Body Mass Index and Accelerometer Measurement Issues for use in the Evaluation of Pedometer-based Physical Activity Interventions in Children

A.C. Routen

A thesis submitted in partial fulfilment of the University's requirements for the Degree of Doctor of Philosophy

2013

University of Worcester

Abstract

Participation in physical activity (PA) of at least moderate intensity may yield important health benefits for children. A popular behavioural tool used to promote increased PA is the pedometer. There is however limited evidence regarding pedometer-based strategies in children. This thesis reports on a series of anthropometric and accelerometer-measurement issue studies which inform the methods used to address the primary aim of this thesis- to determine the effectiveness of goal-setting, self-monitoring and step-feedback pedometer-based interventions for increasing PA in 10-11-year-old children. In addition, each study in their own right provides an original contribution to knowledge within their specific area of scholarship. The first objective of this thesis was therefore to determine diurnal variation of height and weight and the combined effect upon body mass index (BMI) weight status in children via a field based study. Next, the reliability of the Actiwatch 4 (AW4) accelerometer was tested in a mechanical laboratory experiment. Following this laboratory trial a second field based study examined the impact of placement site upon AW4 output, and the validity of a regression equation to predict hip-derived AW4 data from wrist-derived data. Finally, a brief intervention mapping approach was used to develop goal-setting, self-monitoring and step-feedback pedometer-based interventions, the effectiveness of which was evaluated in a small scale controlled trial involving two primary schools. The main findings of this thesis were a) that diurnal variation in height (and in girls alone, weight) impact upon increased BMI and BMI percentile in afternoon versus morning measurements b) AW4 activity counts exhibit acceptable reliability statistics (comparable to other accelerometer models), which improve when raw activity counts are reduced into derived activity intensity variables c) wrist and hip derived AW4 data are not comparable, and the derived regression equation may not be suitable for group level prediction due to inaccurate individual level prediction and the large standard error of the estimate observed d) pilot testing pedometer wear and intervention materials may highlight practical pedometer issues (i.e. pedometer attachment, wearing compliance and acceptability of instruction sheets) that inform

intervention implementation and e) pedometer-based goal-setting, self-monitoring and step-feedback interventions did not increase PA in 10-11-year-old children. However, individual-standardised goal setting may be more promising as this appeared to mitigate any decline in moderate-to-vigorous physical activity (MVPA) in *more-active* children, and increased MVPA in *less-active*. To summarise, the findings of this thesis highlight important issues for physical activity scientists to consider when using BMI-determined weight status as a grouping variable and accelerometers as an outcome measure, when evaluating physical activity interventions in children. With regard to the primary aim of this thesis, future researchers should further examine the effectiveness of the individual-standardised against the group-standardised goal type in a longer-duration intervention and using a larger sample of children, which may permit sub-group analyses to be conducted. Of primary importance is future clarification on the effectiveness of goal setting, self-monitoring and step-feedback pedometer-based interventions *per se* for changing PA in children.

Acknowledgements

I have a number of people to thank for their informal support, advice and friendship over the past 4 years. In no particular order, thanks to Rich, Rach, Big Col, Jenni, Paul C and the inhabitants of the Graduate Research School corridor.

I am lucky to have made some superb friends during my time at Worcester. Thanks to 'The Lads', that is Ben, Chris and Geoff, for accompanying me on various cycling, walking, eating and drinking endeavours. Particular mention must go to Geoff, Neil and Rich for providing motivation when mine had all but ebbed away … A 'chapeau' must go to Mark Corbett and Joe Bevins for their technical support, as well as continual bike maintenance - I greatly enjoyed the cycling news banter and lunchtime rides.

Thank you to Professor Dominic Upton and Dr Martin Edwards for their guidance when called upon. My Director of Studies, Professor Derek Peters, has invested many hours of time, alongside personal research funding to provide me with the opportunity to complete this research training. Thanks to Derek for his patient guidance and for helping me develop key skills that will stand me in good stead for the future.

Most importantly, the biggest thanks should go to my parents, Rod and Marion. Thank you for not giving up on me all those years ago and for everything you have done for me to date. Getting to this submission stage is as much your 'achievement' as it is mine, and as the great alpinist Ed Hillary said of Everest *"Well, we knocked the bastard off"*.

"The first wealth is health"

Ralph Waldo Emerson (American poet, lecturer and essayist, 1803-1882).

"I don't think much of a man who is not wiser than he was yesterday"

Abraham Lincoln (16th president of the United States, 1809-1865).

Declaration

I declare that this thesis is a presentation of my own original research work and all the written work and investigations are entirely my own. Wherever contributions of others are involved, this is clearly acknowledged and referenced.

I declare that no portion of the work referred to in this thesis has been submitted for another degree or qualification of any comparable award at this or any other university or other institution of learning.

Signed:

Date: 05/07/2013

Table of Contents

List of Tables

List of Figures

List of Abbreviations/Units

APE	Absolute percent error
AW	Actiwatch accelerometer
AW4	Actiwatch 4 accelerometer
BCT	Behaviour change technique
BMI	Body mass index
CARS	Children's activity rating scale
CI	Confidence interval
cts	Accelerometer activity count
CV	Coefficient of variation
d	Cohen's *d* effect size statistic
DH	Dominant Hip
DOH	Department of Health, UK
DW	Dominant Wrist
EE	Energy expenditure
f	Critical *f* statistic
FM	Fat mass
hr	Hour
ICC	Intra-class correlation coefficient
kcal	Kilogram calorie
kg	Kilogram
kj	Kilojoule
MET	Metabolic equivalent ratio
min	Minute
MVPA	Moderate-to-vigorous physical activity
NDH	Non-dominant Hip
NDW	Non-dominant Wrist
p	Probability value
PA	Physical activity
PAEE	Physical activity energy expenditure
r	Correlation coefficient
r^2	Coefficient of determination
ROC	Receiver operating characteristic
s	Second
SB	Sedentary behaviour
t	Critical *t* statistic
VO_2	Oxygen consumption

Thesis Research Dissemination

Peer-reviewed publications (see Appendix Five):

Routen, A.C., Upton, D., Edwards, M.G., & Peters, D.M. (2012). Discrepancies in accelerometer-measured physical activity in children due to cut-point non-equivalence and placement site. *Journal of Sports Sciences*, *30*(12), 1303-1310.

Routen, A.C., Upton, D., Edwards, M.G., & Peters, D.M. (2012). Intra- and inter-instrument reliability of the Actiwatch 4 accelerometer in a mechanical laboratory setting*. Journal of Human Kinetics*, *31*, 17-24.

Routen, A.C., Edwards, M.G., Upton, D., & Peters, D.M. (2011). The impact of school-day variation in weight and height on National Child Measurement Programme body mass index-determined weight category in Year 6 children. *Child Care, Health and Development, 37*(3), 360-367.

Peer-reviewed conference presentations:

Routen, A.C., Upton, D., Edwards, M.G., & Peters, D.M. (2011, July). *Intra- and inter-instrument reliability of the Actiwatch 4 accelerometer in a mechanical laboratory setting*. Poster presentation at the 16th Annual Congress of the European College of Sport Science, Liverpool John Moores University, UK.

Routen, A.C., Upton, D., Edwards, M.G., & Peters, D.M. (2011, July). *Accelerometer measured physical activity differs between trunk and limb placements, but not side dominance in children*. Poster presentation at the 16th Annual Congress of the European College of Sport Science, Liverpool John Moores University, UK.

1 Chapter One - Thesis Background

1.1 Introduction

Evidence for the health benefits of physical activity (PA) in adults has been apparent for some time (Morris, Heady, Raffle, Roberts, & Parks, 1953; Paffenbarger, Laughlin, Gima, & Black, 1970). Previous reviews have, however, highlighted the lack of direct support for the same assertion in children (Biddle, Gorely, & Stensel, 2004; Boreham & Riddoch, 2001). Since the publication of these reviews the deployment of objective measures of PA has yielded new data showing inverse, and in some cases independent cross-sectional (and/or prospective) relations between moderate-to-vigorous PA (MVPA), body fatness (Ness et al., 2007; Riddoch et al., 2009) and cardiometabolic risk factors (Andersen et al., 2006; Ekelund et al., 2012). Additionally there is evidence in children that sedentary behaviour (SB) is independently associated with lifestyle disease risk factors including insulin resistance (Sardinha et al., 2008) and lower cardiorespiratory fitness (Mitchell, Pate, & Blair, 2012).

