2 data, using established equations (Montoye et al., 1986). This technique allows PAEE to be estimated following the subtraction of the previously calculated resting metabolic rate (RMR) and thermal effect of food (TEF) from the resultant TEE.

This technique has been used as the criterion measure to evaluate the accuracy of other instruments assessment of physical activity energy expenditure (PAEE), e.g., Heart-rate monitors (Davidson et al., 1997; Moon & Butte, 1996), rate of oxygen consumption (VO_2) (Moon et al., 1996). Despite its accuracy in measuring PAEE, DLW is not the measurement tool of choice in many epidemiological studies because of its limited feasibility in terms of cost and practicability. This has led in many instances to limitations in validation procedures (Starling, Toth, Carpenter, Matthews, & Poehlman, 1998). Validation studies also show that DLW underestimates individual PAEE by 9% (Speakman, 2001). A further limitation of DLW is its inability to gather information regarding the type, frequency, duration and intensity of the activity being engaged

in. DLW will provide information about the total amount of physical activity engaged in, but is an inappropriate method for behaviour pattern surveillance.

Rate of Oxygen Consumption (VO_2)

VO_2 is the criterion measure for cardiorespiratory fitness (ACSM, 2000) and is measured following an exercise protocol on a treadmill or cycle ergometer. A significant change in the marker of cardiorespiratory fitness (i.e., rate of oxygen consumption) would not be expected within the space of one week (Eston, Rowlands, & Ingledew, 1998; ACSM, 2000), thus VO_2 is an inappropriate validation criterion measure for a questionnaire that requests the amount of PA participated in during the previous seven days. The use of VO_2 to validate PA questionnaires has shown poor correlations in other validity studies, e.g., Lipid Research Clinics Questionnaire, r=0.29 (Ainsworth et al., 1993b), Kaiser Physical Activity Survey (occupational activity, r=0.04, caregiving activity, r=0.01) (Ainsworth et al., 2000b) and see also Table 2.5, Chapter Two. VO_2 does not capture data pertaining to the FITT principles, making it an unsuitable validation method for this questionnaire.

Accelerometers

Accelerometers are motion sensors that have the ability to detect the magnitude of triaxial (three-directional) body displacement. This enables an indication of the energy cost required to overcome ground forces, in order to move the body through space. They are usually the size of a matchbox and are worn on the waistband, at the hip. The output data of accelerometers are in units of (i) vector magnitudes (VM), which are the sum of the forces exerted in each of the three spatial dimensions, (ii) energy expenditure (kcals) and (iii) time (minutes). Information about three aspects of FITT (i.e., frequency, time and intensity) can be obtained from the accelerometer data, once an individualised regression equation has been calculated from a VM-VO_2 calibration curve. This is calculated following a submaximal treadmill testing the laboratory and gives person-specific information about the range of VM counts relevant to moderate and high intensity physical activity. From this information, the researcher can gauge when a participant has been active at a moderate or high intensity level, for how long and for how often.

A limitation of accelerometers is their tendency to underestimate PAEE in field studies for particular activities. This occurs because of the accelerometers limited ability to detect

changes in gradient and velocity. Therefore the increased energy expenditure associated with climbing stairs or an incline, or sprinting, is not accurately accounted for (Jakicic et al., 1999; Nichols et al., 1999; Welk et al., 1995). Poor detection of changes in energy expenditure is similarly acknowledged for activities that are non-weight bearing, such as cycling, because of the accelerometers inability to detect that the participant is no longer supporting their own weight (Welk et al., 1995). Accelerometers have been used in validation studies of other PA questionnaires, e.g., SPAQ, r=0.34 (Lowther et al., 1999), Godin Leisure-time questionnaire, r=0.36 (Godin et al., 1985), see also Table 2.5, Chapter Two. The user-friendly physical dimensions and the ability to objectively record movement, facilitates the accelerometer as a practical method of measuring PAEE in the field.

Heart Rate Monitors

Heart rate monitors have been used to estimate PAEE, based upon the assumption that there is a strong linear relationship between heart rate (HR) and VO_2 (Wareham, Hennings, Prentice, & Day, 1997). Therefore, as the amount of oxygen consumed increases during activity, a corresponding linear increase in heart rate will be observed. A limitation of using HR monitors to measure participation in physical activity is that HR is influenced by extraneous factors such as environmental conditions, age, gender, individual fitness level, hydration status as well as type and intensity of activity (Strath et al., 2000). For this reason, elevations in heart rate captured by the monitor in a field situation, may not be due to PA. A further limitation of HR is its variability in estimating PAEE at lower intensity activity, i.e. HR does not increase linearly at low intensity. This includes activity that requires <3 METs of energy expenditure (ACSM, 2000) and can account for errors in PAEE of up to 30% in individuals (Ainslie et al., 2003).

Output from HR monitors provides information about duration, frequency and intensity of activity. Similar to the accelerometer, single and multiple participant HR-VO_2 calibration curves, performed as a submaximal walking treadmill test in the laboratory, offers the most accurate method of predicating PAEE from HR. This provides information that accounts for differences between people because of fitness levels, genetics and gender. These differences manifest themselves in terms of variations in VO_{2max} and HR (Rowlands, Eston, & Ingledew, 1999; Strath et al., 2000). Rowlands and colleagues found individual calibration to be time consuming and costly but more accurate than using a group calibration curve. The results of the

individual calibration curve will provide an individual HR-VO_2 calibration curve that indicates HR ranges for moderate and high intensity activity for that individual (Haskell et al., 1993). Information about time and frequency can then also be obtained. However, the additional contributory factors to elevating HR referred to earlier (i.e., environmental, stress, hydration etc.), mean that its limitations as a single criterion measure are perhaps reflected in the fact that HR monitoring has not been used to validate any of the PA questionnaires detailed in *Table 2.5*. Its value as an objective measurement tool has however been suggested in terms of composing one part of a combined instrument (Ainslie et al., 2003). This is outlined below.

Combined methods: Accelerometer and Heart Rate

Both accelerometer and heart rate monitors record information about PAEE that incorporates three of the four FITT principles (it does not include "type") and thus suggests reasonable appropriateness as a criterion measure for PARQ validation. The limitations of both heart rate monitors and accelerometers have been outlined above. To overcome these limitations, it has been proposed that heart rate monitoring and accelerometers could be successfully employed as an integrated, simultaneous system of measuring PAEE (Healey, 2000; Strath et al., 2000). This combined feature has previously been used in other studies for the purpose of directly assessing energy expenditure (Moon et al., 1996; Haskell et al., 1993). A study of arm and leg ergometry tasks in the lab (N=10, Male =40%, mean age=26.5 +/- 2.2 years) showed that the combined heart rate and motion sensor showed a stronger significant relationship with VO_2 (R^2=0.81) than heart rate alone (R^2=0.63) (Strath, Bassett, Thompson, & Swartz, 2002). The combined method also accurately measured time spent in moderate and high intensity activity, while HR overestimated time spent in moderate intensity activity. A similar study by Haskell and colleagues (1993) showed that the simultaneous analysis of HR and accelerometer data predicts energy expenditure better, rather than separately. This study also used lab based leg and arm ergometry tasks and obtained individual calibration curves (N=19, all males, age range 22-64 years). This study found that when motion sensor data was added to HR data for the prediction of O_2 uptake, R^2 was increased from 0.69 to 0.82. A limitation of this study however, was that calibration curves were obtained only for the HR and not the accelerometer data. This meant that the simultaneous analysis included combining information from the accelerometer that was assumed to be when the participant was active. In this way, the accelerometer data did not have

the same participant-specific VO_2-related regression equations, as the HR data and may therefore have contributed some error to the results..

In a review of the techniques used to measure PAEE, Ainslie, Reilly and Westerterp (2003) suggested that the integrated approach allows the accelerometer to act as a back-up measure to verify that the corresponding elevations in heart rate are due to participation in physical activity. While this combination may provide a more precise means of estimating PAEE, it is limited by the unavailability of a validated, integrated HR-Accelerometer unit. This necessitates the use of valid, individual units of equipment that are worn simultaneously. This introduces the possibility of human error because of the manual matching of data between instruments. The impact of human error can be dissipated with the application of rigorous data collection and analysis procedures. Using this combined method to validate PARQ represents a novel method of validating a PA questionnaire. Therefore the validity of PARQ will be an outcome of the appropriateness of the combined HR-Accelerometer method as a gold-standard criterion measure.

Validation of moderate and high intensity physical activity

Validity of moderate intensity questions, using an accelerometer as the criterion measure, have been reported as r=0.23 and r=0.11 for the Minnesota Leisure time Questionnaire (Taylor et al., 1984; Taylor et al., 1978) and Seven-day PA recall questionnaire (Sallis et al., 1985), respectively. High intensity for the same questionnaires was reported as r=0.23 and r=0.6, respectively. Recently developed questionnaires such as IPAQ (Craig et al., 2003) did not report validity for each intensity, while SPAQ's design does not differentiate between high and moderate intensity activity (Lowther et al., 1999).

In summary, the purpose of this study was to determine the criterion concurrent validity of the PARQ against a more objective criterion method. The study was designed in two parts. Part one assessed the concurrent validity of the questionnaire using the accelerometer as the criterion method. Part two assessed the concurrent validity of PARQ with simultaneous combined heart rate and accelerometer data as the criterion method. This use of combined methods, that were worn simultaneously and temporally (time) matched, is a novel approach to validating a PA questionnaire.

6.1.2. Hypotheses

The following hypotheses were formulated with respect to the validity of PARQ:

i. PARQ (minutes of activity) will demonstrate concurrent validity, with the RT3 accelerometer (vector magnitude) as the criterion method.
ii. The strength of the relationship between recalled data (PARQ) and the criterion measure (RT3), will lessen as time between occurrence of activity and time of recall increases.
iii. PARQ (minutes of activity) will demonstrate concurrent validity, with the simultaneous use of temporally-matched heart rate and RT3 accelerometer data (minutes of activity) as the criterion method.
iv. High intensity activity (minutes of activity) will demonstrate better concurrent validity than moderate intensity (minutes of activity), with the simultaneous heart rate and RT3 accelerometer data (minutes of activity) as the criterion method.

6.2. Methods

Upon ethical approval from DCU, 60 participants volunteered for this study. All potential participants were requested to complete a screening form that sought information relating to age, gender, occupation status and physical activity stage of change (SOC). SOC has been previously explained in the literature review chapter. Participants were selected for this study by stratified random sampling (Thomas et al., 2001). Random selection occurred by firstly dividing the volunteers by one specific criteria and then randomly selecting them for participation. In this study, the dividing criterion was physical activity SOC and participants were divided into group one (stages 1-3) or group two (stages 4-5). This was because stages 1-3 represent those that are not yet regularly physically active, in contrast to those in stages 4 and 5.

6.2.1. *Equipment*

Evidence of validity of a questionnaire is limited by the appropriateness and quality of the criterion measure, i.e., the criterion measure used to validate the questionnaire should also have demonstrated acceptable validity. The criterion equipment used in this study were the RT3 triaxial accelerometer (Version 1.1, Stayhealthy Inc., Monrovia, California, U.S.A.). and the wireless Polar heart rate monitor (Polar team system, Port Washington, New York, U.S.A.).

Accelerometer

The Tritrac-R3D has been the triaxial (i.e., 3-Dimensional) accelerometer of choice in PAEE research (Freedson et al., 2000; Hendelman et al., 2000; Eston et al., 1998; Ainslie et al., 2003). This model is now no longer available. A newer version of the original Tritrac-R3D has recently been developed by Stayhealthy Inc. and is referred to as the RT3 research tracker. Validation of this latest model is limited due to its relatively recent emergence in 2001. Peer-reviewed studies on the newer model are pending publication (personal communication with Stayhealthy Inc.). For this reason, information pertaining to triaxial accelerometers in research, with regard to practicality and validity, generally refers to the original Tritrac-R3D model.

The Tritrac-R3D (Hemokinetics Inc., Madison, WI, USA) has been validated and tested against criterion methods such as VO_2, or compared to other accelerometers, in both laboratory (Louie et al., 1999; Rowlands et al., 1999; Nichols et al., 1999) and field-based experiments (Welk et al., 2000; Leenders et al., 2001; Freedson et al., 2000; Hendelman et al., 2000). As the focus of this current research is the measurement of lifestyle, free-living PAEE, these studies are discussed. The Tritrac-R3D was tested against a portable metabolic (VO_2) instrument to measure PAEE at moderate intensity for several recreational and household activities (Hendelman et al., 2000). Participants (N=25, F=60%, age=30-50 years) completed walking, played golf and performed indoor and outdoor household chores. Results indicated a strong relationship between VO_2 (METs) and accelerometer (vector magnitude counts) for walking (r=0.89), with r=0.62 for combined other activities. Hendleman and colleagues study also noted an underestimation of PAEE (Kcals) by 30-50 % in activities other than walking. The validity of the Tritrac-R3D was also assessed during laboratory walking and running by Nicholls et al. (1999). Results (N=60, Male = 50%, mean age= 23.4 +/-2.9 years) showed that Tritrac-R3D consistently over estimated energy expenditure (Kcals) during fast horizontal treadmill walking and was not sensitive to changes in gradient. These studies highlight the limitations of Tritrac-R3D in measuring PAEE in activities that are high intensity (or on a gradient), or involve upper body movements.

Validity studies that compare Tritrac-R3D to other triaxial and uniaxial accelerometers, have shown the Tritrac-R3D to predict VO_2 better than other accelerometers. In Hendleman and colleagues (2000) study referred to above, the Tritrac-R3D performed better than other accelerometers (r=0.59 for combined activities). In another study (N=52, F=60%, mean age=29 years), relative validity of three accelerometers under field and laboratory based conditions

showed the Tritrac-R3D to have less error in individual estimates of PAEE, although the correlations for all three monitors with VO_2 were higher for laboratory treadmill activity (r=0.86) than for free-living lifestyle activities (r=0.7), (Welk et al., 2000). These results indicate limitations of the Tritrac-R3D in validating PARQ if it is comparatively less accurate in capturing field-based information.

Although the RT3 Research Tracker (Stayhealthy Inc., CA, USA) is a relatively new model to have become available, it is described by its manufacturers, Stayhealthy Inc., as being "... a replacement for the older Tritrac-R3D...following research to improve the algorithms used to calculate the calorie expenditure", (www.stayhealthy.com). Furthermore, an in-house comparison between the two devices (conducted by the manufacturers), established that the RT3 is a more researcher and user-friendly device with an improved level of precision and accuracy compared to the Tritrac-R3D. A comparative study of internal mechanisms between the RT3 and Tritrac-R3D by researchers at Stayhealthy Inc. concluded that the results from the RT3 showed better consistency than the Tritrac-R3D. This does not appear to have been objectively verified by other research.

(ii) Heart Rate Monitor

Polar heart rate monitors have previously been used as an objective method of assessing physical activity. Under controlled laboratory conditions during vigorous, continuous exercise, heart rate and VO_2 are closely related and exhibit a linear relationship (Freedson et al., 2000). In an examination of various activities at mixed intensities in a laboratory setting (N=30, Male=55%, mean age= 33.1 +/-12.2 years), the strength of the significant relationship between HR and VO_2 was 0.67. HR significantly overestimated energy expenditure by 11% (Strath et al., 2000). The comparison of HR monitoring to doubly labelled water for estimating PAEE was shown to be over estimated by 6% in a free-living assessment of energy expenditure in men (N=9, mean age = 25-54 years). This study assessed body calorimetry (doubly labelled water) for a 24 hour period and compared it to the average of a 9-day, minute-by-minute heart rate collection period (Davidson et al., 1997). While this extended HR data collection period was perhaps more representative of usual lifestyle, its direct comparison with just one day of criterion measurement means that results are perhaps not conclusive.

6.2.2 *Pilot study: Validation of criterion measurement instruments for this study*

Due to the relative emergence of both the RT3 accelerometer and the Polar team heart rate monitor, a pilot study was conducted in order to test the validity of both of these measurement tools in collecting data relating to lifestyle physical activity in adults between 18-55 years old (Power, Martin, & Woods, 2003).

Aim

The aim of this study was to assess the validity of HR and an accelerometer in measuring PAEE against the criterion measure of oxygen consumption, during a range of lifestyle physical activities.

Participants

Eighteen participants (male=22%, mean age=27(+/-10.8) years) took part in this pilot study.