Founded upon this mounting epidemiologic evidence there is now international consensus that children should accumulate ≥ 60 min of MVPA per day to maintain healthy growth and development and achieve psychological and physiological health benefits (CDC, 2011; CSEP, 2011; DOH, 2011; WHO, 2010). Estimates of children engaging in ≥ 60 min of MVPA in the UK range from ~2 to ~70% according to large sample (*n* > 1000) accelerometer studies (Owen et al., 2009; Riddoch et al., 2007; Steele et al., 2010). A recent review of international data (self-report and accelerometer measured) concluded that ~30-40% of children and youth are sufficiently active according to the aforementioned recommendations (Ekelund, Tomkinson, & Armstrong, 2011). Further, data also indicate that the prevalence of overweight and obesity is increasing (NHS Information Centre, 2011), and levels of cardiorespiratory fitness are decreasing (Stratton et al., 2007) in primary-school-aged children (5-11 years) in the UK.

Presently there is limited knowledge upon the relations between childhood PA and protection from morbidity (i.e. Type II diabetes, cardiovascular disease etc.) and premature mortality (i.e. death) in adulthood (Biddle et al., 2004; Rowland, 2007). Nonetheless the development of risk factors for some hypokinetic lifestyle diseases is initiated in late childhood/adolescence, such as the formation of atherosclerotic plaques which could develop into overt coronary heart disease in later life (Strong et al., 1999). It is therefore logical to suggest that participation in recommended levels of MVPA during childhood may retard the development of disease risk factors and in tandem increase the likelihood of becoming a more active adult, which could confer protection against chronic disease during adulthood (Rowland, 2007).

Within childhood, the period of late primary school years (10-11 years) is a vital stage for PA promotion. Large sample observational studies have shown that total PA and MVPA decline markedly from ~9 years of age, with this decline continuing throughout adolescence into young adulthood (Kimm et al., 2002; Nader, Bradley, Houts, McRitchie, & O'Brien, 2008). On this basis, with rising levels of overweight and obesity, and with the potential health benefits of regular participation in PA, the promotion of PA in primary-school-aged children is clearly an issue of increasing importance.

A behavioural tool, which is increasingly popular with public health professionals and exercise scientists, and has been shown to effectively promote PA across a wide range of populations is the pedometer. Indeed a recent review of systematic reviews published in *The Lancet* presented pedometer-based interventions as displaying the largest effect (0.68) size when compared to estimates derived for different settings (e.g. workplaces), populations (e.g. healthy adults) and intervention types (e.g. after school) (Heath et al., 2012). These effect size estimates are displayed in Figure 1.1 below.

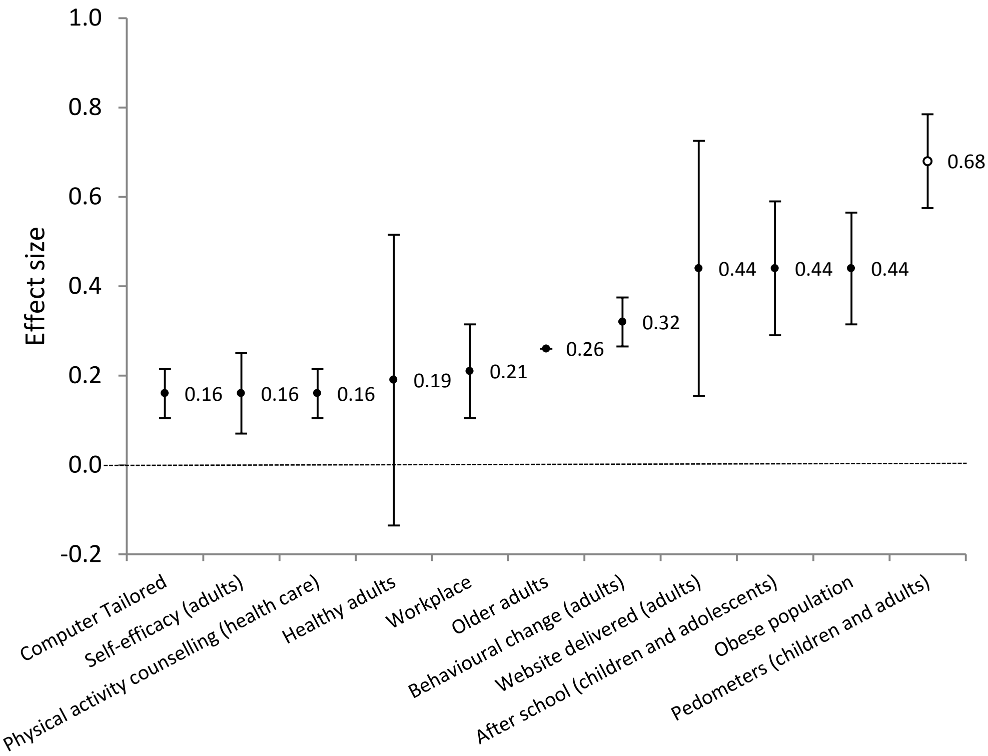


Figure 1.1 Mean effect-size estimates (and 95% CIs) from systematic reviews of physical activity interventions by population, setting and intervention type. Pedometer interventions are displayed in open white marker. Adapted from Heath et al. (2012).

In 2007 evidence from a systematic review of adult studies showed that pedometer use was associated with significant increases in daily steps (steps.day^{-1}) over baseline (+26.9%), as well as decreases in body mass index (BMI) and systolic blood pressure (Bravata et al., 2007). In 2009, a systematic review of youth literature concluded that whilst pedometer-based interventions have generally had positive effects on steps. day^{-1} (12 of 14 studies reviewed showed increases), there is a distinct paucity of research regarding pedometer-based interventions in children (Lubans, Morgan, & Tudor-Locke, 2009b).

Data from systematic reviews of both adult and youth populations (Bravata et al., 2007; Kang, Marshall, Barreira, & Lee, 2009a; Lubans et al., 2009b) show that the key motivational factors for increasing PA using pedometers include setting a daily step goal, viewing current steps (step feedback) and self-monitoring daily step totals using a step diary or chart. Most pedometer-based interventions in children, for example, have used one, or a combination of, self-monitoring, step feedback and step goal setting alongside additional behavioural components such as behavioural counselling, peer modelling and tangible rewards (Berry, Savoye, Melkus, & Grey, 2007; Hardman, Horne, & Lowe, 2011; Horne, Hardman, Lowe & Rowlands, 2009). The inclusion of additional behavioural strategies has restricted understanding regarding the optimal deployment of pedometers to promote PA due to the masking of independent contributions of self-monitoring, step feedback, and step goal components. Only a single study has explicitly examined the independent contribution of self-monitoring, step feedback and step-goal setting in children (Kang & Brinthaupt, 2009). This data is important as it allows judgements to be made upon the utility of pedometer strategies alone for promoting physical activity in children.

The types of step goals that can be set by practitioners/researchers include self-set (in which the user sets their own goal, also known as tailored or personalised), individual-standardised (i.e. increment above individual baseline), group-standardised (i.e. increment above group baseline) or pre-set universally administered (i.e. 10,000 $steps.day^{-1}$) goals (Tudor-Locke & Lutes, 2009). A study in inactive adult women found no differences in $steps.day^{-1}$ between self-set and universal goal groups (Sidman, Corbin, & Le Masurier, 2004). A meta-analyses (comprising mixed adult and child data) by Kang et al. (2009a) reported a marginally greater positive effect size (0.84 vs. 0.72) for individuals administered universal (10,000 $steps.day^{-1}$) compared to individualised step goals. As stated above, only one study has compared step goal types in children, reporting no difference in $steps.day^{-1}$ between individual-standardised and group-standardised goal conditions (Kang & Brinthaupt, 2009). No further studies have examined the relative effectiveness of goal type in children; therefore the optimal step goal for increasing PA remains uncertain.

When developing physical activity initiatives such as pedometer interventions, Tudor-Locke, Myers and Rodger (2001) recommend they be developed using a form of 'intervention theory', which involves the identification and description of the planning/development and evaluation of a behavioural intervention to ensure researchers can explain what happens in an intervention and why (Tudor-Locke, Myers, & Rodger, 2001). A commonly used 'intervention theory' approach in behavioural interventions is the intervention mapping (IM) procedure (Bartholomew, Parcel, & Kok, 1998). Intervention mapping sets out six steps of an iterative process which includes the development, implementation and evaluation of an intervention. Key stages in this process include (1) a needs assessment, (2) identification of outcomes and behaviour change objectives, (3) identification of theoretical models and practical strategies to induce behaviour change, (4) programme development and pilot-testing, (5) development of an implementation and adoption plan and (6) development of an evaluation plan. No previous pedometer-based studies in children have reported the use of an 'intervention theory' to guide the development and evaluation of the interventions implemented.