Methods

Equipment

The RT3 triaxial accelerometer (Version 1.1, Stayhealthy Inc., Monrovia, California, U.S.A.) and the wireless Polar heart rate monitor (Polar team system, Port Washington, New York, U.S.A.), were validated with a portable VO_2 metabolic analyser (Oxycon Mobile, Version 4.6. Jaeger, Hoechberg, Holland). This is a lightweight testing unit that was secured to the upper chest by a series of straps. Participants breathed through a facemask from which samples of expired air were drawn and breath-by-breath measures were analysed for expired oxygen.

Procedure

Upon ethical approval from DCU, the participants completed six tasks, each lasting five minutes, which were chosen to represent lifestyle PA and be inclusive of activities that involve the upper body and inclined gradients. These were reading, walking while carrying a 5kg box, cleaning windows, vacuuming, walking and stair climbing. The inclusion of reading was to enable an assessment of whether the accelerometer and HR monitor could accurately reflect a sedentary activity. Although HR was monitored to ensure it was elevated during the activity, the participant performed each activity at a self-perceived level of moderate intensity. This was

deemed necessary to make the study relevant to PARQ, where respondents are asked to recall activities that they perceive occurred at a moderate intensity. Activities were separated by a rest period, during which participants sat and allowed their heart rates to return to resting values. This pilot study was completed in the laboratories of the Centre for Sport Science and Health at Dublin City University over a four week data collection period.

Data Analysis

All data was entered into SPSS (Version 10) for analysis. Simple linear regression equations were used to predict VO_2 from each equipment measure and a multiple regression equation predicted VO_2 for the paired equipment. Finally Pearson correlation coefficients were calculated to assess the concurrent validity of the accelerometer and HR monitor with VO_2.

Results

Descriptive statistics for this study were N=18, Male=60%, age 28.4+/-12.2 years. Minute-by-minute analysis (of five activities minus reading) showed significant correlations with VO_2 of r=0.65 r=0.43 and r=0.56, for net HR (i.e., activity HR-resting HR), HR and RT3 accelerometer (vector magnitudes), respectively. The correlation of accelerometer vector magnitudes with VO_2 (r=0.56) were stronger than accelerometer kcals and VO_2 (r=0.53) for the combined five activities (minus reading) and for the six activities (including reading) r=0.67 (kcals) and r=0.69 (vector magnitudes). The correlation of net HR to VO_2 increased from r=0.65 to r=0.78 for all six activities (including "reading"). Multiple regression analysis was employed to establish the usefulness of applying two measures simultaneously to measure a construct. The results showed that net HR and accelerometer accounted for 66% of the variance of VO_2 ($p<0.01$).

Discussion

The RT3 accelerometer and Polar HR monitor can validly measure activity that is perceived as moderate intensity by the participant, in a free-living context. Both pieces of equipment are also capable of validly measuring sedentary behaviour, as correlations increased for both HR and accelerometer when "reading" was added to the combination of activities. The accelerometer correlations reported in this study are similar to those reported by Hendleman and

colleagues (2000) and Welk and colleagues (2000), for a combination of free-living activities, as discussed above. The heart rate correlation is similar to that achieved by Strath et al. (2002). The R^2 of 0.66 reported for the combined equipment is less than that reported in other studies, i.e., approximately 0.8 (Welk et al., 2000; Strath et al., 2000). The implications from this study are that while these instruments have demonstrated moderate validity in measuring PAEE, it also highlights that there are limitations to the concurrent use of these pieces of equipment as a criterion measure. This factor should be taken into consideration when assessing the validity of PARQ.

6.3. Part one: Concurrent validity of RT3 and PARQ

This study assessed the concurrent validation of PARQ with the RT3 accelerometer. Anthropometrical and demographic data (i.e., age, height, weight, sex and resting heart rate) were obtained for each participant. Verbal and written instructions (see Appendix eleven) were given to the participant on how to apply and look after the equipment for the duration of the study. Two phone numbers were supplied and all participants were asked to contact the researcher at any time, should they encounter any difficulties with the equipment.

6.3.1. Participants

Thirty four participants took part in this part of the study (Female=67%, mean age= 27.14 years (+/-5.49), mean body mass index (BMI)=23.8 (+/-3.67), % SOC at stage 4 or 5 =69%).

6.3.2. Procedure

Each participant's demographic details were entered into the RT3. The RT3 accelerometer was set into mode three, which collected data in the three dimensions at one-minute intervals. To ensure consistency in data collection, all participants were instructed to wear the RT3 accelerometer on their right hip, on a secure waistband, for four consecutive days (two week days and two weekend days). They were asked to wear it from the time they got out of bed until they returned to bed later that night. Each participant was given a time-sheet for the four days and asked to write in the exact time that they put on/took off the equipment every day. The time on the participant's watch was synchronised with the intrinsic time in the RT3 to ensure accurate time-keeping.

On day 5, participants were asked to return the RT3. Data was saved and participants were asked to complete a copy of the PARQ questionnaire, which was estimated to take approximately 20 minutes. All participants were informed that they could withdraw from this study at any stage, without prejudice and that all information received as a consequence of this study would be treated confidentially. All participants were assigned identification numbers and were paid €15 towards travel expenses.

6.3.3. *Data Analysis*

The four-day RT3 accelerometer activity (vector magnitude counts) and activity reported in the PARQ (minutes) were both entered into SPSS (Version 10) for statistical analysis. RT3 accelerometer data was analysed as vector magnitude counts, as these have shown a good relationship with VO_2, in free-living energy conditions (Nichols et al., 1999; Welk et al., 2000; Hendelman et al., 2000). Accelerometer Kcals have been underestimated by 30-50% for activities other than walking (Hendelman et al., 2000) and overestimated for treadmill activities (Nichols et al., 1999). Accelerometer vector magnitude counts have also been used to validate other PA recall questionnaires, e.g., International Physical Activity Questionnaire (IPAQ), (Craig et al., 2003).

On examination of the data, two sets were identified as outliers by SPSS. Both of these participants had reported between 5 and 7 hours of at least moderate intensity on one day of the questionnaire. Inclusion of these data sets caused the distribution to skew and affected an otherwise normal distribution and were therefore removed prior to analysis. A normality test (Shapiro-Wilks) ensured that the assumptions of parametric data were met for both the accelerometer ($W=0.91$, $p=0.51$) and questionnaire data ($W=0.93$, $p=0.6$). The relationship between the accelerometer (vector magnitude counts) and the PARQ (minutes of activity) was assessed with a Pearson correlation coefficient. The Limits of Agreement (LOA) could not be assessed because of the variation in measurement units between the accelerometer (vector magnitudes counts) and PARQ (minutes of physical activity) (Bland et al., 1986).

6.3.4. *Results*

Two RT3 accelerometer data sets were not available for analysis as one participant's accelerometer malfunctioned, while another participant forgot to wear his on the fourth day. Complete PARQ data was available for all participants.

Descriptive statistics

Descriptive statistics for the participant's PARQ and RT3 data are given in Table 6.1.

Table 6.1. *Descriptive statistics for total four day RT3 accelerometer (vector magnitude counts) and PARQ data (minutes)*

Measurement	N	Mean	Median	SD	Min	Max
RT3 (counts)	34	1234261.97	1113270.5	524019.9	553564	2673676
PARQ (minutes)	34	331.88	285	178.88	84	800

Tables 6.2. and 6.3. show descriptive statistics for each of the individual days for both RT3 and PARQ data, providing an indication of the variability of the data between days. Days one, two, three and four correspond to Saturday, Sunday, Monday and Tuesday, respectively. The questionnaire was administered on Wednesday.

Table 6.2. *Descriptive statistics for individual four day RT3 accelerometer (vector magnitude counts)*

RT3 (vector magnitude counts)	N	Mean	Median	SD	Min	Max
Day one (Sat)	34	306904.9	253534.5	166575.3	95285	865862
Day two (Sun)	34	309612.4	270073	153454.1	122416	743866
Day three (Mon)	34	311773.6	286432.6	158636.4	59051	745127
Day four (Tues)	34	305971	257097	246469	27142	1383012

Table 6.3. *Descriptive statistics for individual four day PARQ data (minutes)*

Questionnaire (minutes)	N	Mean	Median	SD	Min	Max
Day one (Sat)	34	97.26	80	67.37	15	280
Day two (Sun)	34	64.53	43.5	56.83	0	250
Day three (Mon)	34	87.5	62.5	77.66	0	325
Day four (Tues)	34	82.59	60	83.28	0	420

Mean activity minutes for day four (PARQ) are the second lowest but have the highest standard deviation, thus indicating high variation in the amount of activity reported between

participants. This day also has the largest difference between minimum and maximum reported activity. A similar pattern is observable for this day in the accelerometer data.

Pearson product moment correlation coefficients

The four-day vector magnitude counts recorded on the RT3 accelerometer were compared to the self-reported minutes of total physical activity in the PARQ. This was undertaken to assess the extent and strength of the linear relationship between these two variables. Initially an analysis of the total four days data for both measurements was done, with $r=0.52$ ($p<0.01$). This was followed by a comparison of weekend data ($r=0.5$) to weekday data ($r=0.64$), both at $p<0.01$. The correlation is higher for the weekdays which were immediately prior to when the questionnaire was completed. Finally, the day-by-day activity was analysed and is outlined in Table 6.4. The correlation is highest for day four, which was also immediately prior to the day on which the questionnaire was completed.

Table 6.4. *Pearson correlation coefficients between accelerometer (vector magnitude counts) and self-reported physical activity (minutes) for each individual day.*

Day 1 (furthest to recall)	Day 2	Day 3	Day 4 (nearest to recall)
0.544**	0.534**	0.496**	0.782**

Note. **p< 0.01 (1 tailed).

Hypothesis One

This stated that PARQ (minutes of activity) will demonstrate acceptable concurrent validity, with the RT3 accelerometer (vector magnitude) as the criterion method. The significant correlation result of $r=0.52$ supports this hypothesis.

Hypothesis Two

This stated that the strength of the relationship between recalled data (PARQ) and the criterion measure (RT3), will lessen as time between occurrence of activity and time of recall increases. The difference between the Pearson's correlations for day one and day four support this hypothesis.

6.4. Part two: Concurrent validity of combined heart rate and accelerometer with PARQ

The design of PARQ incorporates separate questions about the amount of time spent doing moderate and high intensity PA. This information indicates whether the respondent is meeting the recommended minimum participation rates for health-related benefits (Pate et al., 1995). Part one did not permit an individual analysis of the validity of each intensity question. This was facilitated in part two, where the criterion concurrent validity of PARQ was assessed with the simultaneous, combined use of the RT3 accelerometer and Polar HR monitor.

6.4.1. Participants

Twenty seven participants fully completed this part of the study (F=67%, mean age=26.44 years (SD+/- 5.04), mean BMI =24.22 (SD+/- 4.12))

6.4.2. *Procedure*

This part was completed concurrently by the same participants as part one and thus the procedure is as outlined in Sections 6.3 and 6.3.1, with one additional feature. In addition to wearing the RT3 accelerometer for four days, participants were also requested to wear the Polar heart rate monitors for two days (one week day and one weekend day). The heart rate monitors could not be worn for the four days because of its 12-hour memory capacity. HR monitors were returned to the researcher on alternate days (i.e., day 2 and day 4) so that the data could be saved and the memory could be cleared. The data collection period therefore included two days of simultaneous heart rate and accelerometer data.

<u>Sub-maximal treadmill walking test</u>

On their return visit on day five (the day they completed the PARQ), each participant was invited to complete a sub-maximal walking treadmill test in the exercise performance laboratory at the Centre for Sport Science and Health in DCU. The purpose of this test was to obtain individual information about each participant's response to physical activity at moderate and high intensities from their heart rate and accelerometer data. The protocol employed for a submaximal exercise test should consider (i) the purpose of the test, (ii) the specific outcomes desired and (iii) the individual being tested (ACSM, 2000). According to the ACSM's guidelines for exercise testing and prescription, submaximal tests should avoid large and unequal

increments in workload and should have a usual test time of between 8 and 12 minutes. The test is also usually stopped at a predetermined level.

The submaximal test protocol in this study adhered to the ACSM protocol guidelines as it involved walking at three equally incremental speeds (2.5, 3.5 and 4.5 mph, at a 1% gradient) for a total of four minutes at each speed. These speeds were selected in order to reflect a gradual increase in MET's expended while walking, with the 1% gradient intended to reflect external resistance such as wind and uneven terrain. These speeds were chosen to represent a light, moderate and high intensity of activity for the participants. Attainment of each intensity was verified from the participant's VO_2 data, by ensuring that MET values were between 3-5.9 for moderate intensity and >6 for high intensity, as referred to in the compendium of physical activities (Ainsworth et al., 2000). The four-minute duration was selected to ensure that the participant had reached a steady state heart rate at each stage for increasing intensities. Steady state occurs following the initial first minute rapid increase, in response to the initiation of exercise. The heart rate then levels off and remains fairly stable throughout the remainder of the bout of exercise (Wilmore, 2001). Participants were also asked to express their Rate of Perceived Exertion (RPE) every minute, by indicating their selected score from Borg's 20 RPE scale. This ensured that the increase in speed was subjectively, progressively challenging to the participant.

6.4.3. *Data Analysis*

In part two, PARQ was compared against the combined two-day accelerometer and heart rate data, in order to analyse data by intensity. Initially, data from the submaximal walking treadmill test was analysed (see below) to obtain individual calibration equations for each individual. HR and RT3 data were analysed firstly in MS Excel, where missing data was treated (see below) and also where data from the two instruments was temporally matched (see below). Comparison of data from PARQ with data from the combined criterion measures was undertaken in SPSS (Version 10), by means of a Pearson's correlation coefficient and Limits of Agreement (Bland & Altman, 1995; Bland et al., 1986).

Analysis of Submaximal Test Data and Calculation of Individual Regression Equations

The completion of the submaximal treadmill tests enabled an individual calibration of heart rate (HR) and vector magnitude counts (VM) (from the RT3 accelerometer) against METs, as

calculated by the indirect calorimeter (VO_2). An individual regression equation was calculated for each person using MS Excel, for HR against METs and VM against METs. These equations were then used to predict each individual's own HR and VM response to activity at a moderate intensity and a high intensity. The ACSM (2000) guidelines of 3-5.9 METs for moderate and >6 METs for high intensity were used as cut-off points. Figures 6.1 and 6.2 below provide an example of the individual regressions with equations for HR vs. METs and VM vs. METs for one participant (participant number 10).

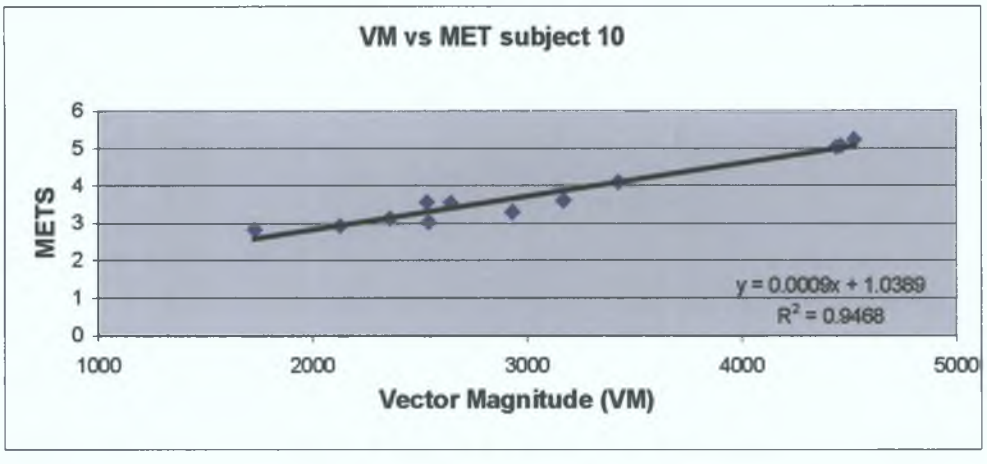

Figure 6.1. Individual regression for Participant No. 10 (METs vs. VM)

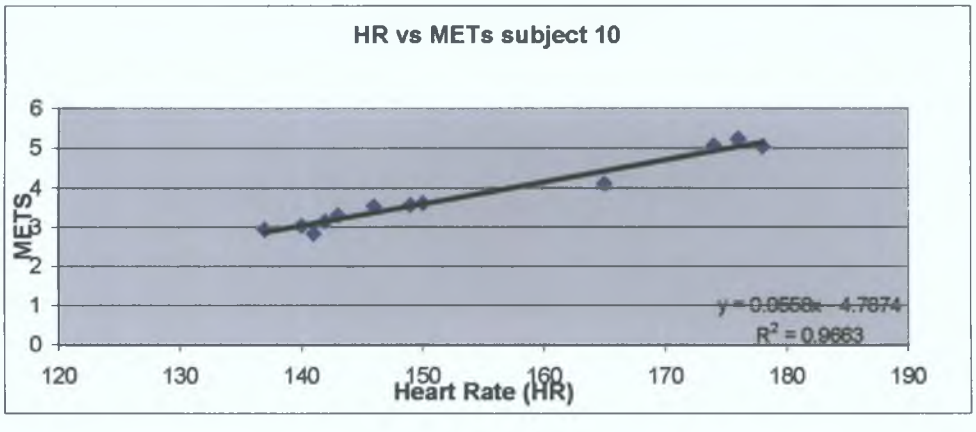

Figure 6.2. Individual regression for Participant No. 10 (METS vs. HR)

The resulting regression equations identified each participant's cut-off points for the HR and RT3 vector magnitude counts, for both moderate and high intensities. All participant's moderate and high intensity accelerometer cut-off points were noted as being within the same ranges identified for these intensities by Nichols and colleagues (1999), (N=20) and Rowlands (personal communication, 2003), (N= 15). This enabled the conversion of the HR and

accelerometer data into "minutes spent in activity", by counting the minutes from the two days of data collection that each participant spent at both moderate and high intensity. This was identified by their individual simultaneous HR and accelerometer range of values, i.e., minutes of activity were included only when the HR and accelerometer data were simultaneously within the range identified for that participant. The conversion of the criterion measure data to minutes, enabled direct comparison with the output data of PARQ (i.e., minutes of activity). This meant that the validity of PARQ could be directly assessed in relation to the specific recommendations for minimum PA participation guidelines for health, decreed by the ACSM/CDC (Pate et al., 1995).