When planning intervention evaluation consideration should also be given to the identification of potential moderating variables; that is variables which may moderate the main intervention effect (e.g. sex, age etc.), and therefore confound conclusions from intervention data. There have been calls within the pedometer literature for future researchers to determine effects on children of differing weight status groups, i.e. *healthy weight*, *overweight* and *obese* (Horne et al., 2009; Tudor-Locke & Lutes, 2009). A common field-based method for estimating weight status is the measurement of weight and height to determine BMI. Despite the widespread use of BMI to estimate weight status, little attention is given to associated measurement error, which could be related to a number of factors including researcher reliability, equipment variation and the use of standardised measurement protocols. Measurement protocols reported in PA interventions typically do not make mention of standardisation of the time of day at which measurements are taken. To date, the extent to which diurnal fluctuation in weight and height may impact on classification of BMI-determined weight status in children has not been studied. If diurnal variation in weight and height influence BMI-determined weight status categorisation, intervention effect estimates in weight status groups could be biased, and comparisons of these effects between intervention groups would be unfair if measurement timing is unstandardised. Examining differential classification of weight status due to the timing of anthropometric measurements will yield important information for the PA researcher using BMI-determined weight status as a moderating variable.

A further consideration when planning intervention evaluation is the identification of a reliable and valid outcome measure. Past pedometer-based interventions have used a pedometer as the outcome tool (Lubans et al., 2009b). However pedometers only provide a volumetric index of PA (i.e. daily step counts) and as such give no indication of the frequency, intensity, or duration of PA (Welk, Corbin, & Dale, 2000). For example, a pedometer cannot discriminate between steps taken whilst slow walking (i.e. light intensity) compared to steps taken whilst running (i.e. vigorous intensity), of which these different intensities of locomotion may be differentially related to health. In addition, dissimilar to accelerometer-based data, there is a paucity of research linking daily steps to various health outcomes in children (Tudor-Locke et al., 2011). In light of the aforementioned benefits of regular participation in MVPA an estimate of MVPA should be considered important when measuring intervention effectiveness in children. The use of accelerometer-based physical activity monitors as an alternative outcome measure may therefore provide requisite data upon this health-enhancing intensity category.

Wrist-worn accelerometers such as the Actiwatch 4 (AW4, Cambridge Neurotechnology Ltd, UK) can provide an objective estimate of MVPA. The AW16 (previous generation of the AW which is analogous to AW4 in all but memory capacity) has been value calibrated against energy expenditure at right-hip (Puyau, Adolph, Vohra, & Butte, 2002; Puyau, Adolph, Vohra, Zakeri, & Butte, 2004) and lower-right-leg placements (Puyau et al., 2002) in children. However for a measurement device to be valid it must first be shown to have a degree of reliability. As yet there are currently no investigations that have explicitly examined the reliability of the AW4 in a child mounted or mechanical reliability experiment. Whilst the AW has been calibrated at trunk- and leg-placement sites, it was originally designed to be worn on the wrist. As data quality is in part dictated by participant compliance to monitoring protocols, wrist-placed devices are of interest as they are inherently less obtrusive than trunk placements and thus may increase participant compliance (Esliger et al., 2011).

The wrist placement has not been utilised in previous AW validation studies in children (Puyau et al., 2002; Puyau et al., 2004). Full-value calibration is consequently required before use of this placement site in future. Alternative to value calibration in the laboratory, a more practical field-based approach would be the characterisation of wrist-to-hip differences which could provide opportunity to build regression models to predict hip-derived intensity minutes (i.e. MVPA min) from wrist data, negating the need for laboratory calibration. Potential placement site differences in accelerometer output have been examined between the hip and lower back in children and adults (Nilsson, Ekelund, Yngve, & Sjostrom, 2002; Yngve, Nilsson, Sjostrom, & Ekelund, 2003). However, no studies have explicitly examined differences in accelerometer output between the wrist and hip placement sites in any population. Thus knowledge of both device reliability and placement site effects must be considered to permit judgement upon the AW4's suitability for use as a PA outcome measure.

As a relatively emergent field few pedometer-based interventions have been undertaken with children. Consequently the optimal pedometer programming 'template' is unknown. Of the available data, inconsistency in study design and the inclusion of additional behavioural strategies cloud understanding of the most effective pedometer strategies for PA promotion in children. It therefore appears appropriate to investigate measurement, development and deployment-related issues associated with pedometer-based PA interventions in children, which is an under-explored line of enquiry within the field of PA epidemiology.

1.2 Thesis aim

To determine the effectiveness of goal-setting, self-monitoring and step-feedback pedometer-based interventions for increasing physical activity in 10-11-year-old children.

1.3 Thesis objectives

1. Determine diurnal variation of height and weight and the combined effect upon BMI-determined weight status in 10-11-year-old children.
2. Determine the intra- and inter-instrument reliability of the Actiwatch 4 accelerometer.
3. Determine the effect of Actiwatch 4 accelerometer position upon physical activity in 10-11-year-old children.
4. Systematically develop goal-setting, self-monitoring and step-feedback pedometer-based interventions for increasing physical activity in 10-11-year-old children.
5. Determine the impact of goal-setting, self-monitoring and step-feedback pedometer-based interventions on physical activity in 10-11-year-old children.

1.4 Thesis structure

The primary aim of this thesis is to examine the impact of goal-setting, self-monitoring and step-feedback pedometer-based interventions for increasing PA in children aged 10-11 years. To achieve this aim and the aforementioned objectives, this thesis has been structured as follows: Chapter Two contextualises the research questions within current scholarship, and, in particular, characterises the complexities of PA, current prevalence/secular trends of PA, PA measurement methods, pedometer-based intervention studies and the use of BMI to determine weight status. This review also highlights the current paucity of literature regarding pedometer-based interventions, and particular measurement issues associated with accelerometer-determined PA in children. Chapter Three presents the anthropometric measurement methods used in this thesis, details the technical specification of the PA monitors (i.e. accelerometer and pedometers) used, and gives an overview of part of the PA data reduction process. Chapter Four subsequently draws on the results of a field-based study to examine the impact of diurnal variation in weight and height on BMI-determined weight status. Chapter Five is a mechanical laboratory investigation of the reliability of the AW4 accelerometer. Chapter Six presents a field-based study that investigates the impact of placement site upon AW4-measured PA. Phase One of Chapter Seven reports on the development of two pedometer-based interventions using the intervention mapping (IM) framework. Phase Two of Chapter Seven presents an experimental study examining the impact of the two goal-setting, self-monitoring and step-feedback interventions, developed in Phase One, upon AW4-measured PA in children. Lastly in Chapter Eight the findings from Chapters Four to Seven are discussed in relation to the thesis aims, recommendations for future work are given and conclusions are drawn.

2 Chapter Two - Literature Review

2.1 Introduction

This chapter reviews literature relevant to PA and children, notably the definition of PA, its associated health benefits, recommended amounts, prevalence, secular trends, measurement, and finally promotion using pedometer programming. For specificity to the population under study in this thesis, literature pertaining to primary-school-aged children aged 5-12 years predominates.

2.2 Physical activity and energy expenditure

Physical activity is any movement behaviour that is associated with an increase in energy expenditure (EE) above basal levels, resulting from skeletal muscle contraction (Howley, 2001). It is important to stress that PA and EE are not the same. PA is a behaviour that results in EE, whilst EE is the associated energy cost of PA (Lamonte & Ainsworth, 2001). This is PA described in its most reductionist sense. It is in fact a multi-dimensional and complex behaviour than can be characterised in numerous ways. The three dimensions of PA are the frequency, intensity and duration (the product being the total volume of PA). Frequency can be defined as the number of activity sessions per time unit (e.g. three times per week). The duration is the length of time of each individual session (e.g. 60 min) and the intensity is the physiologic effort of PA, often expressed as a metric of EE (Kesaniemi et al., 2001). PA may also be characterised by the type or mode used to perform the behaviour e.g. walking or cycling (Corder, Ekelund, Steele, Wareham, & Brage, 2008). Further, PA can be partitioned into components such as leisure time PA, that is activity that takes place in one's free time, and is usually purposeful (i.e. structured exercise); occupational PA (i.e. PA carried out as part of one's occupation) and lifestyle-embedded PA (i.e. incidental walking, unstructured play etc.) (Howley, 2001; Tremblay, Esliger, Tremblay, & Colley, 2007).

Physical activity researchers ordinarily use absolute measures of oxygen uptake (e.g. VO_2 $L.min^{-1}$) or EE (e.g. $kcal.min^{-1}$) to characterise the intensity of PA. To account for differences in body size these values are expressed relative to body mass or fat-free mass (e.g. VO_2 $ml.kg.min^{-1}$ and $kcal.kg.min^{-1}$). A commonly used metric of EE/intensity is the metabolic equivalent ratio (MET) (Schutz, Weinsier, & Hunter, 2001). METs are defined as the ratio-of-work metabolic rate to a standard-resting metabolic rate (1 MET, which in adults is usually taken as 3.5 ml of $O_2.kg.min^{-1}$) (Schutz et al., 2001). METs are often used in the field to denote the intensity level (e.g. light, moderate etc.) or predict the EE of observed or self-reported PA. MET values are assigned using MET compendia, which provide MET values for >600 different activities, based on adult-derived VO_2 data (Ainsworth et al., 2000). Caution must be taken when using these in children however as mass-related work is greater in children than adults *e.g.* moving a box of fixed weight requires greater EE for a child compared to an adult. A child-specific MET compendium has been published (Ridley, Ainsworth, & Olds, 2008), however 65% of the MET values in this are adult derived and their accuracy in children is unknown. The use of actual-measured resting metabolic rate (as opposed to the adult standard), alongside the child compendia values may however produce acceptable estimates of free-living EE (Harrell et al., 2005).

2.3 Sedentary behaviour

There is growing interest in research related to the health impact of sedentary behaviour (Tremblay, Colley, Saunders, Healy, & Owen, 2010). Evidence now shows that increased time spent sedentary (e.g. sitting/reclining) is associated with deleterious health in children and adolescents regardless of engagement in MVPA and other confounders (Mitchell et al., 2012; Sardinha et al., 2008; Tremblay et al., 2011). Currently there is inconsistency in the use of the term 'sedentary behaviour' in the 'exercise science' literature. Traditionally 'exercise scientists' have referred to sedentary behaviour as an absence of recommended levels of MVPA (i.e. not meeting PA guidelines), and consequently have classified research participants as 'sedentary' on this basis. This categorisation of sedentary behaviour as an absence of sufficient MVPA negates the contribution of light-intensity PA and sedentary behaviour to health. However with the appreciation that sedentary behaviour now represents an independent risk factor for chronic lifestyle diseases a recent review paper has attempted to re-conceptualize the low end of the PA spectrum (Tremblay et al., 2010). Tremblay and colleagues (2010) have provided a frame of reference called the 'movement continuum' (Figure 2.1) which illustrates the differences in energy cost (METs) and absolute intensity descriptor that separate sedentary and exercise physiology.

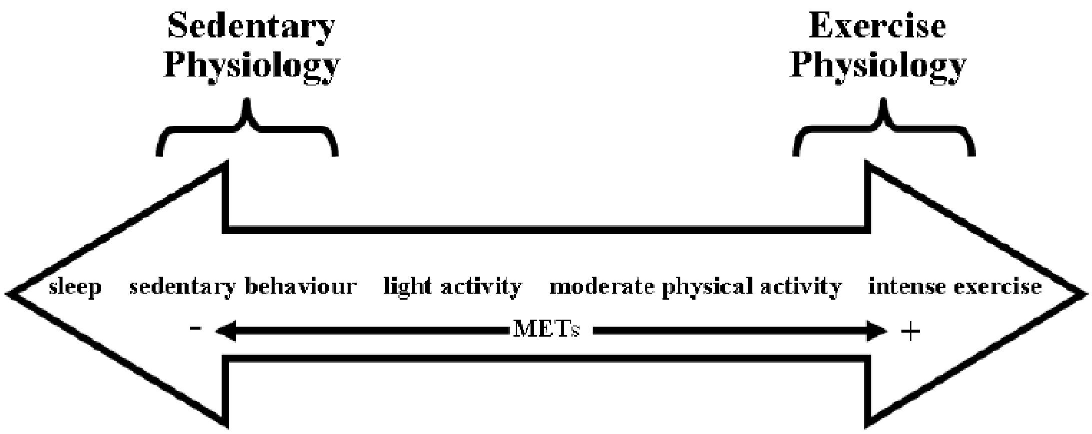


Figure 2.1 The movement continuum, showing the range of METs and intensity categories between sedentary and exercise physiology (Tremblay et al., 2010).

In early 2012 the Sedentary Behaviour Research Network (SBRN) published a standardised definition of the term 'sedentary' and 'sedentary behaviours' (SBRN, 2012). Sedentary behaviour is defined by the SBRN as '*any waking behaviour characterised by an energy expenditure ≤1.5 METs while in a sitting or reclining posture*'. Therefore the term 'physical inactivity' or 'inactive' is suggested to be used to describe individuals who do not meet recommended levels of MVPA (SBRN, 2012).

2.4 Physical activity and health in youth

From the middle of the 20th century research evidence has accumulated to show that frequent participation in PA in adult life has beneficial effects on health (Blair et al., 1989; Morris et al., 1953; Paffenbarger, Hyde, Wing, & Hsieh, 1986; Paffenbarger et al., 1970). The long-held links between habitual PA in adults and health status in adulthood were previously believed to be unconfirmed in children (Boreham & Riddoch, 2001). Following this review by Boreham and Riddoch (2001) however, a substantial review of evidence relating PA to health outcomes in children and youth in 2005 found evidence that PA has beneficial effects on aerobic fitness, cardiometabolic risk factors and body fatness (Strong et al., 2005). Furthermore in the last half decade (2005-) studies utilising objective measures of PA (mainly accelerometers) have shown inverse associations between participation in PA and a selection of health-related outcomes such as obesity (Ness et al., 2007; Riddoch et al., 2009), and metabolic risk (Andersen et al., 2006; Ekelund et al., 2012). However, the evidence to promote PA for health is not unequivocal and findings often differ between studies. This is not surprising as the clinical manifestations of chronic lifestyle disease are generally absent in childhood (Biddle et al., 2004). Therefore where disease end-points are commonly used in adults to establish PA and health relationships, disease risk markers are used in children. This has made it particularly hard to highlight the deleterious effect of physical inactivity in youth.

The indistinct PA-health relations are compounded by the use of many different PA measurement techniques within the literature. Further, relations between health and habitual PA are more complex in children than adults, with the difficulty of accurate PA assessment (Corder et al., 2008), lack of disease outcomes in children and mediating effect of physical fitness (Boreham & Riddoch, 2001) inhibiting clear conclusions being drawn. Despite this discord, the rationale for promoting PA in childhood can be built on the premise that PA may track into adulthood, although tracking coefficients for this phenomenon have been shown to be moderate at best (Trudeau, Laurencelle, &

Shephard, 2004). As Rowland (2007) purports however, the potential benefits of PA may manifest later in an individual's life:

> *An alternative approach focuses on the concept of promoting physical activity in children not as a means of reducing risk indicators, but rather as a precursor to adult activity. That is, the goal is to establish early habits of regular exercise that will carry over to the adult years. A lifetime of habitual exercise would then be expected to pay dividends in reduced health risk*.

Despite the lack of clarity upon the proposed benefit of PA on health in children, a recent review of 86 studies also found dose-response relations between both the volume (and to a lesser extent intensity) and a variety of health outcomes in children (Janssen & Leblanc, 2010). The most substantive health benefits were derived from engagement in PA of at least moderate intensity (Janssen & Leblanc, 2010), indicating that engagement in MVPA is a key outcome when studying children's PA behaviour. This assertion is supported by findings that engagement in MVPA is negatively associated with objectively measured fat mass (Ness et al., 2007; Riddoch et al., 2009) and clustered cardiometabolic risk (Andersen et al., 2006). On the basis of evidence such as this, and with the potential of future health benefits (Rowland, 2007) a number of expert panels and organisations have produced guidelines for recommended participation in PA for children and adolescents focused on ≥ moderate activity.

2.5 Youth physical activity guidelines

Physical activity guidelines offer evidence-based PA goals for the public, which are associated with a reduced risk of chronic disease and mortality. The guidelines are framed by a recommended 'dose' of PA, informing individuals how much, how frequently and how hard they should engage in PA (Brawley & Latimer, 2007). The first of such guidelines were issued by the American College of Sports Medicine (ACSM) in 1988, which stated that children and youth should attain 20-30 min of vigorous intensity exercise each day. This original position has changed over time however, with greater focus upon the benefit of at least moderate intensity activity in children. It can be observed from Table 2.1 that current recommendations suggest that children and youth should participate in ≥60 min per day of MVPA on each day of the week.

2.5.1 Guideline compliance

The latest Health Survey for England (HSE) 2008 reported that 32% of boys and 24% of girls in England aged 2-15 years were achieving 60 min of MVPA on all 7 days of the week according to a researcher-led interview (Department of Health, 2008). The validity of data from the HSE study is arguably questionable, as in 130 6-7 year olds the mean bias between similar 2004 HSE questionnaire-derived data compared to accelerometer-measured data was +122 min of MVPA per day (Basterfield et al., 2008). A number of observational studies have employed accelerometers to estimate compliance with the guidelines in the UK. Riddoch et al (2007) found that 2.5% of 5595 children aged 10-11 years (5.1% boys, 0.4% girls) engaged in ≥60 min of MVPA (weekly average). In contrast Steele et al. (2010) found that 81% of boys and 60% of girls (*n* = 1568, 9-10 years) engaged in ≥60 min of MVPA (weekly average). Similarly Owen et al. (2009) reported high compliance, with 64% of boys and girls aged 9-10 years (*n* = 2071) achieving ≥60 min (on a single day). Elsewhere internationally, Colley et al. (2011) reported that 80% of Canadian children and youth (*n* = 1608) aged 6-19 years engage in ≥60 min on at least 1 day, when this was raised to at least 6 of 7 days only 6.7% (9.0% boys, 4.1 % girls) met the guidelines.