Treatment of missing data

Prior to the analysis of results for part two, missing data received from the HR, RT3 accelerometer and PARQ needed to be identified. There were no incidences of any missing data for the accelerometer or questionnaire. As the HR monitor relies upon constant skin contact with the sensor for the purposes of measuring the person's heart rate, there were a few incidences where contact was momentarily broken. When the data was downloaded from the monitor, it was possible to see for how long this occurred (in intervals of 5 seconds). One data set lost just over 2 minutes, while 6 other data sets lost up to 15-20 seconds either during the day or at the end of the day when the monitor stopped recording in the middle of a minute. In order to have a complete data set for each minute, missing values were substituted with a heart rate value that was calculated by getting the average of the 12 data points immediately preceding the missing value, i.e., the average of the preceding minute (sum of previous 12 data points (in 5 second epochs) / 12).

Reliability of temporally matched data

Specific steps were undertaken to ensure that the simultaneous analysis of the heart rate and accelerometer data in part two was accurate. Initially, pilot work was undertaken to enable the researcher to become familiar with matching the equipment and the data. This was facilitated by four volunteers (Males=50%, age 20-30 years) wearing both instruments for a few hours and performing a range of activities, such as climbing stairs, walking and sitting. From this pilot work, the researcher could identify the specific instructions that needed to be given to the

participant when applying the equipment. For this current study, each participant met with the researcher the day before the data collection period for an hour. During this time, they received instruction about wearing and applying the equipment and instructions on how to complete the time sheet for each day.

Secondly, as two separate pieces of equipment were used, specific steps were undertaken in the analysis phase to ensure that the heart rate data was temporally and reliably matched to the accelerometer data. The RT3 accelerometer collected data in one-minute epochs compared to every five seconds for the HR data. A macro was created in MS Excel to account for an average heart rate for each minute of the two days collection. This data was then temporally matched against the corresponding accelerometer time, for each individual. To ensure that data from the heart rate monitors and the accelerometers were accurately matched for time, an inter-rater reliability test was undertaken. A research assistant randomly picked four participants from the list of identification numbers. He was then asked to randomly choose 30 time points on the accelerometer data for each participant and to identify the corresponding HR for that specific time from the separate HR data. Results were analysed by means of a percentage agreement (98.3%) between the researchers.

6.4.4. *Results*

Thirty one (of 38) sets of HR data were available for analysis. The other seven data sets could not be included because two HR monitors malfunctioned while downloading and four HR monitors collected only one day of data. The remaining HR data belonged to the participant whose accelerometer had malfunctioned and therefore was not analysed because it had no corresponding motion sensor data. Of these 31 participants, only 27 completed the sub-maximal treadmill walking test. Reasons for non-completion of this part of the study were given as personal preference (two), emigration (one) and unavailability (one).

Descriptive statistics

Descriptive statistics for the 27 participants (F=67%, mean age=26.44 years (SD+/- 5.04), mean BMI =24.22 (SD+/- 4.12)), is given in Table 6.7. The results showed that mean PARQ scores reported were nearly three times that of the objectively measured activity (HR/RT3). The extent to which the distribution of PARQ scores differ from their mean, (Standard Deviation

(SD)), is approximately twice that of the HR/RT3. These results provided an indication that the two different methods do not fully agree. The extent to which they do not agree was explored by the Limits of Agreement (LOA) analysis (see below and descriptive statistics in Tables 6.9 and 6.10).

Table 6.7. *Descriptive statistics for total combined HR/RT3 (minutes) and PARQ (minutes)*

Measurement	N	Mean	Median	SD	Min	Max
HR/RT3 (minutes)	27	54.67	47	42.29	0	172
PARQ (minutes)	27	155.48	140	91.08	15	405

Note. SD= Standard Deviation, RT3=Accelerometer data, HR=Heart Rate data.

Descriptive statistics for moderate and high intensity, for both the criterion measure (HR/RT3) and PARQ, are given in Table 6.8. Results indicated a similar trend for both intensities as for total activity above, in terms of comparative mean and SD of PARQ against HR/RT3. The largest variation in distribution of scores from the mean occurred in moderate intensity for both measurement tools. This indicates a comparative difficulty in measuring this intensity of activity for both subjective and objective instruments.

Table 6.8. *Descriptive statistics for total combined HR/RT3 (minutes) and PARQ (minutes) by intensity over two days of data collection.*

Measurement	N	Mean	Median	SD	Min	Max
PARQ moderate intensity (minutes)	27	112.3	85	75.8	10	335
HR/RT3 moderate intensity (minutes)	27	42.6	31	38.16	0	172
PARQ high intensity (minutes)	27	43.2	30	41.17	0	128
HR/RT3 high intensity (minutes)	27	12.07	0	22.64	0	77

Table 6.9 displays the mean and SD of the differences between the two methods and for total activity and moderate and high intensity. This is for the purpose of measuring the Limits of Agreement (LOA) between the two methods, i.e., mean difference (MD) +/- 2SD's (described below).

Table 6.9. *Descriptive statistics of the differences between the HR/RT3 and PARQ (minutes).*

Intensity	N	Mean Difference (MD) (mins)	SE of MD (mins)	SD of MD (mins)	MD + 2 SDs (mins)	MD – 2SDs (mins)

Total	27	100.81	15.2	78.98	258.77	-57.15
Moderate	27	69.7	13.45	69.88	209.46	-70.06
High	27	31.11	6.6	34.3	99.71	-37.49

Note. SD= Standard Deviation, MD=Mean Difference, SE= Standard Error

Pearson product moment correlation coefficients

The relationship between total minutes of activity reported in the PARQ was compared to minutes of activity recorded by the criterion equipment (heart rate and accelerometer). This produced a correlation of $r=0.5$ ($p<0.01$). Pearson correlation coefficients were calculated for high and moderate intensity, at $r=0.55$ and $r=0.4$ ($p<0.01$), respectively.

Limits of Agreement (LOA) Analysis

Total Activity

The extent of the agreement between the criterion measure and PARQ was measured by following Bland and Altman's (1986,1995) Limits of Agreement (LOA) instructions (see Appendix ten). The LOA was conducted firstly for the total activity and secondly for each intensity.

The LOA for total activity (Table 6.9) estimated that there was a 95% possibility that these participants could report between 57.1 minutes below and 258.8 minutes above the amount of physical activity recorded by the criterion measure. Figure 6.3 plots the differences against the averages of both measurement methods. This shows that 95% of the data lies within two standard deviations of the mean difference, thus indicating a normal distribution (British Standards Institution, 1979).

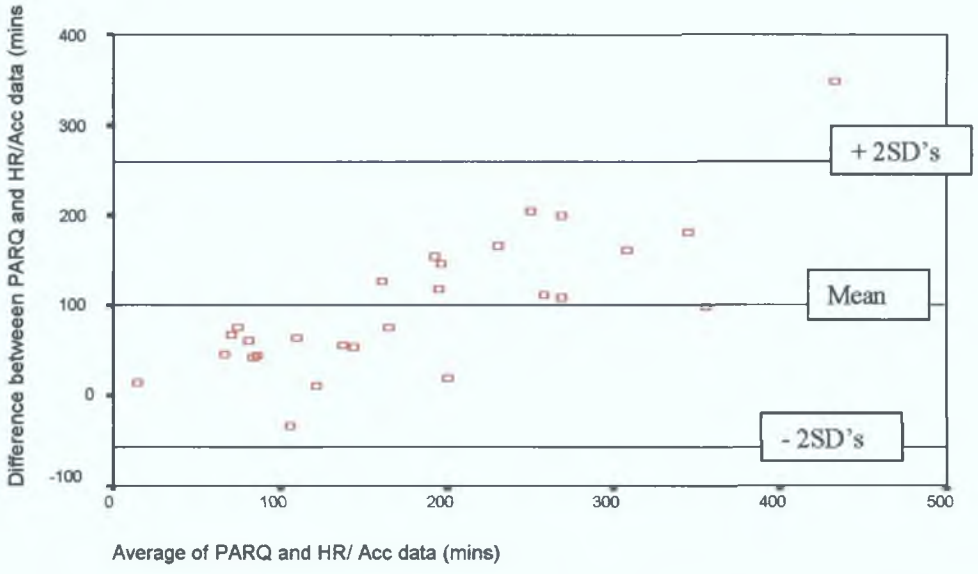

Figure 6.3. Difference against average of PARQ and HR/ RT3 methods with 95% limits of agreement.

The scatter of the differences increases as the amount of activity performed increases, i.e., as the magnitude of activity gets larger, the differences between the two measurement methods increasingly moves away from zero. These results suggest that one of the pieces of equipment becomes less accurate when large amounts of data are recalled. The relative contributions of moderate and high intensity activity to this total figure are displayed below.

To assess whether the LOA for this sample were representative of the general population, the 95% confidence intervals for the upper and lower limits of agreement were then calculated. The standard error of the mean differences is 15.2 minutes. For the 95% confidence interval at 26 degrees of freedom, t=2.056.

Upper limit: 258.77 – (2.056*15.2) to 258.77 + (2.056*15.2) = 227.52 minutes to 290.02 minutes.

Lower limit: -57.15 – (2.056*15.2) to –57.15 + (2.056*15.2) = -88.37 minutes to –25.86 minutes.

The difference between PARQ and the criterion measure (LOA) may be as much as 258 minutes (2SD's) and there is a 95% certainty that if further tests were undertaken in similar circumstances, activity would be underestimated by up to 88 minutes or overestimated by 290

minutes. This is similar to the LOA reported, indicating good precision in the estimated Limits of Agreement.

Moderate and High intensity

The LOA for moderate intensity estimated that for 95% of participants, the amount of physical activity self-reported in PARQ will be between 70 minutes below and 209.5 minutes above the amount of physical activity recorded by the criterion measure.

The LOA for high intensity estimated that for 95% of participants, PARQ will be between 37.5 minutes below and 99.7 minutes above the amount of physical activity recorded by the criterion measure. Figures 6.4 and 6.5 plot the differences against the averages of both measurement methods and show that 95% of the data for both intensities lie within two standard deviations (thus indicating a normal distribution).

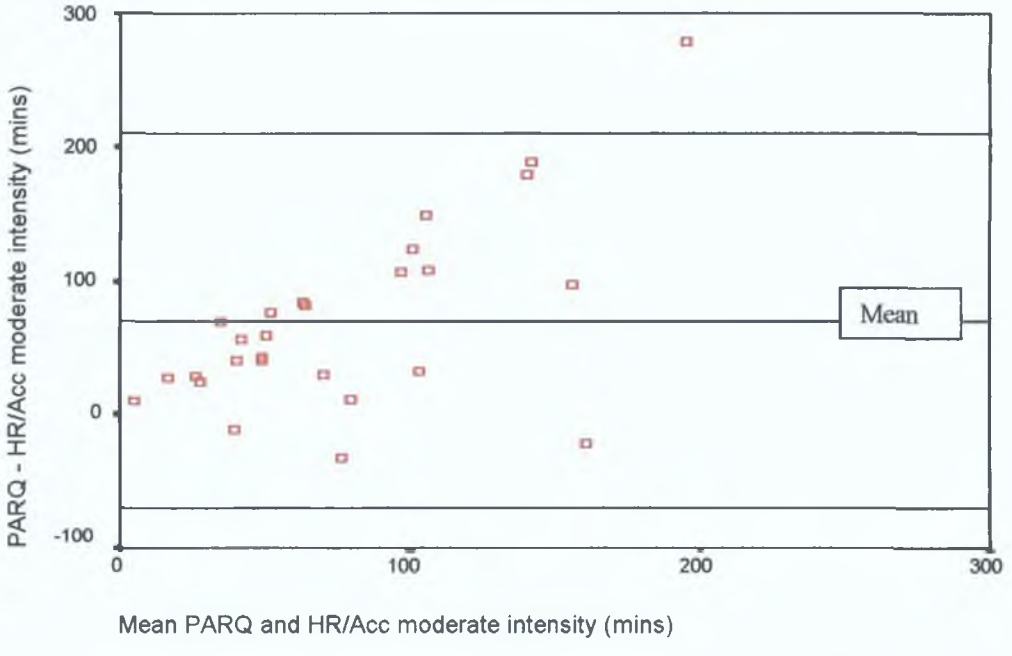

Figure 6.4. Difference against average of PARQ and HR/ Acc methods for moderate intensity with 95% limits of agreement.

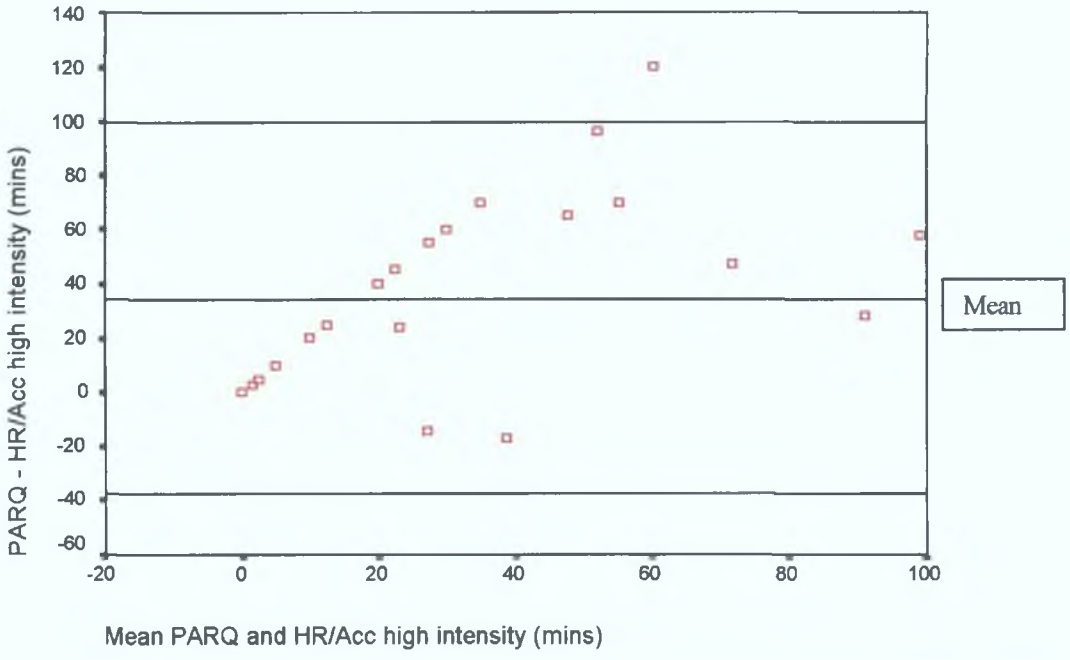

Figure 6.5. Difference against average of PARQ and HR/ Acc methods for high intensity with 95% limits of agreement.

The scatter of the differences increases for both intensities as the amount of activity performed increases. For moderate intensity, the scatter of data is roughly clustered around the zero difference point until approximately 90 minutes, after which the data appears randomly scattered. In contrast, the scatter of data for high intensity appears to increase linearly until about 40 minutes. Average minutes that are >60, generally move back towards the zero difference point. These results indicated a higher volume of activity for moderate intensity activity, which showed less agreement as the volume of activity increases.

The 95% confidence intervals for the upper and lower limits of agreement were then calculated for each intensity, to assess whether the sample LOA was representative of the general population.

Moderate intensity - The standard error of the mean differences is 13.45 minutes. For the 95% confidence interval at 26 degrees of freedom, t=2.056.

Upper limit: 209.46 – (2.056*13.45) to 209.46 + (2.056*13.45) = 181.8 minutes to 237.11 minutes.

Lower limit: -70.06 – (2.056*13.45) to –70.06 + (2.056*13.45) = -97.71 minutes to – 42.4 minutes.

The difference between the PARQ and criterion measure (LOA) may be as much as 209 minutes (2SD's) for moderate intensity activity and there is a 95% certainty that if further tests were undertaken in similar circumstances, that moderate intensity activity would be overestimated by up to 237 minutes.

High Intensity - The standard error of the mean differences is 6.6 minutes. For the 95% confidence interval at 26 degrees of freedom, t=2.056.

Upper limit: 99.71 – (2.056*6.6) to 99.71 + (2.056*6.6) = 86.14 minutes to 113.18 minutes.
Lower limit: -37.49 – (2.056*6.6) to –37.49 + (2.056*6.6) = -23.92 minutes to – 42.4 minutes.

The difference between the PARQ and criterion measure (LOA) may be as much as 99 minutes (2SD's) for high intensity and there is a 95% certainty that if further tests were undertaken in similar circumstances, that high intensity activity would be overestimated by up to 113 minutes.

Hypothesis Three

This stated that PARQ (minutes of activity) will demonstrate concurrent validity, with the simultaneous use of temporally-matched heart rate and RT3 accelerometer data (minutes of activity) as the criterion method. Results from the Pearson's correlation support this hypothesis, while LOA results suggest that the two instruments are not in total agreement.

Hypothesis Four

This stated that high intensity activity (minutes of activity) will demonstrate better concurrent validity than moderate intensity (minutes of activity), with the simultaneous heart rate and RT3 accelerometer data (minutes of activity) as the criterion method. Results from both the Pearson's correlation and LOA support this hypothesis.

6.5. Part three: Construct validity of PARQ with the Stages of Change model of intentional behaviour change

This part of the study was concerned with the validation of PARQ with a construct that measures a person's intention with regard to a specific behaviour, in this case physical activity. The aim of this study was to assess the concurrent validity of PARQ with the Stages of Change physical activity behaviour model (Prochaska & Marcus, 1994).

6.5.1. Participants

A total of 109 participants data (mean age=29.9 (+/-9.1), male=39%) was analysed in this part of this study. This included completed PARQ data from the 73 participants in the reliability study and 36 participants from the criterion validity study. Participants identified themselves as being in the following stages of change; (i) precontemplation (n=4), (ii) contemplation (n=17), (iii) preparation (n=24), (iv) action (n=13), (v) maintenance (n=51).

6.5.2. Procedure

Data from the four days reported total, high and moderate intensity physical activity from each of the 109 participants of the two reliability and validity studies was combined and entered into SPSS (version 10) for analysis. Prior to analysis, the participants were classified as either "not regularly active" (NRA) because they identified themselves as being at Stages 1-3 of the model (i.e., precontemplation, contemplation and preparation), or "regularly active" (RA) because they identified themselves as being at stages 4 and 5 of the model (i.e., action and maintenance).

6.5.3. Data Analysis

Prior to conducting analysis, both NRA and RA data was tested for normality with a Kolomogorov-Smirnov test. Data from NRA participants was normally distributed ($Z=0.12$, $p=0.1$) but data from RA participants was not normally distributed ($Z=0.2$, $p<0.001$). To determine construct validity, an independent t-test would be undertaken to assess whether any significant differences existed between NRA's and RA's for reported total, high and moderate intensity physical activity. However, due to the non-parametric RA data, a Mann-Whitney test was undertaken to assess for significant differences between the two groups.

6.5.4. Results

Descriptive statistics for the 109 participants are presented in Table 6.10 below.

Table 6.10. *Descriptive statistics for regularly active and non-regularly active participants total, moderate and high intensity activity.*

Classification of activity status	N	Mean (mins)	SD (mins)	Min (mins)	Max (mins)
Regularly active	63				
Total		381.4	279.24	64	1200
High intensity		96.13	106.91	0	630
Moderate intensity		285.27	236.19	0	1200
Not regularly active	46				
Total		282.24	213	0	840
High intensity		46.65	65.18	0	390
Moderate intensity		235.7	184.88	0	790

Figure 6.6. presents a graphical representation of the differences in the amount of physical activity reported by those who consider themselves regularly physically active, compared to those who do not consider themselves regularly physically active. For each category, i.e. total, high and moderate intensity, RA's report higher amounts of activity. To assess whether these differences are significant, a Mann-Whitney test was undertaken.

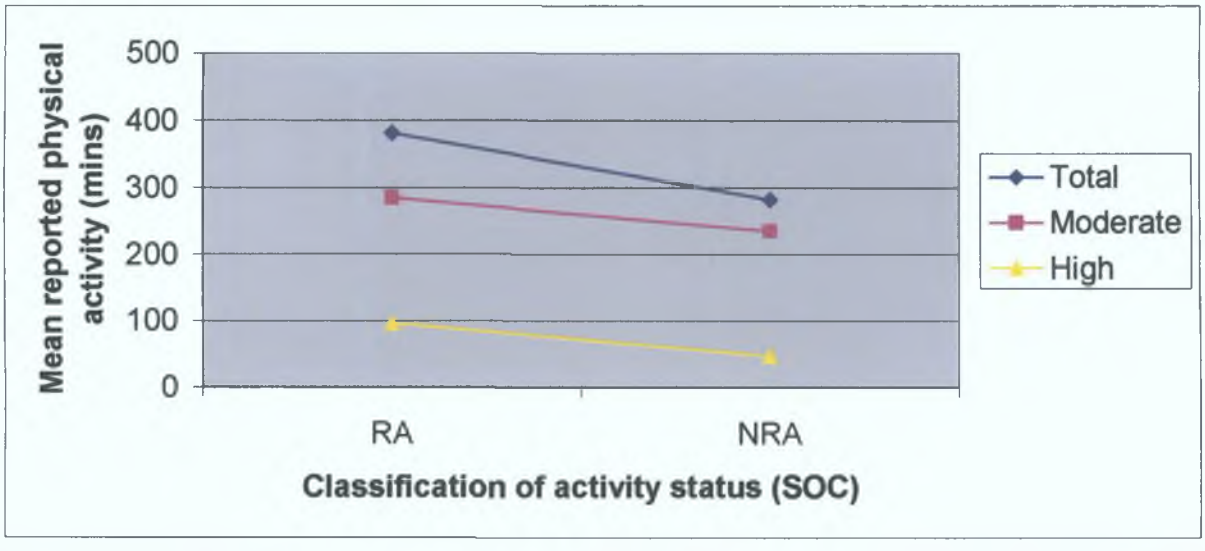

Figure 6.6. *Relation between stages of exercise behaviour change and total, moderate and high intensity physical activity (mean minutes).*
Note. NRA= Not regularly active, RA= Regularly active and SOC= Stage of Change

Mann Whitney U test

Results from this statistical analysis indicates that there is a significant difference between NRA and RA participants for the ranks of total, moderate and high intensity physical activity (see Table 6.11). These results indicate that PARQ has construct validity with the stages of change model, in terms of classifying participants as regularly active and not regularly active, as they show the expected relation between physical activity and the stages of exercise behaviour change.

Table 6.11. *Mean ranking of regularly active and not regularly active participants for total, high and moderate intensity physical activity.*

Classification of activity status	Mean rank	U
Total		1113*
NRA	47.7	
RA	60.3	
High Intensity		942*
NRA	43.9	
RA	63	
Moderate Intensity		1257*
NRA	50.8	
RA	58	

*$p<0.05$

6.6. Discussion

This study reported the procedures undertaken to assess the criterion concurrent validity of PARQ. Initially, the validity of total physical activity was assessed with a RT3 accelerometer as the criterion method. This analysis enabled comparison to other PA questionnaires that had also been validated with an accelerometer, IPAQ (Craig et al., 2003); Kaiser PA Survey for women (Ainsworth et al., 2000b); SPAQ (Lowther et al., 1999); Godin Leisure-time Questionnaire (Godin et al., 1985). A second study was undertaken to individually assess the criterion validity of the moderate and high intensity questions. This represented a novel approach to the validation of a PA questionnaire, as it incorporated the simultaneous use of temporally matched heart rate monitor and accelerometer as criterion methods. The results of both of these

studies are discussed separately below, followed by a discussion on the implications of these findings for the future design of PA questionnaires.

6.5.1. *Criterion method: RT3 Accelerometer*

The validity of this newly designed recall questionnaire (r=0.52) is equal or better than other published physical activity questionnaires, validated with an accelerometer (Table 2.5). The comparison of results does however need to be undertaken with caution, as each of these studies represent a variety of research design, using *different* accelerometer output for validation purposes. For example, in this study, PARQ was validated with accelerometer vector magnitude counts as the criterion measure. Some researchers have preferred to use estimated energy expenditure from the accelerometer (Kcals) as the criterion measure with reported correlations of r=0.34 (N=30) for the Scottish Physical Activity Questionnaire (SPAQ) (Lowther et al., 1999) and r=0.3 (N=50) for the Kaiser Physical Activity Questionnaire with women (Ainsworth et al., 2000b). This has been shown to be an unreliable method of predicting energy expenditure because of the accelerometers inability to calculate increased Kcal expenditure during certain activities, e.g. inclines (Hendelman et al., 2000; Nichols et al., 1999). Validation of the SPAQ found that the underestimation of energy expenditure (Kcals) was related to the accelerometers poor ability to account for the participant's resting metabolic rate (RMR). Participation in lower body resistance training did not register on the accelerometer as increased energy expenditure (Kcals) (Lowther et al., 1999). This led to a decreased estimate of PAEE. Vector magnitude counts are not influenced by the contribution made by the participant's estimated RMR and were found to be comparatively more accurate than Kcals in predicting PAEE (VO_2) (Hendelman et al., 2000). A specific comparison of results from this study with other past-seven day recall PA questionnaires, that use accelerometer counts as the criterion method, shows favourable results for PARQ, as r=0.3 for IPAQ (Craig et al., 2003) and r=0.32 for the Paffenbarger PA questionnaire (Sallis et al., 1985).

The structure of Conway's (1996) model of Autobiographical Memory (ABM) provides a theoretical basis for understanding the results for the breakdown of individual days. The higher correlation for the day immediately previous to the completion of the questionnaire (r=0.78) supports hypothesis two and is reflective of the accessibility to details about recent events at the Event Specific Knowledge (ESK) level and are rich in sensory detail. With appropriate cueing

these memories enter working memory and are quickly available to the respondent for recall. Details of the accessibility of information, offered by Conway's structural model, may explain the gradual decrease in correlation results for the days furthest away from the reference date (i.e., the day the questionnaire was completed). For events that occur further back in time, in this case days, information is also available at ESK through General Event (GE) level but may take longer to enter working memory. This is because of natural forgetting processes, associated with memory decay. These include factors such as the amount of events to be recalled and the amount of time over which the respondent must remember, which are known to increase with time. This results in the merging of similar events, so that common characteristics of that type of event (e.g., walking to work) are remembered but distinctive details about each individual walk to work are forgotten (Wagenaar, 1986; Linton, 1982).

Merging of similar events also encourages respondents to be more likely to estimate an answer, thus reducing accuracy (Burton et al., 1991; Blair et al., 1987; Sudman et al., 1973). An encouraging observation however is that the correlation stabilises at approximately r=0.53 for days three and four. This result reflects a similar forgetting function observed by Ebbinghaus (1885, cited in Eysenck & Keane, 2002), where forgetting was observed to be rapid after a few hours but then slowed down considerably thereafter, with little difference between 48 and 72 hours. Although Ebbinghaus' study involved remembering non-sense syllables and not autobiographical memories, the levelling of correlation results in this study suggests that the rate of forgetting, for a past-seven-day recall questionnaire, may be stable after day three. This result lends support for using a past-seven day recall format, which will provide a more representative picture of PA pattern than a shorter period, e.g., Bouchard's three-day PA record (Bouchard et al., 1983). Further investigation is however required to establish the optimum recall period, which balances the accuracy of information recalled with the representativeness of information about patterns of physical activity among the population.

Memories for the weekend were less accurately recalled (r=0.5) in comparison to the weekday data (r=0.64). Similar to the individual day results, a recency effect is possible, i.e., the two week days were more recent to the reference date than the weekend data. However, the lower weekend correlation may also be a consequence of having to recall more information, particularly on Saturday (Table 6.3.). Unless this information is distinctive and personally salient to the participant, it is more likely to be merged with details of similar, regularly performed

events (Conway, 1996). Weekends also tend to include more leisure-time and be less structured than weekdays. Naturally occurring structures in a day, e.g., starting work, lunchtime, finishing work, appear to provide natural recall cues which assist the participant to recall information (McKenna et al., 2004). Segmenting the day into known contexts also enhanced memory accuracy for children (8-12 years, N=22). More accurate memories were self-recalled for the previous two days when the day was segmented into various sections (e.g., before school, at school, after school, after dinner) compared to when it was not (Baranowski, 1988).

Tourangeau's (1984; Tourangeau et al., 2000) model of the question-answer process provides theoretical support for the decomposition of the recall of PA into individual days and also individual categories (i.e., housework, family). Retrieval strategies for frequency events are more likely to involve an enumerative (individual counting) rather than rate-based (estimation) technique, when the number of events to be recalled is less than ten (Blair et al., 1987; Burton et al., 1991). Decomposition into separate days and categories, provides natural recall cues and reduces the cognitive burden of the respondent. For weekend data, implications from the theoretical basis suggests that a higher volume of data to be remembered, which was not assisted by recall cues associated with the natural structure of the day, did not encourage the employment of an enumerative retrieval strategy. This resulted in a less accurate recall of information.

Implications from these findings suggest that for the most accurate weekend data, PARQ should be completed on Monday (in this way the weekend days are immediately preceding the day the questionnaire is completed). Further research should investigate ways in which people segment or store information about weekend days. This would provide the respondent with naturally occurring recall cues that reflect the structure of those days. A limitation of this study was the inability of the RT3 accelerometer to provide information about the validity of the subdivisions of total PA, i.e., the moderate and high intensity questions. This would provide information about how validity is affected by each intensity and was the reason for the analysis with the combined heart rate and accelerometer units.

6.5.2. *Criterion method: RT3 Accelerometer and Polar Heart Rate Monitor*

This study reports on novel research where PARQ (minutes of activity at moderate and high intensities) was validated with the concurrent use of a temporally matched heart rate monitor and RT3 accelerometer. A review of literature revealed the limitations associated with

using either of these criterion instruments individually to assess lifestyle PA (Jakicic et al., 1999; Lowther et al., 1999; Eston et al., 1998; Nichols et al., 1999). A need was identified for the concurrent use of two instruments that could compensate for the other's limitations (Ainslie et al., 2003). However, the novel aspect of this research means it is not currently possible to directly compare the significant relationship found (r=0.5) with other questionnaire validation studies.

High intensity showed a higher correlation (r=0.55) than moderate intensity (r=0.4), suggesting that memories are more accurately recalled for activities performed at a more vigorous level. This supports hypothesis three and results from Chapter Five (reliability study), which showed in one instance (Pearson's correlation) that high intensity activity was more reliably recalled than moderate intensity. Ease of recall of specific memories is known to be related to the distinctiveness of and the emotionality attached to the activity by the participant at the time of encoding (Wagenaar, 1986). In his diary study, Wagenaar recalled memories he rated as intense more accurately than those he rated as less intense. Studies one and two (Chapters Three and Four) of this thesis support this premise, as participants attached particularly distinctive adjectives to the performance of high intensity activity. In comparison to moderate intensity, high intensity activity evoked more extreme physiological and psychological responses from the participant. Emotive reactions like this make the activity more subjectively memorable for the participant. This affords the opportunity for a more elaborate initial encoding and thus the possibility of a more accurate retrieval (Tourangeau et al., 2000).