It is evident that boys show greater compliance with the PA recommendations than girls, which is consistent with data showing boys to be more active than girls (Riddoch et al., 2007). Of note is the clear discord between studies in estimates of the percentage of children meeting the guidelines. This disagreement may be due to the differing populations studied (i.e. distribution of age, gender, ethnicity), but will largely be the result of dissimilar data reduction procedures, and how guideline compliance is operationalised. For example Riddoch et al. (2007) used an accelerometer count cut-point of 3600 $cts.min^{-1}$ in the Actigraph to define MVPA, whilst Steele et al. (2010) and Owen et al. (2009) used ~2000 $cts.min^{-1}$ (currently believed to be too low (see (Guinhouya, Hubert, & Zitouni, 2011; Guinhouya, Lemdani, Vilhelm, Durocher, & Hubert, 2009)) which could account for the higher prevalence of sufficiently active children reported in these studies.

As noted above guideline compliance can be operationalised in many ways which can affect compliance estimates (e.g. average weekly MVPA ≥60 min or ≥60 min on each day of the week - which is a more strict representation of the guidelines) (Esliger & Tremblay, 2007). To illustrate this point, a child could for example accrue most of their weekly MVPA at the weekend, but when this is averaged over 7 days they would erroneously appear to be sufficiently active i.e. ≥60 min MVPA per day. Until a standardised computation of MVPA using objective measures is obtained, the prevalence of children meeting PA guidelines will vary considerably between research groups.

Table 2.1 Physical activity guidelines for children and youth. Updated and adapted from (Janssen, 2007).

Organisation/Body	Country	Title of Guideline	Reference	Year	Age range	Recommendations
American College of Sports Medicine	United States of America	Opinion statement on physical fitness in children and youth	American College of Sports Medicine (1998)	1988	Children and youth	Obtain 20 to 30 min of vigorous exercise each day
International Consensus Conference on Physical Activity Guidelines for Adolescents	United States of America	Physical activity guideline for adolescents: consensus statement	Sallis and Patrick (1994)	1994	11-21 years	Be physically active daily, or nearly daily, as part of play, games, sports, work, transportation, recreation, physical education, or planned exercise; engage in ≥3 sessions per week of moderate to vigorous activities that last ≥20 min
US National Institutes of Health	United States of America	Consensus development panel on physical activity and cardiovascular health	NIH Consensus Development Panel on Physical Activity and Cardiovascular Health (1995)	1995	All	Accumulate 30 min of moderate physical activity on most, preferably all, days of the week
US Surgeon General	United States of America	Physical activity and health	US Department, Health & Human Services (1996)	1996	≥2 years	Accumulate 30 min of moderate physical activity on most, preferably all, days of the week
UK Health Education Authority	United Kingdom	Young people and health-enhancing physical activity: evidence and implications	Biddle et al. (1998)	1998	Children and youth	Participate in physical activity that is of at least a moderate intensity for an average of 1 hr per day; participate in physical activities that enhance and maintain strength in the musculature of the trunk and upper arm

						girdle ≥2 days per week; the above recommendation should be met by participating in developmentally appropriate activities
Australia Department of Health and Ageing	Australia	National physical activity guidelines for Australians	Department of Health and Ageing (1999)	1999	5-18 years	At least 60 min, and up to several hours, of moderate to vigorous physical activity every day; limit screen time ≤2 hr per day
UK Expert Consensus Conference	United Kingdom	Health-enhancing physical activity for young people: statement of the UK Expert Consensus Conference	Cavill et al. (2001)	2001	5-18 years	Participate in 60 min of at least moderate intensity physical activity every day, those who currently participate in little physical activity should attain at least 30 min of physical activity of at least moderate intensity
American Cancer Society	United States of America	Guidelines on nutrition and physical activity for cancer prevention	Byers et al. (2002)	2002	Children and youth	Engage in ≥60 min per day of moderate to vigorous physical activity at least 5 days per week
Health Canada and the Canadian Society for Exercise Physiology	Canada	Canada's physical activity guide for children and youth	Health Canada and the Canadian Society for Exercise Physiology (2002a, 2002b)	2002	6-14 years	Increase time currently engaged in physical activity by at least 30 min per day (in periods of at least 5 to 10 min), progressing to ≥90 min per day more physical activity; the 90 min per day increase in physical activity should include both moderate (60 min) and vigorous (30 min) activities; decrease time spent doing sedentary activities (television, video games, internet), initially by 30 min per day, eventually by ≥90 min per day
Weight Realities Division of the Society for Nutrition Education	United States	Guidelines for childhood obesity prevention programs	Weight Realities Division of the Society for Nutrition Education (2003)	2003	Children	Be active for at least 60 min per day; limit screen time to <2 hr per day and replace with more active activities

US National Association for Sports and Physical Education	United States	Guidelines for appropriate physical activity for elementary school children	Corbin and Pangrazi (2004)	2003	5-12 years	Accumulate at least 60 min, and up to several hours, of age-appropriate physical activity on all, or most days of the week; daily accumulation should include moderate and vigorous physical activities, with the majority being intermittent in nature
*UK Department of Health	United Kingdom	At least five a week: evidence on the impact of physical activity and its relationship to health	Department of Health, Physical Activity, Health Improvement and Prevention (2004)	2004	All	Children and youth should accumulate ≥60 min of at least moderate intensity physical activity each day. At least twice a week this should include activities to improve bone health, muscle strength and flexibility. This can be achieved continuously or alternatively cumulatively in bouts no shorter than 10 min in length
US Department of Agriculture	United States	Dietary guidelines for Americans	US Department of Health and Human Services and US Department of Agriculture (2005)	2005	Children and youth	Accumulate ≥60 min of physical activity on most, preferably all, days of the week
Divisions of nutrition and physical activity and adolescent and school health of the US Centres for Disease Control	United States	Evidence-based physical activity for school-age youth	Strong et al. (2005)	2005	6-18 years	Participate in ≥60 min per day of moderate to vigorous physical activity; activities should be developmentally appropriate, enjoyable and involve a variety of activities
*British Association of Sport and Exercise Sciences	United Kingdom	The ABC of physical activity for health: A consensus statement from the British Association of Sport and Exercise Sciences	O’Donavon et al. (2010)	2010	All	Children and young people aged 5-16years should accumulate at least 60 min of moderate to vigorous intensity aerobic activity per day, including vigorous-intensity aerobic activities that improve bone density and muscle strength

*Not applicable	Canada	Systematic review of the health benefits of physical activity and fitness in school-aged children and youth	Janssen & LeBlanc (2010)	2010	Children and youth (5-17 years)	Children and youth 5-17 years should accumulate on average at least 60 min of MPA per day, and up to several hours of at least MPA. Some health benefits can be accrued through 30 min of MPA per day. More VPA should be added when possible, including activities that strengthen the muscle and bones (on at least 3 days per week)
*UK Department of Health	United Kingdom	Start active, stay active: a report on physical activity from the four home countries' Chief Medical Officers	Department of Health, Physical Activity, Health Improvement and Protection	2011	All (5-65+ years)	Children and youth 5-18years should engage in moderate-to-vigorous physical activity for ≥60 min, and up to several hours each day. Vigorous activities which strengthen muscle and bone should be performed 3 days per week. This can be achieved continuously or alternatively cumulatively in bouts no shorter than 10 min in length

*Note these guidelines were not included in the original Janssen (2007) review.

2.6 Secular trends in physical activity

It is often assumed by the public and researchers alike that children's PA levels have been declining due to contemporary society's reliance on motorised transport, labour-saving devices and increases in forms of sedentary entertainment (Rowland, 2002). However, in absence of sufficient baseline data it is not possible to conclusively pass judgement on trends in PA over time (Dollman, Norton, & Norton, 2005). A systematic review of primary data found that of nine eligible studies, six showed PA had declined from the mid-1970s to the early 2000s in youth (Knuth & Hallal, 2009), however these studies used self-report questionnaires and were mainly focused upon constructs such as leisure time physical activity (LTPA), or physical education provision.