The superior high intensity correlation result may be further explained by the comparatively lower amount of time reported doing high intensity (Table 6.8). Both the PARQ and objective criterion data indicate that time spent doing moderate intensity activity was up to three times more than high intensity activity. When repeated experiences occur, they are often harder to remember than those that are unique (Means et al., 1991; Jobe et al., 1993; Wagenaar, 1986; Linton, 1982). Although it is not possible to ascertain from this data whether the same activity reoccurred for these participants, the increased amount of activity performed in the moderate intensity category may suggest that not all activities were unique. Thus the possibility that memories of similar activities may have merged and therefore were not reported, may explain the difference in correlation results. Measuring moderate intensity activity appears to be also problematic for the criterion instrument. The descriptive statistics for the criterion

instrument (Table 6.8.) show a larger standard deviation for moderate intensity than for high intensity activity. This indicates that the distribution of scores for moderate intensity and the extent to which they deviate from the mean score and from each other. An implication from these findings is that moderate intensity is more difficult to accurately measure than high intensity, for both objective and subjective measurement instruments.

The Pearson's correlation indicated that the relationship between PARQ and the criterion measure in study two (r=0.5) was comparable to validity results of study one (r=0.52). Mean differences (minutes) between the two measurements indicate that they are not however in total agreement (Tables 6.10 and 6.11). LOA results for total physical activity indicate that respondents may overestimate by up to four hours. The increasing magnitude in activity offers a reason for this large variation. The difference between the two measurements increases as the average minutes of activity increases (Figure 6.3), i.e., as the amount of activity performed or recalled increases, the difference between the two measurements also increases. The large variation in total physical activity is mainly caused by the influence of the moderate intensity component. Overestimations of up to three and a half hours and underestimations of up to 70 minutes are possible for the moderate intensity section of the questionnaire, in comparison to between minus 30 minutes and plus 90 minutes for high intensity. These results further demonstrate that moderate intensity is harder to recall with accuracy, most probably because of the higher frequency and less emotively salient properties compared to high intensity.

It is possible that the large LOA intervals are not an indication of poor validity but rather a consequence of the limitations of using a novel criterion method. Findings of this study have already shown that the criterion method had difficulty accurately capturing moderate intensity activity. The use of two separate pieces of equipment instead of an integrated unit with an inbuilt synchronised timing system, may have also contributed to an increase in measurement error. Maximum care was taken to temporally match both pieces of equipment, however the possibility of human and measurement error was always a prospect. Results from the pilot study also indicated that the concurrent use of the heart rate monitor and RT3 accelerometer was prone to error (explained 66% of variance). The lack of comparative results from other questionnaire validity studies makes it impossible to appraise the value of the results found here. However, the significant correlation coefficient for each of the total, moderate intensity and high intensity activities does provide an argument for a valid questionnaire.

The results of this study (particularly part two) have shown that PARQ has demonstrated validity in terms of minutes of reported physical activity for both moderate and high intensity. This reflects the unit of measurement used by the ACSM/CDC to assess whether respondents are participating in the minimum amount of physical activity required for achieving health benefits. Therefore, PARQ can be considered a valid instrument that is appropriate for assessing the ACSM/CDC guidelines of accumulating "30 minutes or more of moderate intensity physical activity over the course of most, preferably all, days of the week", (Pate et al., 1995). A limitation of all PA questionnaires is their inability to directly assess physical activity energy expenditure (PAEE). PARQ does however offer the facility to indirectly assess PAEE because of the information that it offers in relation to the FITT principles. The validity that PARQ has demonstrated in terms of *time* spent in different *types* of activity at different *intensities*, means that PAEE can be calculated by assigning units of energy expenditure to each intensity and then multiplying those values by the duration of participation. Pate and colleagues (1995) have assigned kilocalorie values to each intensity. These are 4-7 kcals/min for moderate intensity and > 7 kcals/min for high intensity and are based on a 70kg person. As PARQ has demonstrated validity for "minutes of activity", it permits the extrapolation of the information about time spent in activity of at least moderate intensity to units of energy expenditure. The validity of this procedure will however be limited by the approximate estimates of energy expenditure assigned to each intensity and the assumption of a 70 kg person. The outcome of this calculation will enable comparison of the respondent's PAEE in relation to the recommended PAEE per week, as suggested by the ACSM (2000). This states "the ACSM recommends a target range of 150 to 400 kcal of energy expenditure per day in physical activity. The lower end of this range represents a minimal caloric threshold of approximately 1,000 kcal per week from physical activity" (p.151).

In summary, the theories of memory structure and retrieval (Conway, 1996; Tourangeau et al., 2000) would suggest that the lower performance of moderate intensity compared to high intensity was not unexpected. In terms of the present study, the correlation results for moderate intensity (r=0.41), high intensity (r=0.55) and total activity (0.5), are better than results for many published questionnaires (see Tables 2.5 and 2.6). In this respect, PARQ represents an improvement in the design of past-seven-day physical activity recall questionnaires. The inclusion of various memory cues such as the personal calendar, the decomposition of types of

activities and the affectual descriptors of both intensities, provided the participants with cues to what activity they were doing, when and where they were doing it and how they felt while doing it. The provision of multiple cues, instead of singular cues, is known to positively enhance the retrieval of accurate memories (Wagenaar, 1986) as it reflects the temporal and thematic storage of autobiographical memories (Conway, 1996; Belli et al., 2001). This is the first time that this specific combination of cues has been used in a past seven-day physical activity recall questionnaire.

The stability in correlation as the time increased away from the reference date (day of questionnaire administration), suggests that a past 7-day recall period is an appropriate time frame to seek information. Confining recall to the previous day would offer obvious benefits in terms of accurate memories recalled, however the depth and representativeness of the information received may be compromised. A longer time frame such as seven days provides an indication of pattern of activity, without a significant concession in the quality and validity of retrieved information.

6.5.3. Concurrent validity with stages of exercise behaviour model

The results from this analysis indicate that PARQ demonstrated strong concurrent validity with the stage of exercise behaviour change model of the TTM. A limitation of this study was the collapsing of the five stages into two categories, (i) not regularly active (NRA) and (ii) regularly active (RA). This occurred because of insufficient numbers of pre-contemplators volunteers. Collapsing into two distinct categories permitted equal numbers in the NRA and RA groups. Creating RA and NRA groups was also relevant and appropriate to the research question and the outcome measure of this thesis, i.e. PARQ, which is concerned with quantifying whether individuals are meeting the ACSM/CDC minimum recommended guidelines for regular physical activity (Pate et al., 1995).

Participants who identified themselves as being in the action or maintenance stage (RA) recalled more total, high and moderate intensity activity than the NRA participants, as would be expected. The differences between these two groups were also significant, indicating good validity with the SOC model as it reflected the expected relationship between physical activity and the stage of exercise behaviour change. Similar results were found in concurrent validity studies of self-reported physical activity when the SOC model was collapsed into grouped

categories (Wyse, 1995; Marcus & Simkin, 1993). Wyse found that validity was demonstrated with SOC only for moderate and high intensity activity but not for lower intensity activities. This is possibly due to the characteristics of this type of information, i.e., its indistinctive nature and high frequency (Durante & Ainsworth, 1996) and the associated difficulty in storing and recalling this type of information from memory (Conway, 1996; Tourangeau et al., 2000). The results of this study indicate that PARQ is valid for both RA and NRA individuals.

6.6. Limitations and Future Research

Two different methods were used to assess the criterion validity of PARQ. Despite the different methodological approaches to both studies, the correlation coefficient was consistent at approximately r=0.5 for total physical activity. A breakdown of physical activity by intensity permitted an insight into the relative contributions made by each intensity to the overall validity of the questionnaire. An inclination for moderate intensity activity to negatively affect the overall result was apparent. This is an important discovery in terms of questionnaire development as it suggests that further exploration of the properties of moderate intensity activity is required. In particular, an investigation of the contribution that each of the individual types of activities (e.g., occupational, household) made to the overall validity of the moderate intensity section, may be helpful in identifying problematic areas. This would involve limiting data collecting to specific environments, e.g., work, in order to isolate and validate job-related activity specifically.

The use of individual instruments to simultaneously assess the criterion validity of PARQ has previously been described as a limitation of this study. Future research is therefore required to (i) develop a valid, integrated unit that can concurrently measure heart rate and motion forces, and (ii) use this unit to assess the validity of PARQ and other PA recall questionnaires.

This study monitored only four days of accelerometer and two days of HR recording, because of the limited recording capacity of the HR monitors. Also, in order to reflect possible variations in PA patterns, it was felt necessary to include weekend and weekday data in the recording period, which reduced the number of monitors available. Therefore, although PARQ is a past-seven day recall questionnaire, only four days have been validated. Despite this limitation, PARQ has still managed to demonstrate validity that is equal or better than other PA recall questionnaires. Further testing is required to assess the validity of the seven days of PA that are recalled in PARQ.

CHAPTER 7 – CONCLUSIONS AND FUTURE RESEARCH.

7.1. Conclusions and Implications for Questionnaire Design

The purpose of this research was to design a physical activity recall questionnaire (PARQ) that was reliable and valid in terms of providing information about time spent performing day-to-day physical activity, at different intensities, over the past-seven days. The construct of interest measured by PARQ is "minutes of physical activity (PA) participation", which can then be compared to the recommended minimum amount of participation recommended by the ACSM/CDC for the purposes of gaining health benefits (Pate et al., 1995). Because of its capacity to measure all aspects of the FITT principles (i.e., frequency, intensity, time and type) of activity, PARQ can also indirectly provide information about physical activity energy expenditure (PAEE). This is important in terms of assessing the dose-response relationship between the PA and health (Lamonte & Ainsworth, 2000).

The outcome of this research is a PA recall questionnaire that is reliable and valid in terms of measuring minutes of different types of activity at both moderate and high intensity. The reliability and validity scores of PARQ are a product of its design and the methodologies used to assist the respondents to enhance their memory and recall accurate information. The examination of the structure and processes of memory and recall of information was facilitated by the application of two cognitive models of memory (Conway, 1993, 1996; Tourangeau 1984; Tourangeau et al., 2000). These models formed the foundation of this research and were used to assist in the design process of PARQ (Studies one and two) and to provide an explanation of emergent results (Studies 1-4).

The following sections discuss the conclusions reached at the end of each study, in terms of the findings from each of the individual studies, the implications for the design of PA recall questionnaires in general and suggestions for future research. The chapter concludes with recommendations and limitations for using PARQ, identified as a result of this current research.

7.1.1. Study One

The first study was reported in Chapter Three and was a focus group study that examined the descriptions and images associated with PA-related terms, relevant to the ACSM/CDC health related minimum participation guidelines (Pate et al., 1995). These phrases were "physical

activity" and "moderate and high intensity". This study provided an insight into how participants encode and recall information about PA and its related terms. The outcome of this study was a draft version of PARQ that used the vocabulary of the participants in the questions and in the wording of the recall cues. The physical activity "types" and "examples" in the questionnaire are a product of the data that emerged from the discussions between the participants. These types and examples of activity emerged freely in response to an open question about what the term "physical activity" meant to the participants. This indicated that physical activity is naturally encoded (i.e., understood and stored) within memory according to the type of activity or context in which it occurs. The opportunity for improved recall is enhanced when the recall conditions match the original encoding conditions (Tourangeau et al., 2000). Thus, implications of this finding for questionnaire design are that, as a minimum, questionnaires need to include a variety of different "types" of activity, that are representative of the situations in which the target population participate in physical activity.

The results of this study indicate the importance of including the target population in the design process, yet documentation of item generation for other PA questionnaires is limited. The paradigm shift towards PA for health rather than fitness (Pate et al., 1995), means that many of the pre-1995 PA questionnaires are not appropriate for use in assessing the dose-response relationship between PA and health. This is because they do not truly reflect the moderate intensity range of types of activity that many people engage in daily (Ainsworth, 2000). The design of future PA questionnaires should include a consultation procedure with the target population that enables the discovery of the specific types of activity that are relevant to that population. This discussion environment will also enable the researcher to investigate the target population's understanding of relevant terms that are central to the presumption of completing a PA-related questionnaire, i.e., that the respondent has an understanding of the construct that is being examined. Tudor-Locke and colleagues (2003) found that there was a wide disparity between the definitions of words such as PA, exercise and intensity used for academic purposes and the interpretation of those same terms by a sample population. This current study found that the participants differentiated between moderate and high intensity physical activity through the use of descriptives that detailed the physiological and psychological responses to participation at both intensities.

These findings suggest that participants encode PA at different intensities by the affectual response that they experienced at the time. Implications of this finding for questionnaire design suggests that an array of affectual responses should be offered to the respondent as a recall cue, to assist in the retrieval of memories that are relevant to both intensities. These results also provide further evidence of the importance of involving the target population in the design process as many of the types and affective descriptors used by the participants to differentiate between high and moderate intensity, involved the addition or omission of just one word, e.g., "slight" for moderate and "very" for high intensity. Designers should become familiar with the jargon and vocabulary of the intended population of their questionnaire, to ensure maximal appropriateness and relevance.

Differences between high and moderate intensity were also found to be related to the degree of conscious intent involved when undertaking the activity. This finding had implications for the recall of moderate intensity activity as it was generally discussed in terms of usually occurring in less structured environments and more frequently than high intensity. This type of information is stored in a less accessible area of autobiographical memory (ABM) than the type of activity that characterises high intensity, i.e., less frequent and more salient (e.g., more extreme affect responses). Consequently, designers of PA recall questionnaires should consider the structure of ABM when developing the layout and content of the questionnaire. The natural decomposition of PA into different types of activity by the participants, provided one means of breaking down moderate intensity activity into smaller, less frequent occurrences that could be individually identified. The provision of multiple cues can initiate retrieval pathways in different directions. This should therefore be a consideration in designing the questionnaire, with due consideration given to the relevance of these cues to the target population. Provision of these cues (e.g., types or categories of activities, the specific examples within those categories and the different physiological and psychological effects for each intensity), using the participants own language to describe the context, encourages a question format which should in some way reflect the original encoding conditions and therefore facilitate an appropriate retrieval strategy.

7.1.2. Study Two

The second study was reported in Chapter Four and was a qualitative study that examined the cognitive processes that participants used to answer questions, by means of a cognitive

interview technique. Through a series of observations and verbal responses, the functionality of PARQ was evaluated through the medium of a cognitive model of the question-answer process and as each participant completed it, problems associated with each of the four stages of the model were identified. Questions were consequently amended, so that the outcome of this study was an improved version of PARQ that would be later examined in terms of its psychometric properties.

The findings of this study indicate that cognitive interviewing (CI) process was an effective method, which contributed to understanding what areas of the questionnaire were most problematic, in terms of obtaining accurate and relevant information. However, this specific process of pre-testing the questionnaire is not without its limitations. Specifically, issues pertaining to interviewer competency, complex analysis and relevance to wider populations are valid (Beatty, 2002). The process of cognitive interviewing is also ill-defined in terms of methods used to evaluate CI techniques, with suggestions that data reduction methods are too dependent upon the researchers subjective judgements. This may affect reproducibility and generalisability to the field (Willis & Schechter, 1996). Evidence-based research involving CI methods of questionnaire pre-testing is limited and therefore, more research needs to be conducted to investigate (i) methods to assess which types of probes are efficient in identifying problems associated with particular cognitive processes, (ii) ways of evaluating and standardising analysis procedures, and (iii) the evaluation of the effectiveness of questionnaires that have been pre-tested with CI techniques in the laboratory, in the field.

The findings of this particular study suggest that despite these limitations, newly designed questionnaires require some method of pre-testing prior to being assessed for reliability and validity. Tourangeau's (1984; Tourangeau et al., 2000) model of the question-answer process provides a framework for evaluating participant responses and reactions to completing the questionnaire that is based on established theory of ABM structure and retrieval processes (Conway, 1993, 1996). Implications from this particular study suggest that, for future PA questionnaires, designers should apply this model of the question-answer process to pre-test the questionnaire, in order to ensure that it is a functional instrument that is understood and can be completed by the target population. Pertinent issues for PA questionnaire design, raised as a result of this study, are reflected in the diversity and frequency of problems that arose. Initially, problems associated with understanding the context and structure of the questionnaire were

paramount (the "encoding" stage). This appeared to adversely affect the frequency of problems associated with the "judgement" stage, because as the encoding issues were addressed and diminished in number, so too did the occurrence of issues associated with judgement. The dominant judgement issue appears to have been related to the participants motivation to recall the information asked, which was perceived by the participants as potentially cognitively burdensome and also perhaps related to the encoding issues and inability to understand what was asked. Designers of PA questionnaires need to consider the effect that the recall of difficult and taxing information from memory may have on the accuracy of answers given. Methods of improving the respondents motivation to complete the questionnaire and provide accurate information need to be further investigated. This is of particular significance given the naturally burdensome quality of recalling information about frequently performed activities from memory.

The calculation of time spent doing particular activities presented problems for participants. The properties of frequency and duration of activities are of paramount importance to the dose-response relationship between PA and health. Therefore further research needs to investigate methods of accessing this type of information in ABM. Particular problems were identified with calculating the duration of activities that were not continuous and occurred at various times over the day, e.g., climbing stairs. Conway's (1993, 1996) model of the structure of ABM suggests that calculation can be enhanced by parallel and sequential cueing. For example, a cue that reminded a respondent where they were with a particular person on a given day, might help them remember the start time of the walk they did when they left that person to go home. Further investigations are required to examine the encoding and retrieval processes involved in calculating the duration and frequency of different types of physical activities, e.g., continuous and discontinuous, frequent and infrequent. The properties of frequency and duration were not specifically included for discussion in the focus group study of this research. This type of discussion forum may have provided more detailed information about encoding and retrieval properties than was evident from the CI process.

7.1.3. Study Three

The third study was reported in Chapter Five and reported on the assessment of the test-retest reliability of PARQ. It also investigated the effect of different design features on the

reliability of information obtained, namely (i) the order that questions were presented in the questionnaire and (ii) the presence or absence of a recall cue called a "personal calendar".

Results from the first part of this study provide strong support for the inclusion of a personal calendar (PC) in a PA recall questionnaire. The reliability of information recalled improved when participants first completed the calendar with details of personal significance that had occurred each day over the past seven days. Conway's (1993, 1996) model of ABM suggests that this is because completing the PC (i) provides personal landmarks that mirrors the personal goals that shape the structure of ABM, (ii) provides multiple types of retrieval cues (when, where, with who) that reflect the thematic and temporal pathways of ABM that provide access to relevant information, and (iii) provides a contextual timeframe for the recall of relevant information, helping to exclude activities that are not within the requested timeframe and to include activities that are. The stronger relationship of total activity as well as high and moderate intensity for the questionnaire with the PC, compared to without, has implications for the design of other PA questionnaires. Further research could investigate the effect of different types of information (cues) included in the PC, in improving the accuracy of information recalled.

A comprehensive approach to assessing PARQ's reliability was undertaken in this study, by means of (i) interclass correlation, (ii) intraclass correlation and (iii) repeatability coefficients. Results from these analyses show that total physical activity is highly correlated but that agreement between repeated measurements of PARQ diminishes as the volume of activity recalled increases. This result highlights the advantage in carrying out more than one type of statistical analysis, as has typically been done to evaluate the reliability of other PA questionnaires. The results of the analysis of activity by intensity has further implications for the design of PA questionnaires. This study concluded that the lack of agreement between test and retest total activity was primarily due to the large volume of moderate intensity activity recalled, in comparison to high intensity activity. Examination of the types of activity, revealed that the large volume of moderate intensity sport and exercise activity was the main cause of the lack of agreement between test and retest. Despite the apparent problem given by sport and exercise in terms of repeatability, it was decided not to remove it from PARQ, because this category of activity had also demonstrated acceptable scores for both the interclass and intraclass correlations (Vincent, 1999). Additionally, the occupational activity had demonstrated the lowest correlation but had shown the second best repeatability score. This illustrated inconsistencies in

the reporting of the coefficient of repeatability. It was concluded from these results that the most consistent and appropriate method of statistical analysis for repeated measures was the intraclass correlation. However, a recommendation for assessment of other PA questionnaires is to undertake a variety of statistical analyses, in order to obtain a comprehensive profile of the reliability properties of the new instrument.

The results of this study showed that moderate intensity activity was equally as reliable as high intensity. This result is in contrast to all other PA questionnaires reviewed during this thesis (Table 2.5.), thus it was concluded that the encoding properties of PA and the recall cues obtained as a result of studies one and two (focus group and cognitive interviews) were highly effective and appropriate. Future research could make additional enquiries into the properties of moderate intensity, in particular, to further the understanding of what cues are most personally salient and effective for accessing information in memory about frequently performed activity. This could serve the dual purpose of enhancing reliability further and reducing the cognitive burden associated with retrieving large amounts of information. For the same reasons, experimental studies could also examine the comparative effectiveness of different types of cues in helping the respondent to retrieve information.

The investigation of the impact the order of questions within the questionnaire on the reliability indicate that there is no benefit to be obtained from placing the high intensity question first. The conclusion from this study is that no order effect exists for high and moderate intensity questions, however future research could investigate whether an order effect exists in terms of what order the types of activity questions are placed. Other topics requiring some investigation include the effect that changes in (i) physical environment, (ii) instructions and (iii) timeframe, between test and retest might have on the reliability scores of a new PA questionnaire.

7.1.4. Study Four

The fourth study was reported in Chapter Six and detailed the procedures undertaken to assess the criterion concurrent validity of PARQ. Two separate analyses were completed to assess (i) the relationship between total physical activity (minutes) and a RT3 accelerometer (vector magnitude) and (ii) the relationship between total, moderate and high intensity activity (minutes) and combined heart rate and accelerometer data (minutes of activity).

The results of the first analysis indicated that the correlation score for PARQ was better than the scores of other PA recall questionnaires that had been correlated with accelerometer vector magnitude counts. This result denoted that PARQ is equally, if not more, valid an instrument than other PA questionnaires, but did not specify it as valid in terms of measuring minutes of physical activity. An acceptable validity score in terms of measuring minutes of activity of at least moderate intensity, is important as it demonstrates PARQ's ability to accurately capture information that is pertinent to the ACSM/CDC (Pate et al., 1995) health related PA participation guidelines. The findings from the second analysis (with combined heart rate and accelerometer data), support the conclusion that PARQ is a valid instrument for measuring minutes of moderate and high intensity activity.

A breakdown of scores by intensity reiterated the problematic areas of recalling and recording moderate intensity activity that were highlighted in study three. It was concluded that the lower correlation score and limits of agreement demonstrated by moderate intensity compared to high intensity, was in part affected by the larger volume of recorded (by heart rate and accelerometer) and recalled (PARQ) moderate intensity physical activity. The difficulty that the combined heart rate and accelerometer instruments exhibited in measuring moderate intensity activity, suggested that this was not an optimal criterion measure. The limitations of using this criterion measure included the influence of human measurement error in manually, temporally matching the two instruments and a suggestion for future research is the development of an integrated unit that simultaneously measures heart rate and accelerometer data.

Despite these limitations, the correlation for moderate intensity activity was higher in PARQ than for other PA recall questionnaires, e.g., Seven-day PA recall questionnaire (Sallis et al., 1985), Godin Leisure-time Exercise Questionnaire (Godin & Shephard, 1985), see Table 2.5. Moderate intensity also showed decreased levels of agreement as the volume of information recalled increased. Its comparatively lower result to high intensity does however suggest that there are still difficulties in accurately recalling activity of a lower intensity. Further research that examines the specific encoding and retrieval strategies for moderate intensity is thus warranted. Because of the higher frequency and volume of information that is a feature of moderate intensity, examination of possible methods of helping the respondent to decompose the large amount of information in memory may reduce the cognitive burden. Research should also investigate the prominent calculation strategies that are used to recall different types of moderate

intensity activity. Identification of these strategies may provide a suggestion for a particular structure or wording of a question that could act as a recall cue and potentially enhance accuracy.

The collective results of this study indicate that PARQ is a valid recall questionnaire that demonstrates comparatively better accuracy in directly measuring minutes of total, moderate and high intensity physical activity than other PA recall questionnaires. The ability of PARQ to obtain accurate information about duration and intensity of different types of activity, also means that PARQ can indirectly ascertain the energy expended by the respondent as a result of participation in PA. In this way, PARQ is a valid instrument for obtaining information that is relevant to (i) the ACSM/CDC (1995) health related guidelines, in terms of minutes of participation of moderate intensity activity, and (ii) the dose-response relationship between health and PA, guided by the ACSM (2000) recommendations of minimum kcal energy expenditure required to incur health benefits. Further research is however required to validate each specific type of activity measured within PARQ, i.e., commuting, occupation etc. This could be achieved by limiting the data collection to one specific area, e.g., work. However, validation of categories such as housework may present some difficulties because of its discontinuous nature. Further research is also required to ascertain the optimal recall period. Although findings in this study showed that forgetting stabilised after three days, this result was in the context of having measured only four days of data in total. The optimal recall period would ideally be long enough to provide information about patterns of activity, yet be short enough to maximise accurate recall and minimise respondent burden. PARQ should also be further validated with seven days data.

7.2. Recommendations and Limitations of PARQ

This thesis has provided evidence that various features such as a personal calendar, decomposed categories and descriptors of intensity of activity, should be included as standard in the design of past-seven day physical activity recall questionnaires. A comprehensive overview of all the findings of the studies within this thesis, shows that the acceptable reliability and validity results achieved, particularly for moderate intensity activity, are a product of the application of the two cognitive models of the structure and process of autobiographical memory (Conway, 1993, 1996; Tourangeau, 1984, Tourangeau et al., 2000). It is a recommendation from this study, that other PA questionnaires should incorporate these models into the developmental

stage of the questionnaire. They should also be used as a theoretical framework to explain results and data, obtained from the qualitative studies that form the item generation and pre-testing stages of design.

PARQ has demonstrated itself to be a reliable and valid instrument for measuring total, moderate and high intensity physical activity. There are however, some limitations in its ability to accurately capture some types of information. The limitations associated with both the coefficient of repeatability and the limits of agreement results have been previously outlined, in studies three and four respectively. Despite these shortcomings, the agreement of PARQ, firstly against a repeated measure of itself and secondly, against a criterion measure, was observed to decrease as the volume of activity recalled increased. Separation of the total activity data by intensity, revealed that the principle contributor to the lack of agreement was moderate intensity. In the reliability study, further analysis identified moderate intensity sport and exercise as the category most affected by the recall difficulties associated with moderate intensity, because of the large volume of activity involved. Possible variations of up to 140 minutes for this category, if the questionnaire was repeated under similar circumstances, might support a proposal for its removal from the questionnaire. However, this category also performed very highly in the interclass ($r=0.9$) and intraclass (0.95) correlation results and so it was decided that sport and exercise had demonstrated acceptable reliability and should remain in the questionnaire.

This result does however point to a feature of this questionnaire that should be taken into consideration by any researchers or practitioners that may intend to use it. The comparative difficulties in recalling moderate intensity activity, observed for all types of activity and in both the reliability and validity studies, should be noted. This may be of significance for studies that wish to focus primarily upon the measurement of moderate intensity activity. Of equal significance however, is the superior performance of moderate intensity and each of the individual categories (including sport and exercise) compared to the reliability and validity scores of other published PA recall questionnaires. This shows that while problems in accessing the recall of moderate intensity are still apparent, the design of PARQ has facilitated an improved method of assisting the respondent to retrieve accurate information. Research into methods of improving access to and understanding of the storage facility of moderate intensity physical activity within ABM is needed. However, a valid argument for the consistent poorer performance of moderate intensity recall may be that the very nature and characteristics of this

level of intensity, simply means that it is naturally a very cognitively burdensome type of information that is inherently difficult to retrieve accurately.

The methodologies applied to the studies in this thesis were intended to assist in developing a reliable and valid past-seven-day recall questionnaire that could obtain accurate information about participation in different types of day-to-day physical activity. The outcome of this thesis is PARQ, which has the facility to obtain accurate, direct information about the minutes of PA participation and indirectly obtain information about energy expenditure, both of which are appropriate for comparison to the recommended minimum participation guidelines for health benefits (Pate et al., 1995; ACSM, 2000). It is also hoped that the methodologies employed and the findings of these studies will contribute to the improved development of future PA recall questionnaires.

References

Aadahl, M. & Jorgensen, T. (2003). Validation Of A New Self-Report Instrument For Measuring Physical Activity. *Medicine And Science In Sports And Exercise, 33,* 1196-1202.

ACSM (2000). *ACSM's Guidelines For Exercise Testing And Prescription.* (6th Ed.) Baltimore: Lippincott, Williams And Wilkins.

Ainslie, P. N., Reilly, T., & Westerterp, K. R. (2003). Estimating Human Energy Expenditure: A Review Of Techniques With Particular Reference To Doubly Labelled Water. *Sports Medicine, 33,* 683-698.

Ainsworth, B. E. (2000). Challenges In Measuring Physical Activity In Women. *Exercise And Sport Sciences Reviews, 28,* 93-96.

Ainsworth, B. E., Haskell, W. L., Leon, A. S., Jacobs, D. R. Jr., Montoye, H. J., Sallis, J. F. Et Al. (1993a). 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., Haskell, W. L., Whitt, M. C., Irwin, M., Swartz, A. M., Strath, S. J. Et Al. (2000a). Compendium Of Physical Activities: An Update Of Activity Codes And MET Intensities. *Medicine And Science In Sports And Exercise, 32,* S498-S516.

Ainsworth, B. E., Jacobs, D. R. Jr., & Leon, A. S. (1993b). Validity And Reliability Of Self-Reported Physical Activity Status: The Lipid Research Clinics Questionnaire. *Medicine And Science In Sports And Exercise, 25,* 92-98.

Ainsworth, B. E., Richardson, M. T., Jacobs, D. R. Jr., Leon, A. S., & Sternfeld, B. (1999). Accuracy Of Recall Of Occupational Physical Activity By Questionnaire. *Journal Of Clinical Epidemiology, 52,* 219-227.

Ainsworth, B. E., Sternfeld, B., Richardson, M. T., & Jackson, K. (2000b). Evaluation Of The Kaiser Physical Activity Survey In Women. *Medicine And Science In Sports And Exercise, 32,* 1327-1338.

Allison, R. B. (1996). *Cognitive Interviewing: A Critical Evaluation* Newsletter Of The American Academy Of Psychiatry And The Law.

Allor, K. M. & Pivarnik, J. M. (2001). Stability And Convergent Validity Of Three Physical Activity Assessments. *Medicine And Science In Sports And Exercise, 33,* 671-676.

Anderson, S. J. & Conway, M. (1993). Investigating The Structure Of Autobiographical Memories. *Journal Of Experimental Psychology: Learning, Memory And Cognition, 19,* 1178-1196.

Atkinson, G. & Nevill, A. (1998). Statistical Methods For Assessing Measurement Error (Reliability) In Variables Relevant To Sports Medicine. *Sports Medicine, 26,* 217-238.

Atkinson, G. & Nevill, A. (2000). Typical Error Versus Limits Of Agreement. *Sports Medicine, 30,* 375-381.

Baddeley, A. (1996). Exploring The Central Executive. *The Quarterly Journal Of Experimental Psychology, 49,* 5-28.

Baecke, J. A. H., Burema, J., & Frijters, J. E. R. (1982b). A Short Questionnaire For The Measurement Of Habitual Physical Activity In Epidemiological Studies. *American Journal Of Clinical Nutrition, 36,* 936-942.

Baecke, J. A. H., Burema, J., & Frijters, J. E. R. (1982a). A Short Questionnaire For The Measurement Of Habitual Physical Activity In Epidemiological Studies. *American Journal Of Clinical Nutrition, 36,* 936-942.

Bain, L. (1989). Interpretive And Critical Research In Sport And Physical Education. *Research Quarterly For Exercise And Sport, 60,* 21-24.

Baranowski, T. (1988). Validity And Reliability Of Self Report Measure Of Physical Activity: An Information Processing Perspective. *Research Quarterly For Exercise And Sport, 59,* 314-327.

Baranowski, T. (1985). Methodological Issues In Self Report Of Health Behaviour. *Journal Of School Health, 55,* 179-182.