Studies of transport patterns (i.e. mode of school transport) may offer some insight into children's PA habits over time. Ham, Martin, and Kohl (2008) analysed data on active transport in US children and adolescents (aged 5-18 years) from transport surveys in 1969 and 2001. The percentage of children who actively commuted i.e. walked/biked to school declined from 42% in 1969 to 16% in 2001, with the number of children travelling by car increasing from 16% to 46%. Much of this decline was attributed to increased travel distances from school (Ham et al., 2008). Similarly trends in children's discretionary time are of interest. Sturm (2005) pooled primary data and reported that the discretionary time of children declined by ~12% between 1981 and 1997 (3-12 years), but sports participation increased by 92 min in 9-12 year olds (Sturm, 2005). Arguably the most interesting data come from work with Old Order Amish (OOA) communities (who reject a modern lifestyle and labour-saving devices) which has revealed that OOA children are less obese and engage in ~50 min of weekday MVPA more than their contemporary counterparts (Esliger et al., 2010).

With the availability of some empirical data, the concept that children's PA levels have declined (and conversely sedentary time increased) due to society's drive to engineer PA out of our everyday lives has some credence. However, due to dissimilar findings between studies, reliance upon proxy/self-report data and the only recent deployment of objective activity measures in nationally representative samples, temporal trends of children's PA may remain equivocal for some time.

2.7 Measuring physical activity in children

Physical activity measurement in children is believed to be more complex than in older populations as their movement patterns are frenetic and intermittent, with bursts of MVPA interspersed with longer bouts of lower intensity PA (Baquet, Stratton, Van Praagh, & Berthoin, 2007). Further, children move less efficiently, requiring greater energetic output to conduct a given task (Harrell et al., 2005) and they are believed to have shorter attention spans and lower cognitive development (Welk et al., 2000).

The degree of reliability and validity logically dictate the accuracy of a given PA measurement tool. Consideration of both issues is important if researchers are to determine compliance to PA recommendations, characterise PA and health relations and test the effectiveness of interventions (Freedson, Pober, & Janz, 2005). To add further difficulty it has been suggested that measurement tools should be utilisable in large populations and be valid across ages, cultures and ethnicities (Hands, Parker, & Larkin, 2006). Not surprisingly there is at present no unilaterally agreed 'gold standard' measure of free-living PA. It is contested by Esliger and Tremblay (2007) that measurement tool selection should be based on the suitability of the method for a given study goal and population, which often results in a trade-off between feasibility, participant acceptability and data accuracy (Esliger & Tremblay, 2007).

As noted earlier PA is not analogous to EE. The global construct of 'movement' can be sub-divided into both PA (behaviour) and EE (energetic cost of behaviour). Both subjective and objective tools exist for the measurement of PA and EE (Sirard & Pate, 2001). Researchers usually convert raw data values (e.g. activity counts) to units of EE or time in a given intensity threshold (Lamonte & Ainsworth, 2001). The conceptual diagram illustrated in Figure 2.2 outlines the predominant field and laboratory methods used to measure PA and EE.

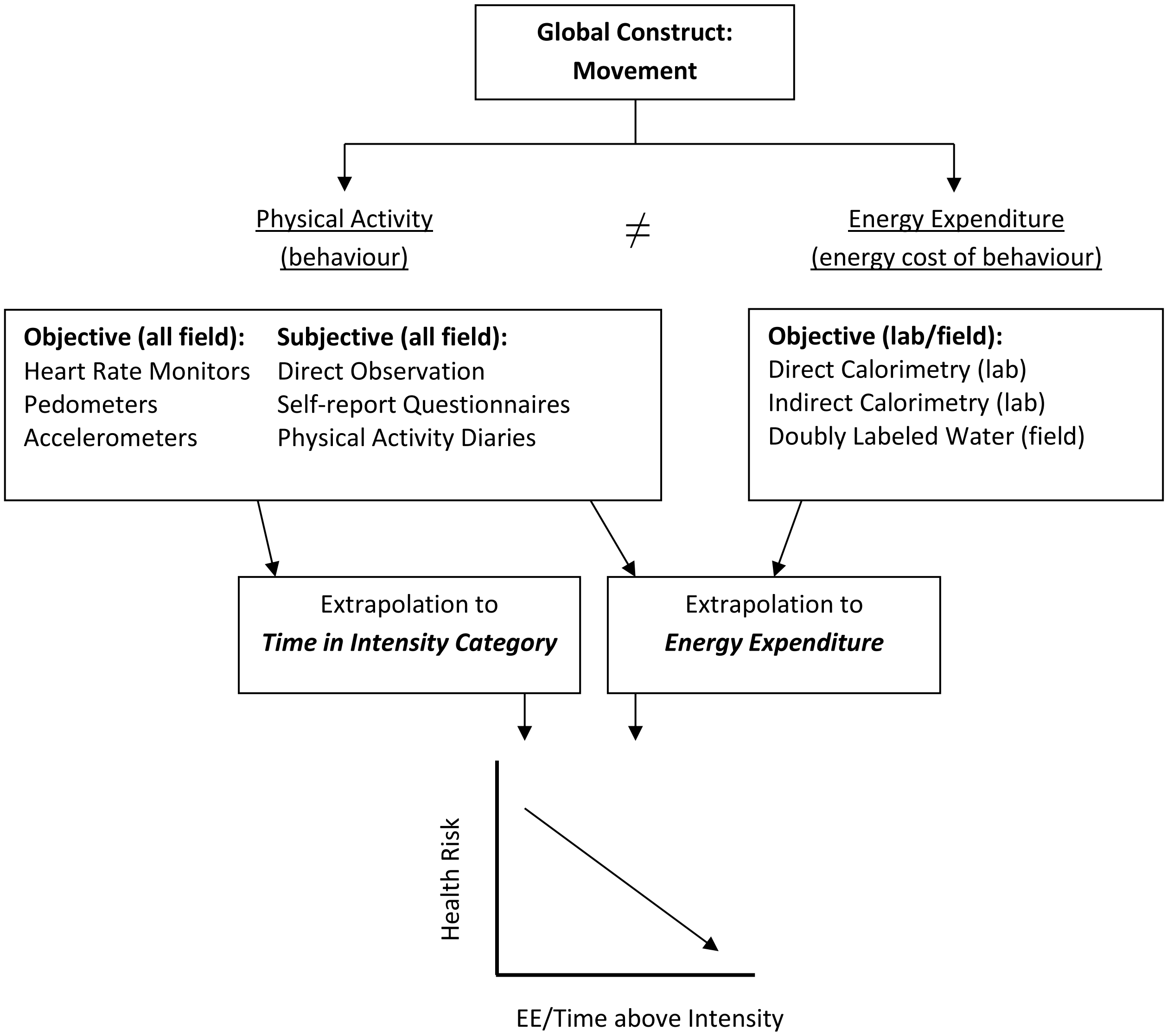


Figure 2.2 The inter-relationship between movement, physical activity and energy expenditure and common measurement methods. Adapted from (Lamonte & Ainsworth, 2001).

Physical activity epidemiology has traditionally relied upon self-report tools, as these are simple to implement, are cost effective and present low participant burden - making them suitable for large-scale studies (Mattocks, Tilling, Ness, & Riddoch, 2008b). Nevertheless it is well reported that children may not be sufficiently cognitively developed to self-report their PA habits accurately; indeed compelling data show children and youth vastly overestimate via recall methods (Adamo, Prince, Tricco, Connor-Gorber, & Tremblay, 2009).

Objective measurement tools are however not as susceptible to researcher bias (e.g. intra-observer reliability) or participant characteristics (e.g. over-reporting of PA by obese individuals). As such they provide a more stable and accurate estimate. Objective monitors can capture and record (within on-board memory) PA data both in real time (i.e. accelerometers and heart rate monitors) and on a daily basis (i.e. pedometers), allowing for detailed examination of temporal patterns. As such the data objective devices generate are more information rich than traditional self-report methods (Corder et al., 2008; Esliger, Copeland, Barnes, & Tremblay, 2005).

Due to their relative expense, participant burden (i.e. extended monitoring protocols) and researcher burden (i.e. device initialisation and data mining) objective PA monitors were initially only suitable for smaller sample investigations. However with technological advancements objective PA monitors have become progressively smaller, less expensive, more sophisticated and increasingly user friendly for the PA researcher. This has narrowed the gap between accuracy, feasibility and usability and prompted a shift in objective tools moving out of the laboratory and into the field (Esliger & Tremblay, 2007). Indeed the suitability of objective measurement on a large scale has been proven by the use of accelerometers in population level surveillance studies (Troiano et al., 2008). Objective tools are not without limitations however as differences in device specifications, data capture methods and data reduction protocols make comparison between studies problematic. The section that follows will provide a brief description of field-based measures of PA, with the section on accelerometry the most extensive as it is essential to the research in this thesis.