Baranowski, T. & De Moor, C. (2000). How Many Days Was That? Intra-Individual Variability And Physical Activity Assessment. *Research Quarterly For Exercise And Sport, 71,* 74-78.

Baranowski, T., Smith, M., Thompson, W. O., Baranowski, J., Hebert, D., & Demoor, C. (1999). Intraindividual Variability And Reliability In A 7-Day Exercise Record. *Medicine And Science In Sports And Exercise, 31,* 1619-1622.

Barbour, R. S. & Kitzinger, J. (1999). *Developing Focus Group Research.* (1st Ed.) London: Sage Publications.

Barsalou, L. W. (1988). The Content And Organisation Of Autobiographical Memories. In U.Neisser & E. Winograd (Eds.), *Remembering Reconsidered: Ecological And Traditional Approaches To The Study Of Memory* (1st Ed., Pp. 193-243). Cambridge: Cambridge University Press.

Bassett, D. R. Jr. (2000). Validity And Reliability Issues In Objective Monitoring Of Physical Activity. *Research Quarterly For Exercise And Sport, 71,* 30-36.

Bassett, D. R. Jr., Ainsworth, B. E., Swartz, A. M., Strath, S. J., o'brien, W. L., & King, G. A. (2000). Validity Of Four Motion Sensors In Measuring Moderate Intensity Physical Activity. *Medicine And Science In Sports And Exercise, 32,* S471-S480.

Beatty, P. (2004). The Dynamics Of Cognitive Interviewing. In S.Presser, J. M. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods In Testing And Evaluating Survey Questionnaires* (1st Ed., Pp. 45-66). New Jersey: John Wiley And Sons Inc.

Belli, R. F. (1998). The Structure Of Autobiogrpahical Memory And The Event History Calendar:Potential Improvements In The Quality Of Retrospective Reports In Surveys. *Memory, 6,* 383-406.

Belli, R. F., Shay, W. L., & Stafford, F. P. (2001). Event History Calendar And Question List Surveys. A Direct Comparison Of Interviewing Methods. *Public Opinion Quarterly, 65,* 45-74.

Belson, W. A. (1981). *The Design And Understanding Of Survey Questions.* (1st Ed.) Aldershot, UK: Gower.

Biddle, S. J. H., Markland, D., Gilbourne, D., Chazisarantis, N. L. D., & Sparkes, A. C. (2001). Research Methods In Sport And Exercise Psychology: Quantitative And Qualitative Issues. *Journal Of Sports Sciences, 19,* 777-809.

Blair, E. A. & Burton, S. (1987). Cognitive Processes Used By Survey Respondents To Answer Behavioural Frequency Questions. *Journal Of Consumer Research, 14,* 280-288.

Blair, E. A. & Ganesh, G. K. (1991). Characteristics Of Interval-Based Estimates Of Autobiographical Frequencies. *Applied Cognitive Psychology, 5,* 237-250.

Blair, J. & Presser, S. (1993). Survey Procedures For Conducting Cognitive Interviews To Pretest Questionnaires: A Review Of Theory And Practice.

Blair, S. N., Cheng, Y., & Holder, J. S. (2001). Is Physical Activity Or Physical Fitness More Important In Defining Health Benefits? *Medicine And Science In Sports And Exercise, 33,* S379-S399.

Blair, S. N., Cheng, Y., & Scott Holder, J. (2001). Is Physical Activity Or Physical Fitness More Important In Defining Health Benefits? *Medicine And Science In Sports And Exercise, 33,* 379-399.

Blair, S. N., Haskell, W. L., Ho, P., Paffenbarger, R. S. Jr., Vranizan, K. M., Farquhar, J. W. Et Al. (1985). Assessment Of Habitual Physical Activity By A Seven-Day Recall In A Community Survey And Controlled Experiments. *American Journal Of Epidemiology, 122,* 794-804.

Blair, S. N., Jacobs, D. R. Jr., & Powell, K. E. (1985). Relationships Between Exercise Or Physical Activity And Other Health Behaviours. *Public Health Reports, 100,* 172-180.

Bland, J. M. & Altman, D. G. (1986). Statistical Methods For Assessing Agreement Between Two Methods Of Clinical Measurement. *The Lancet, 1,* 307-310.

Bland, J. M. & Altman, D. G. (1995). Comparing Methods Of Measurement: Why Plotting Difference Against Standard Method Is Misleading. *The Lancet, 346,* 1085-1087.

Booth, M. (2000). Assessment Of Physical Activity: An International Perspective. *Research Quarterly For Exercise And Sport, 71,* 114-120.

Booth, M., Owen, N., Bauman, A., & Gore, C. J. (1996). Retest Reliability Of Recall Measures Of Leisure Time Physical Activity In Australian Adults. *International Journal Of Epidemiology, 25,* 153-159.

Boreham, C., Twisk, J., Murray, L., Strain, J. J., & Cran, G. (2001). Fitness, Fatness And Coronary Heart Disease Risk In Adolescents: The Northern Ireland Young Hearts Project. *Medicine And Science In Sports And Exercise, 33,* 270-274.

Bouchard, C. (2001). Physical Activity And Health: Introduction To The Dose-Response Symposium. *Medicine And Science In Sports And Exercise, 33,* 347-350.

Bouchard, C., Tremblay, A., Leblanc, C., Lortie, G., Savard, R., & Theriault, G. (1983). A Method To Assess Energy Expenditure In Children And Adults. *American Journal Of Clinical Nutrition, 37,* 461-467.

Bradburn, N., Rips, L. J., & Shevell, S. K. (1987). Answering Autobiographical Questions: The Impact Of Memory And Inference On Surveys. *Science, 236,* 157-161.

Brewer, W. F. (1996). What Is Recollective Memory? In D.C.Rubin (Ed.), *Remembering Our Past-Studies In Autobiographical Memory* (First Ed., Pp. 19-66). USA: Cambridge University Press.

Brewer, W. F. (1988). Memory For Randomly Sampled Autobiographical Events. In U.Neisser & E. Winograd (Eds.), *Remembering Reconsidered: Ecological And Traditional Approaches To The Study Of Memory* (1st Ed., Pp. 21-90). Cambridge: Cambridge University Press.

British Standards Institution, P. O. T. M. (1979). 1: Guide For Determination And Reproducibility For A Standard Test Method. In *BS 5497, Part 1* (London: BSI.

Brown, N. R., Rips, L. J., & Shevell, S. K. (1985). The Subjective Dates Of Natural Events In Very Long Term Memory. *Cognitive Psychology, 17,* 139-177.

Burton, S. & Blair, E. A. (1991). Task Conditions, Response Formulation Processes And Response Accuracy For Behavioral Frequency Questions In Surveys. *Public Opinion Quarterly, 55,* 50-79.

Campbell, K. L., Crocker, P. R., & Mckenzie, D. C. (2002). Field Evaluation Of Energy Expenditure In Women Using Tritrac Accelerometers. *Medicine And Science In Sports And Exercise, 24,* 1667-1674.

Carroll, S., Cooke, C. B., & Butterly, R. J. (2000). Metabolic Clustering, Physical Activity And Fitness In Non-Smoking, Middle-Aged Men. *Medicine And Science In Sports And Exercise, 32,* 2079-2086.

Caspersen, C. J. (1989). Physical Activity Epidemiology: Concepts,Methods And Applications To Exercise Science. *Exercise And Sport Sciences Reviews, 17,* 423-473.

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,* 126-131.

Chen, K. Y. & Sun, M. (1997). Improving Energy Expenditure Estimation By Using A Triaxial Accelerometer. *Journal Of Applied Physiology, 83,* 2112-2122.

Conrad, F. & Blair, J. (2004). Data Quality In Cognitive Interviews: The Case Of Verbal Reports. In S.Presser, J. M. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods For Testing And Evaluating Survey Questionnaires* (1st Ed., Pp. 67-88). New Jersey: John Wiley And Sons Inc..

Conway, J. M., Seale, J. L., Jacobs, D. R. Jr., Irwin, M., & Ainsworth, B. E. (2002). Comparison Of Energy Expenditure Estimates From Doubly Labelled Water, A Physical Activity Questionnaire And Physical Activity Records. *American Journal Of Clinical Nutrition, 75,* 519-525.

Conway, M. (2001). Sensory-Perceptual Episodic Memory And Its Context: Autobiographical Memory. *Phil Trans R Soc, 356,* 1375-1384.

Conway, M. & Rubin, D. C. (1993). The Structure Of Autobiographical Memory. In A.F.Collins, S. E. Gathercole, M. Conway, & P. E. Morris (Eds.), *Theories Of Memory* (1st Ed., Pp. 103-137). Hillsdale:USA: Lawrence Erlbaum Associates.

Conway, M. A. (1996). Autobiographical Knowledge And Autobiographical Memories. In D.C.Rubin (Ed.), *Remembering Our Past- Studies In Autobiographical Memory* (First Ed., Pp. 67-93). USA: Cambridge University Press.

Cote, J., Salmela, J. H., Baria, A., & Russell, S. J. (1993). Organising And Interpreting Unstructured Qualitative Data. *The Sport Psychologist, 7,* 127-137.

Craig, C. L., Marshall, A. L., Sjostrom, M., Bauman, A., Booth, M., Ainsworth, B. E. Et Al. (2003). International Physical Activity Questionnaire: 12-Country Reliability And Validity. *Medicine And Science In Sports And Exercise, 35,* 1381-1395.

Creswell, J. W. (1998). *Qualitative Inquiry And Research Design: Choosing Among Five Traditions*. (1st Ed.) Thousand Oaks: California: Sage Publications Inc.

Davidson, L., Mcneill, G., Haggarty, P., Smith, J. S., & Franklin, M. (1997). Free-Living Energy Expenditure Of Adult Men Assessed By Continuous Heart-Rate Monitoring And Doubly Labelled Water. *British Journal Of Nutrition, 78,* 695-708.

DeMaio, T. & Landreth, A. (2004). Do Different Cognitive Interview Techniques Produce Different Results? In S.Presser, J. M. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods For Testing And Evaluating Survey Questionnaires* (1st Ed., Pp. 89-108). New Jersey: John Wiley And Sons Inc.

Dubbert, P. M. (2002). Physical Activity And Exercise: Recent Advances And Current Challenges. *Journal Of Consulting And Clinical Psychology, 70,* 526-536.

Dunn, A. L., Trivedi, M. H., & O'Neal, H. A. (2001). Physical Activity Dose-Response Effects On Outcomes Of Depression And Anxiety. *Medicine And Science In Sports And Exercise, 33,* 587-597.

Durante, R. & Ainsworth, B. E. (1996). The Recall Of Physical Activity: Using A Cognitive Model Of The Question-Answering Process. *Medicine And Science In Sports And Exercise, 28,* 1282-1291.

Ericsson, K. & Simon, H. (1993). *Protocol Analysis: Verbal Reports As Data.* (2nd Ed.) Cambridge, MA: MIT Press.

Ericsson, K. & Simon, H. (1980). Verbal Reports As Data. *Psychological Review, 87,* 215-251.

Eston, R. G., Rowlands, A. V., & Ingledew, D. K. (1998). Validity Of Heart Rate, Pedometry And Accelerometry For Predicting The Energy Cost Of Children's Activities. *Journal Of Applied Physiology, 84,* 362-371.

Eysenck, M. W. & Keane, M. T. (2001). *Cognitive Psychology: A Student's Handbook.* (1st Ed.) East Sussex: Psychology Press Ltd.

Fahey, T., Layte, R., & Gannon, B. (2004). *Sports Participation And Health Among Adults In Ireland* Dublin: The Economic And Social Research Institute (Ireland).

Flick, U. (2002). *An Introduction To Qualitative Research.* (2nd Ed.) London:UK: Sage Publications Inc.

Foddy, W. (1998). An Empirical Evaluation Of In-Depth Probes Used To Pretest Survey Questions. *Sociological Methods And Research, 27,* 103-133.

Frankland, J. & Bloor, M. (1999). Some Issues Arising In The Systematic Analysis Of Focus Group Materials. In R.S.Barbour & J. Kitzinger (Eds.), *Developing Focus Group Research* (1st Ed., Pp. 144-155). London: Sage Publications.

Freedson, P. S. & Miller, K. (2000). Objective Monitoring Of Physical Activity Using Motion Sensors And Heart Rate. *Research Quarterly For Exercise And Sport, 71,* 21-29.

Friedenreich, C. M. (1994). Improving Long-Term Recall In Epidemiologic Studies. *Epidemiology, 5,* 1-4.

Friedenreich, C. M., Courneya, K. S., & Bryant, H. E. (2001). Influence Of Physical Activity In Different Age And Life Periods On The Risk Of Breast Cancer. *Epidemiology, 12,* 604-612.

Friedenreich, C. M., Courneya, K. S., & Bryant, H. E. (1998). The Lifetime Total Physical Activity Questionnaire: Development And Reliability. *Medicine And Science In Sports And Exercise, 30,* 266-274.

Friedenreich, C. M. & Rohan, T. E. (1995). A Review Of Physical Activity And Breast Cancer. *Epidemiology, 6,* 311-317.

Gerber, E. (2002). Cognitive Interviewing: A Short Course Presented At QDET Conference, South Carolina, USA. In.

Godin, G. & Shephard, R. J. (1985). A Simple Method To Assess Exercise Behaviour In The Community. *Canadian Journal Of Applied Sports Science, 10,* 141-146.

Haskell, W. L., Yee, M. C., Evans, A., & Irby, P. (1993). Simultaneous Measurement Of Heart Rate And Body Motion To Quantitate Physical Activity. *Medicine And Science In Sports And Exercise, 25,* 109-115.

Healey, J. (2000). Future Possibilities In Electronic Monitoring Of Physical Activity. *Research Quarterly For Exercise And Sport, 71,* 137-145.

Hendelman, D., Miller, K., Baggett, C., Debold, E., & Freedson, P. S. (2000). Validity Of Accelerometry For The Assessment Of Moderate Intensity Physical Activity In The Field. *Medicine And Science In Sports And Exercise, 32,* S442-S449.

Hintzman, D. L. (1984). Episodic Versus Semantic Memory: A Distinction Whose Time Has Come - And Gone? *Behavioural And Brain Sciences, 7,* 240-241.

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.

Jakicic, J. M., Winters, C., Lagally, K., & Et Al. (1999). The Accuracy Of The Tritrac R3D Accelerometer To Estimate Energy Expenditure. *Medicine And Science In Sports And Exercise, 31,* 747-754.

Jobe, J. B., Pratt, W. F., Tourangeau, R., Baldwin, A., & Rasinski, K. (1997). Effects Of Interview Mode On Sensitive Questions In A Fertility Survey. In L.Lyberg, M. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey Measurement And Process Quality* (1st Ed., Pp. 311-329). New York: Wiley.