2.8 Subjective tools

2.8.1 Self-report

For the past 40 years epidemiologists have relied upon self-report questionnaires, diaries/logs and interviews to gather data on habitual PA on a population level (Shephard, 2003). A critical review by LaPorte, Montoye, and Caspersen (1985) clearly summarised the four main components that characterise self-report approaches. The first is the time frame participants are asked to recall, it may range from 5 min to 3 decades; the second is the nature and detail of one's PA habits (i.e. frequency, intensity, duration, type); the third is the mode of collection (i.e. interview, questionnaire, dairy); the final component is the use of a summary index based on an estimate of total PA volume (i.e. total energy expenditure in kcal) or an ordinal scale of PA level. Self-report methods can also provide information on PA pattern and are a popular approach as they cost very little, are practical, and place modest burden on participants (Adamo et al., 2009).

A number of reviews have critically evaluated the literature pertaining to the reliability and validity of self-report measures. Corder et al. (2008) reported validity coefficients of r = 0.2 to r = 0.72 for self-report diaries (against accelerometry and doubly labeled water), r = 0.33 to r = 0.51 for self-report questionnaires (against accelerometry) and r = 0.09 to r = 0.73 for interview-administered self-report tools (against accelerometry and doubly labeled water) (Corder et al., 2008). In terms of absolute validity Adamo et al. (2009) found that indirect measures of PA (activity logs or diaries, questionnaires, surveys and recall interviews) overestimated objectively measured PA (heart rate monitoring, doubly labeled water, accelerometry, pedometry and direct observation) by an average of 114% (range = -57% to 2695%) in boys and 584% (range= -95% to 13,025%) in girls. Sallis and Saelens (2000) reviewed 17 children's self-report instruments and found that reliability coefficients for self-report questionnaires, self-report diaries/logs and interview-administered questionnaires ranged from r = 0.60 to 0.98 (Sallis & Saelens, 2000).

The moderate validity of self-report methods reported in the aforementioned studies is largely due to the limitations of the human memory. The issue of recall is arguably the largest source of bias as children (who are less cognitively developed than adults) may have an exaggerated perception of time and effort related to their PA habits. The over-reporting of PA observed by Adamo et al. (2009) may also be contributed to by the sporadic nature of children's PA (Baquet et al., 2007), resulting in children reporting multiple short bouts as a long continuous bout. LeBlanc and Janssen (2010) offer a good example, whereby a child may report a 60-min physical education lesson as a full 60 min of MVPA when it actually consists of instruction, demonstration, getting changed and time on the sidelines watching others participate. An additional issue that may hinder self-report methods is the influence of social desirability, whereby respondents give answers that they perceive to be socially acceptable, as opposed to the true volume of PA (Adamo et al., 2009).

Many questionnaires use EE compendia to convert self-report data into estimates of EE. Despite the existence of a MET compendium in children (Ridley et al., 2008), only 35% of the compendia were child-derived MET values, the rest being adult-derived estimates, the use of which may underestimate EE due to the higher energy cost of movement in children compared to adults, although utilising child-specific resting metabolic rate (1 MET) may reduce error (Harrell et al., 2005; Ridley et al., 2008). Corder et al. (2009) assessed the validity of four self-report questionnaires to predict PAEE in children and youth (4-17 years) using doubly labeled water as the criterion. Results showed that some questionnaires could accurately assess group level EE, but were quite erroneous on an individual level (95% CI: -51 to 11.6 kj.kg.day^{-1}), and could rank individuals relatively well. It appears therefore that questionnaires do not provide an accurate measure of activity volume when expressed as EE. Despite the inaccuracy of self-report methods for the estimation of EE (Corder et al., 2009), intensity and duration (Adamo et al., 2009) they can still provide information on the relative amounts of PA (ranking) between children (Corder et al., 2009) and afford contextual information on the setting and mode of PA behaviours (Trost, 2007).

2.8.2 Direct observation

Observational techniques have been developed by behavioural scientists to monitor PA. These methods require a trained individual to observe a child (or group of children) and classify PA into intensity categories over a specified time period, usually in a natural setting such as in school or at home. Typically observation is conducted and coded with the naked eye, or alternatively a video recording is taken for future coding in the laboratory. The direct observation method is considered to be a criterion method by some researchers and has been used to calibrate accelerometers in young children where indirect calorimetry may not be feasible (Sirard, Trost, Pfeiffer, Dowda, & Pate, 2005). The validity and reliability of direct observation as a measure of PA has been found to be acceptable. McKenzie (2002) reviewed nine of the then currently available measurement systems and reported that eight of the nine systems had considerable evidence of concurrent validity when compared to heart rate monitoring, accelerometry and EE (as measured by indirect calorimetry). In addition, the inter-observer agreement (reliability) was good with percent agreement ranging from 84 to 99% (McKenzie, 2002).

Several observation systems have been developed to measure children's PA behaviour. One example is the children activity rating scale (CARS) where the observer rates the child's activity intensity level from sedentary through to vigorous on a scale of 1 to 5 (1 = static - no movement; 2 = static - with movement; 3 = translocation - slow/easy; 4 = translocation - moderate; and 5 = translocation - fast) (DuRant et al., 1993). The direct-observation approach is best suited for data collection on small numbers of children, as it is particularly time consuming for the researcher and requires time-intensive training and data coding (Loprinzi & Cardinal, 2011). Further, data are usually collected in 1-min epochs and therefore may underestimate time spent in high-intensity activities due to their sporadic nature and short duration (van Cauwenberghe, Labarque, Trost, de Bourdeaudhuij, & Cardon, 2011). Whilst considered an objective criterion method of PA by some (Sirard & Pate, 2001) direct observation can only determine time spent in a given activity intensity (if validated against a measure of physiologic work), and is highly dependent upon the researcher's ability to observe and record behaviour in a consistent manner. In addition the presence of the researcher may disrupt regular activity patterns of the participant(s), much like reactivity (i.e. hawthorne effect) to a wearable motion sensor (Trost, 2007).

However the advantages of this approach are that observational procedures are flexible and they permit the researcher the ability to measure additional factors related to PA such as the context (environment) in which they take place, and the presence of significant others and the types of activities most frequently participated in (Trost, 2007).

2.9 Objective tools

2.9.1 Doubly labeled water

Doubly labeled water (DLW) is normally used as a criterion measure of PA in children. The doubly labeled water technique provides an integrated measure of EE over a given time unit (LaPorte et al., 1985). The DLW method was first used in humans by Schoeller in the early 1980s (Schoeller & van Santen, 1982), and involves the ingestion of 'heavy water' ($^{2}H^{1}H^{18}O$) containing two stable isotopes: deuterium-labeled water ($^{2}H^{2}O$) and oxygen-18-labeled water ($H^{2}\ ^{18}O$). Following ingestion, deuterium-labeled water is excreted from the body through the loss of water, whilst oxygen-18-labeled water is excreted as carbon dioxide and water loss, and thus the difference between the elimination rates of these isotopes can provide a measure of carbon dioxide flux (Goran, Peters, Herndon, & Wolfe, 1990). Usually urine or saliva samples are collected at regular intervals in the monitoring period. From this, carbon dioxide production and therefore EE can be estimated using standard indirect calorimetric calculations (Schoeller, 1983; Schoeller & van Santen, 1982).

In children and adults doubly labeled water has been validated against whole room calorimetry with total EE estimates ranging in absolute accuracy by 5-12% (Goran, 1994), therefore due to its high degree of accuracy for predicting EE, DLW is often referred to as a criterion measure (Loprinzi & Cardinal, 2011). The advantages of the doubly labeled water method are that it does not restrict free living, requires minimal participant compliance and is generally acceptable to the participant (LaPorte et al., 1985). However the DLW method is limited as it only assesses a single construct of PA (total volume), not providing any information on the pattern of PA which may be important variables when assessing the relationship between PA and health. Further, the main disadvantage is the expense and availability of the stable isotopes, which can cost around £200 per child (Goran, 1994; LaPorte et al., 1985). Therefore the doubly labeled water method is more suited to small-scale investigations than measuring free-living PA in large epidemiologic studies.

2.9.2 Heart rate monitoring

Prior to the advent of wearable motion sensors heart rate monitoring was the only feasible objective method of measuring PA *en masse* outside of the laboratory. Heart rate monitors provide an indirect objective indicator of the physiologic effect of PA (Welk et al., 2000). Continuous heart rate monitoring works on the principle that there is a linear relationship between heart rate (HR) and oxygen consumption (VO_2) during dynamic activities, from which one can estimate PAEE (Strath et al., 2000). Modern heart rate monitors have the capability to sample and store data in an ambulatory fashion for extended periods of time, with data being stored in on-board memory for later download to a personal computer. This time-stamping ability permits the end-user to quantify the intensity, duration and frequency of PA. When coupled with their ease of use and relative inexpense (compared to accelerometers) they are an attractive option for PA researchers.