Jobe, J B , Tourangeau, R , & Smith, A F (1993) Contributions Of Survey Research To The Understanding Of Memory *Applied Cognitive Psychology, 7,* 567-584

Just, M A & Carpenter, P A (1992) A Capacity Theory Of Comprehension *Psychological Review, 99,* 122-149

Kalton, G & Schuman, H (1982) The Effect Of The Question On Survey Responses A Review *Journal Of The Royal Statistical Society, 145,* 42-73

Kalton, G & Schuman H (1982) The Effect Of The Question On Survey Response A Review *Journal Of The Royal Statistical Society, 145,* 42-73

Kelleher, C , Nic Gabhainn, S , Friel, S , Corrigan, H , Nolan, G , Sixsmith, J Et Al (2003) *Survey Of Lifestyles, Attitudes And Nutrition (SLAN)* Dublin Health Promotion Unit, Department Of Health And Children, Centre For Health Promotion Studies, NUI Galway, Department Of Public Health Medicine And Epidemiology, U C D

Kelley, D E & Goodpaster, B H (2001) Effects Of Exercise On Glucose Homeostasis In Type 2 Diabetes Mellitus *Medicine And Science In Sports And Exercise, 33,* S495-S501

Kolodner, J (1985) Memory For Experience In G H Bower (Ed), *The Psychology Of Learning And Motivation* (Pp 1-57) Orlando, FL Academic Press

Kriska, A & Caspersen, C J (1997) A Collection Of Physical Activity Questionnaires For Health-Related Research Medicine And Science In Sports And Exercise 29 (6), S3-S205

Krosnick, J A (1991) Response Strategies For Coping With The Cognitive Demands Of Attitude Measures In Surveys *Applied Cognitive Psychology, 5,* 213-236

Krueger, R A (1998a) *Analysing And Reporting Focus Group Results* (1st Ed) (Vols 6) London Sage Publications

Krueger, R A (1998b) *Moderating Focus Groups* (1st Ed) (Vols 4) London Sage Publications

Lamonte, M J & Ainsworth, B E (2001) Quantifying Energy Expenditure And Physical Activity In The Context Of Dose-Response *Medicine And Science In Sports And Exercise, 33,* 370-378

Lee, I -M & Paffenbarger, R S Jr (1996) How Much Physical Activity Is Optimal For Health? Methodological Considerations *Research Quarterly For Exercise And Sport, 67,* 206-208

Lee, I -M , Sesso, H D , & Paffenbarger, R S Jr (2000) Physical Activity And Coronary Heart Disease Risk In Men Does The Duration Of Exercise Episodes Predict Risk? *Circulation, 102,* 981-986

Leenders, N Y , Sherman, W M , Nagaraja, H N , & Kien, C L (2001) Evaluation Of Methods To Assess Physical Activity In Free-Living Conditions *Medicine And Science In Sports And Exercise, 33,* 1233-1240

Linton, M (1982) Transformations Of Memory In Everyday Life In U Neisser (Ed), *Memory Observed Remembering In Natural Contexts* (1st Ed , Pp 77-91) San Francisco Freeman And Company

Litwin, M (1995) *How To Measure Survey Reliability And Validity* (1st Ed) USA Sage Inc

Livingstone, M B E , Robson, P J , McCarthy, S , Kiely, M , Harrington, K , Browne, P Et Al (2001) Physical Activity Patterns In A Nationally Representative Sample Of Adults In Ireland *Public Health Nutrition, 4,* 1107-1116

Locke, L F (1989) Qualitative Research As A Form Of Scientific Inquiry In Sport And Physical Education *Research Quarterly For Exercise And Sport, 60,* 1-20

Louie, L , Eston, R G , Rowlands, A V , Tong, K K , Ingledew, D K , & Fu, F H (1999) Validity Of Heart Rate, Pedometry And Accelerometry For Estimating The Energy Cost Of Activity On Hong Kong Chinese Boys *Pediatric Exercise Science, 11,* 229-239

Lowther, M , Mutrie, N , Loughlan, C , & McFarlane, C (1999) Development Of A Scottish Physical Activity Questionnaire A Tool For Use In Physical Activity Interventions *British Journal Of Sports Medicine, 33,* 244-249

McCauley, R N (1984) Inference And Temporal Coding In Episodic Memory *Behavioural And Brain Sciences, 7,* 246-247

McKenna, J , Foster, L J , & Page, A (2004) Exploring Recall Of Physical Activity In Young People Using Qualitative Interviewing *Pediatric Exercise Science*

Means, B & Loftus, E (1991) When Personal History Repeats Itself Decomposing Memories For Recurring Events *Applied Cognitive Psychology, 5,* 297-318

Milne, R & Bull, R (2002) Back To Basics A Componential Analysis Of The Original Cognitive Interview Mnemonics With Three Age Groups *Applied Cognitive Psychology, 16,* 743-753

Montoye, H J , Kemper, H C G , & Sarris, W H M (1986) *Measuring Physical Activity And Energy Expenditure* (1st Ed) Champaign, IL Human Kinetics Publisher

Moon, J K & Butte, M F (1996) Combined Heart Rate And Activity Improve Estimates Of Oxygen Consumption And Carbon Dioxide *Journal Of Applied Physiology, 81,* 1754-1761

Morgan, D L (1998a) *Planning Focus Groups* (First Ed) USA Sage Publications

Morgan, D L (1998b) *The Focus Group Guidebook* USA Sage Publications

National Centre For Health Statistics (1991) National Health Interview Survey *Vital Health Statistics, 10,* 56-58

Nichols, J F , Morgan, C G , Sarkin, J A , Sallis, J F , & Calfas, K J (1999) Validity, Reliability And Calibration Of The Tritrac Accelerometer As A Measure Of Physical Activity *Medicine And Science In Sports And Exercise, 31,* 908-912

Paffenbarger, R S Jr , Blair, S N , Lee, I -M , & Hyde, R T (1993) Measurement Of Physical Activity To Assess Health Effects In Free-Living Populations *Medicine And Science In Sports And Exercise, 25,* 60-70

Paffenbarger, R S Jr & Lee, I -M (1994) Physical Activity And Personal Charateristics Associated With Depression And Suicide In American College Men *Acta Psychiatrica Scandinavica, 377,* 16-22

Paffenbarger, R S Jr , Wing, A L , & Hyde, R T (1978) Physical Activity As As Index Of Heart Attack Risk In College Alumni *American Journal Of Epidemiology, 108,* 161-175

Pate, R , Pratt, M , Blair, S N , Haskell, W L , Macera, C , Bouchard, C Et Al (1995) Physical Activity And Public Health A Recommendation From The Centre For Disease Control And Prevention And The American College Of Sports Medicine *Journal of the American Medical Association, 273,* 402-407

Patterson, P (2000) Reliability, Validity And Methodological Response To Assessment Of Physical Activity Via Self-Report *Research Quarterly For Exercise And Sport, 71,* 15-20

Philips, W , Pruitt, L , & King, A C (1996) Lifestyle Activity *Sports Medicine, 22,* 1-7

Pols, M A , Peeters, P H M , Kemper, H C G , & Grobbee, D E (1998) Methodologicl Aspects Of Physical Activity Assessment In Epidemiological Studies *European Journal Of Epidemiology, 14,* 63-70

Power, K , Martin, N , & Woods, C (2003) *The Validity Of Heart Rate And Motion Sensors In Measuring The Energy Used By Irish Adults During Activities Of Daily Living* B Sc Unpublished, Dublin City University

Presser, S & Blair, J (1994) Survey Pretesting Do Different Methods Produce Different Results? In P V Marsden (Ed), *Sociological Methodology* (1st Ed , pp 73-104) Beverley Hills, CA Sage

Presser, S , Couper, M , Lessler, J , Martin, E , Martin, J , Rothgeb, J M Et Al (2004) Methods For Testing And Evaluating Survey Questions In S Presser, J M Rothgeb, M Couper, J Lessler, E Martin, J Martin, & E Singer (Eds), *Methods For Testing And Evaluating Survey Questionnaires* (1st Ed , pp 1-22) New Jersey John Wiley And Sons Inc

Prochaska, J O & Marcus, B H (1994) The Transtheoretical Model Applications To Exercise In R K Dishman (Ed), *Advances In Exercise Adherence* (1st Ed , Pp 161-172) Champaign IL Human Kinetics

Prochaska, J O & Norcross, J C (1999) Comparative conclusions towards a transtheoretical therapy In Systems of Psychotherapy A Transtheoretical Analysis (4th Ed, p 487-538) Pacific Grove, CA Brooks/Cole Publishing

Reiser, B J , Black, J B , & Abelson, R P (1985) Knowledge Structures In The Organisation And Retrieval Of Autobiographical Memories *Cognitive Psychology, 17,* 89-137

Richardson, M T , Ainsworth, B E , Wu, H -C , Jacobs, D R Jr , & Leon, A S (1995) Ability Of The Atherosclerosis Risk In Communities (ARIC)/Baecke Questionnaire To Assess Leisure-Time Physical Activity *International Journal Of Epidemiology, 24,* 685-693

Richardson, M T , Leon, A S , Jacobs, D R Jr , Ainsworth, B E , & Serfass, R (1994) Comprehensive Evaluation Of The Minnesota Leisure Time Physical Activity Questionnaire *Journal Of Clinical Epidemiology, 47,* 271-281

Rowlands, A V , Eston, R G , & Ingledew, D K (1999) Relationship Between Activity Levels, Aerobic Fitness And Body Fat In 8-10 Year Old Children *Journal Of Applied Physiology, 86,* 1428-1435

Sage, G H (1989) A Commentary On Qualitative Research As A Form Of Scientific Inquiry In Sport And Physical Education *Research Quarterly For Exercise And Sport, 60,* 25-29

Sallis, J F , Haskell, W L , Wood, P D , Fortmann, S P , Rogers, T , Blair, S N Et Al (1985) Physical Activity Assessment Methodology In The Five-City Project *American Journal Of Epidemiology, 121,* 91-106

Sallis, J F & Saelens, B E (2000) Assessment Of Physical Activity By Self-Report Status, Limitations And Future Directions *Research Quarterly For Exercise And Sport, 71,* 1-14

Schuman, H & Ludwig, J (1983) The Norm Of Even Handedness In Surveys As In Life *American Sociological Review, 48,* 112-120

Schwarz, N (1999) Self-Reports How The Questions Shape The Answers *American Psychologist, 54,* 93-105

Shaughnessy, J J & Zechmeister, E B (1997) *Research Methods In Psychology* (4th Ed) Singapore McGraw-Hill Companies Inc

Smith, A F , Jobe, J B , & Mingay, D J (1991) Retrieval From Memory Of Dietary Information *Applied Cognitive Psychology, 5,* 269-295

Speakman, J R (2001) The History And Theory Of The Doubly Labelled Water Technique *American Journal Of Clinical Nutrition, 68,* 932S-938S

Starling, R D , Toth, M J , Carpenter, W H , Matthews, D E , & Poehlman, E T (1998) Energy Requirements And Physical Activity In Free-Living Older Women And Men A Doubly Labelled Water Study *Journal Of Applied Physiology, 85,* 1063-1069

Stephens, T , Jacobs, D R Jr , & White, C (1985) A Descriptive Epidemiology Of Leisure-Time Physical Activity *Public Health Reports, 100,* 147-158

Strack, F , Schwarz, N , & Gschneidinger, E (1985) Happiness And Reminiscing The Role Of Time Perspective, Affect And Mode Of Thinking *Journal Of Personality And Social Psychology, 47,* 1460-1469

Strath, S J , Bassett, D R , Thompson, D L , & Swartz, A M (2002) Validity Of The Simultaneous Heart-Rate Motion Sensor Technique For Measuring Energy Expenditure *Medicine And Science In Sports And Exercise, 34,* 888-894

Strath, S J , Bassett, D R Jr , Swartz, A M , & Thompson, D L (2001) Simultaneous Heart Rate-Motion Sensor Technique To Estimate Energy Expenditure *Medicine And Science In Sports And Exercise, 33,* 2118-2123

Strath, S J , Swartz, A M , Bassett, D R Jr , o'brien, W L , King, G A , & Ainsworth, B E (2000) Evaluation Of Heart Rate As A Method For Assessing Moderate Intensity Physical Activity *Medicine And Science In Sports And Exercise, 32,* S465-S470

Strauss, A & Corbin, J (1998) *Basics Of Qualitative Research Techniques And Procedures For Developing Grounded Theory* (2nd Ed) Thousand Oaks California Sage Publications Inc

Subar, A F , Thompson, F , Smith, A F , Jobe, J B , Ziegler, R , Potischman N Et Al (1995) Improving Food Frequency Questionnaires A Qualitative Approach Using Cognitive Interviewing *Journal Of The American Dietetic Association, 95,* 781-788

Sudman, S (1980) Reducing Response Error In Surveys *The Statistician, 29,* 237-273

Sudman, S & Bradburn, N (1982) *Asking Questions* San Francisco Jossey-Bass Inc Publishers

Sudman, S & Bradburn, N (1973) Effects Of Time And Memory Factors On Response In Surveys *Journal Of The American Statistical Association, 68,* 805-815

Taylor, C B , Coffey, T , Berra, K , Iaffaldano, R , Casey, K , & Haskell, W L (1984) Seven-Day Activity And Self-Report Compared To A Direct Measure Of Physical Activity *American Journal Of Epidemiology, 120,* 818-824

Taylor, H L , Jacobs, D R Jr , Schucker, B , Knudsen, J , Leon, A S , & Debacker, G (1978) A Questionnaire For The Assessment Of Leisure Time Physical Activities *J Chron Dis, 31,* 741-755

Thomas, J R & Nelson, J K (2001) *Research Methods In Physical Activity* (4th Ed) Champaign IL Human Kinetics

Thompson, C P , Skowronski, J J , Larsen, S F , & Betz, A L (1996) *Autobiographical Memory* (1st Ed) Mahwah, NJ Erlbaum

Thune, I & Furberg, A -S (2001) Physical Activity And Cancer Risk Dose-Response And Cancer, All Sites And Site-Specific *Medicine And Science In Sports And Exercise, 33,* 530-550

Tourangeau, R (1984) Cognitive Science And Survey Methods In T Jabine, M Straf, J Tanur, & R Tourangeau (Eds), *Cognitive Aspects Of Survey Design Building A Bridge Between Disciplines* (1st Ed , Pp 73-100) Washington, DC National Academy Press

Tourangeau, R , Rips, L J , & Rasinski, K (2000) *The Psychology Of Survey Response* (1st Ed) New York Cambridge University Press

Tudor-Locke, C , Henderson, K A , Wilcox, S , Cooper, R S , Durstine, J L , & Ainsworth, B E (2003) In Their Own Voices Definitions And Interpretations Of Physical Activity *Women's Health Issues, 13,* 194-199

Tudor-Locke, C & Myers, A M (2001) Challenges And Opportunities For Measuring Physical Activity In Sedentary Adults *Sports Medicine, 31,* 91-100

Tulving, E (1983) *Elements Of Episodic Memory* (1st Ed) Oxford Oxford University Press

Tulving, E (1984) Relations Among Components And Processes Of Memory *Behavioural And Brain Sciences, 7,* 257-263

U S Department Of Health And Human Services (1996) *Physical Activity And Health A Report Of The Surgeon General* Atlanta, GA US Department Of Health And Human Services

U S Department Of Health And Human Services (1999) *A New Agenda For Interdisciplinary Survey Research Methods* Hyattsville, MD USA U S Department Of Health And Human Services

Vincent, W J (1999) *Statistics In Kinesiology* (Second Edition Ed) USA Human Kinetics

Wagenaar, W A (1986) My Memory A Study Of Autobiographical Memory Over Six Years *Cognitive Psychology, 18,* 225-252

Wareham, N J , Hennings, S J , Prentice, A M , & Day, N E (1997) Feasibility Of Heart-Rate Monitoring To Estimate Total Level And Pattern Of Energy Expenditure In A Population-Based Epidemiological Study The Ely Young Cohort Feasibility Study 1994-5 *British Journal Of Nutrition, 78,* 889-900

Warnecke, R , Johnson, T , Chavez, N , Sudman, S , O'Rourke, D , Lacey, L Et Al (1997) Improving Question Wording In Surveys Of Culturally Diverse Populations *Ann Epidemiol, 7,* 334-342

Welk, G J , Blair, S N , Wood, K , Jones, S , & Thompson, R W (2000) A Comparative Evaluation Of Three Accelerometry-Based Physical Activity Monitors *Medicine And Science In Sports And Exercise, 32,* S489-S497

Welk, G J & Corbin, C B (1995) The Validity Of The Tritrac R3D Activity Monitor For The Assessment Of Physical Activity In Children *Research Quarterly For Exercise And Sport, 66,* 202-209

Williams, M D & Hollan, J D (1981) The Process Of Retrieval From Very Long-Term Memory *Cognitive Science, 5,* 87-119

Willis, G B (1999) Cognitive Interviewing A "How To" Guide (Personal communication)

Willis, G B (2004) Cognitive Interviewing Revisited A Useful Technique In Theory? In S Presser, J M Rothgeb, M Couper, J Lessler, E Martin, J Martin, & E Singer (Eds), *Methods For Testing And Evaluating Survey Questionnaires* (1st Ed , Pp 23-44) New Jersey John Wiley And Sons Inc

Willis, G B (2002) Cognitive Interviewing And Questionnaire Design -"Better Questions Are Ours For The Asking" (Personal communication)

Willis, G B & Schechter, S (1997) Evaluation Of Cognitive Interviewing Techniques Do The Results Generalize To The Field? *Bulletin De Methedologie Sociologique, 55,* 40-66

Wilmore, J H (2001) Dose-Response Variation With Age, Sex And Health Status *Medicine And Science In Sports And Exercise, 33,* 622-634

Winett, R A (1998) Developing More Effective Health-Behaviour Programs Analyzing The Epidemiological And Biological Bases For Activity And Exercise Programs *Applied And Preventive Psychology, 7,* 224