However, there are a number of well-documented limitations to using heart rate to quantify PA and associated EE. Firstly there is considerable inter-individual variation in heart rate at a given work load due to differences in age, fitness and gender. To overcome the limitations of fitness, age and gender researchers have used net HR (activity HR – resting HR), or percent heart rate reserve (activity HR – resting HR/maximum HR – resting HR x 100) which correlates well with percent VO_2 reserve (Strath et al., 2000) and from which child-specific MVPA cut-points have been derived (Stratton, 1996). A second limitation is that the HR-VO_2 relationship is weaker and non-linear at low and very high intensities, potentially leading to inaccurate predictions of both PA intensity and EE (Livingstone et al., 1992). One approach is to only employ heart rate monitors to measure PA of at least moderate intensity, with heart rate above a certain level (Corder et al., 2008).

Indeed there is the HR-flex method, whereby one identifies the 'flex point' above which the HR-VO_2 relationship becomes linear (Bassett, 2000). Below this point the HR-VO_2 relationship is quite variable due to the contribution of factors other than body movement such as environmental stress (e.g. ambient temperature), psychological state and hydration status (Corder et al., 2008). The flex HR has been defined in various ways, but is often termed as the average of the lowest heart rate during exercise and the highest heart rate during rest (Corder et al., 2008). As the flex HR involves individual calibration it does account for age, gender and fitness variance, but is limited by its time-consuming nature and is therefore not feasible for large-scale epidemiologic studies (Loprinzi & Cardinal, 2011).

There are further limitations of heart rate monitoring that are more specific to child populations. Heart rate suffers from a temporal lag in response to the initiation of a bodily movement and therefore may not be suited to capturing intermittent bursts of PA which are of course characteristic of children's PA pattern (Rowlands & Eston, 2007). To summarise, there are numerous disadvantages to heart rate monitoring namely inter-individual variation in heart rate response and the contribution of non-movement factors that influence heart rate. Despite limitations, heart rate monitors are a fairly inexpensive and unobtrusive method of objectively quantifying PA and EE.

2.9.3 Wearable motions sensors

The field of PA measurement has benefited from the rapid development of wearable motion sensors that can measure body movement. Both pedometers and accelerometers measure displacement of body segments and tally these movements (provided they are of sufficient magnitude) into PA units per time period (i.e. steps, counts etc.). This has allowed the examination of habitual PA over long periods due to their unobtrusive size, long battery life and on-board memory.

2.9.3.1 Pedometers

It is believed that Leonardo da Vinci designed the pedometer almost 500 years ago (Rowlands & Eston, 2007) but it has taken until the explosion of interest in the health benefits of PA in the latter part of the 20th century for there to be interest in the use of pedometers to measure human movement. A pedometer is a small step-counting device that is usually affixed to the waistline of the wearer. Pedometers are designed to sense the vertical accelerations of the trunk during ambulation using a horizontal spring-suspended pendulum arm that moves up and down with displacement of the trunk (Bassett, 2000). A ‘step count’ is recorded when a sensed vertical acceleration is greater than a pre-set force sensitivity threshold, with movement of the lever arm completing an electronic circuit (Tudor-Locke & Lutes, 2009). The number of steps taken by the end-user are recorded by the pedometer and displayed as a running tally on a digital screen. Therefore pedometers provide a volumetric index of total PA over the time period assessed (i.e. steps per day). Newer piezoelectric pedometers (e.g. Kenz Lifecorder), which comprise a piezoelectric accelerometer measure sinusoidal wave forms and tally the number of these waves to determine the number of steps. These are still considered pedometers however as the end metric is usually steps (McClain & Tudor-Locke, 2009).

The force-sensitivity threshold of a pedometer is an important component as it is designed to filter out movements that are not due to ambulation such as car vibrations

(Tudor-Locke & Lutes, 2009). The choice of threshold is significant as the correct trade-off between sensitivity and specificity is required for the production of an accurate pedometer, that is the threshold must be sensitive enough to detect a step (i.e. maximise true-positives, sensitivity) but not too insensitive as to include non-steps (i.e. minimise false-positives, specificity) (Tudor-Locke & Lutes, 2009). Due to the emphasis on filtering out non-steps, many pedometers have a high sensitivity threshold so that accuracy of some step counters is reduced at slower walking speeds (~2 mph) (Crouter, Schneider, Karabulut, & Bassett, 2003). The force-sensitivity threshold differs between pedometer models, affecting accuracy and therefore making comparisons between populations using different models difficult. Greatest accuracy has been found in the Yamax Digiwalker range (e.g. SW-701, SW-200) which has been shown to have <3% error compared to actual observed steps in adults at a normal walking pace and a force-sensitivity threshold of 0.35g, which is recommended for optimal accuracy (Crouter et al., 2003; Tudor-Locke et al., 2006).

In terms of convergent validity in children (8-10 years) pedometers have been compared against a number of measures such as oxygen uptake (r = 0.80) and heart rate (r = 0.62) (Eston, Rowlands, & Ingledew, 1998). Kilanowski, Consalvi, and Epstein (1999) compared Yamax DW SW-200 to the Tritrac vector magnitude (total counts) in children (7-12 years) during sedentary and moderate intensity activities, observing a correlation coefficient of r = 0.98 for moderate activities and r = 0.50 for sedentary activities (Kilanowski et al., 1999). The strong correlation between pedometers and other indirect measures of physical activity show that pedometers provide a valid measure of PA in children. In relation to absolute validity Nakae, Oshima, and Ishii (2008) examined the accuracy of both spring-levered (Yamax EC-200) and piezoelectric (Kenz Lifecorder & Omron HJ-700IT) pedometers in a large sample of children (n = 394; 7-12 years) over three self-paced walking trials against researcher-observed steps. The spring-levered pedometers underestimated steps across all three trials (by up to 25%) whereas the piezoelectric pedometers were within ±3%. This is in contrast to the results of Smith and Schroeder (2010) who found the spring-levered Yamax SW-701 to exhibit <3% error during self-paced walking in children aged 10-11 years. Similarly

Beets, Patton, and Edwards (2005) found the Yamax DW-200 to display <5% error at speeds >67m.min^{-1} in a treadmill experiment in children (5-11 years). The differences in degree of accuracy between studies reflect differences in pedometer design, which has been shown in a multi-brand comparison of 10 pedometers in adults (Crouter et al., 2003). A recent systematic review of the criterion validity of pedometers in youth concluded that pedometers are deemed valid indicators of PA at moderate and fast speeds (as determined from treadmill studies), which may not be of detriment as the potential health benefit of light intensity PA is unknown in children (McNamara, Hudson, & Taylor, 2010).

Studies have assessed the inter-unit reliability using contralateral hip placements in children and during 7 days of free living, reporting high intra-class correlation values (ICC = 0.96-0.98) (Barfield, Rowe, & Michael, 2004). During treadmill walking at speeds >67 m.min^{-1} Beets et al. (2005) reported high agreement between contralateral pedometers (ICC = >0.9) in children aged 5-11 years. ICC values were however low when tested at speeds of 54 m.min^{-1}. Similarly Jago et al. (2006) also found no systematic bias between contralateral hip placements (p = 0.56) during over-ground walking and running trials in children aged 10-15 years, with ICC values ranging from 0.73 to 0.80. In the same study intra-instrument reliability was good with no systematic bias between repeat trials and ICC values ranging from 0.51 to 0.92 (Jago et al., 2006). Findings from studies of reliability in adults during self-paced walking corroborate these child-derived observations (Schneider, Crouter, Lukajic, & Bassett, 2003). Whilst it appears that pedometers offer a reproducible measure of PA in children for normal walking speeds, these findings were based on *in-vivo* protocols that may introduce inherent error issues such as side dominance. Therefore there is a clear need for future studies to utilise mechanical shakers, and a priority must be to test within pedometer reliability over repeat trials.

As with any PA measurement method there is the issue of reactivity (i.e. Hawthorne effect), but this may be heightened with wearable motion sensors as participants are constantly aware of their behaviour being monitored. Vincent and Pangrazi (2002) found there to be no reactivity in children aged 6-12 years when wearing sealed pedometers over an 8-day period. Similarly in the Canadian Physical Activity Levels among Youth study (Craig, Tudor-Locke, Cragg, & Cameron, 2010) no differences in daily step counts were apparent between days, over a 7-day monitoring period in 11,669 5-19 year olds. Interestingly the same holds for unsealed pedometers, as Ozdoba, Corbin, and Le Masurier (2004) found no differences in step counts between sealed and unsealed pedometers over an 8-day free-living period in children aged 8-9 years. These data combined suggest that in children at least, wearing an open or closed pedometer does not induce a change in usual PA.

A limitation of most pedometer models is the lack of internal clock so that data cannot be time stamped, and therefore the frequency, intensity and duration of PA cannot be assessed (Bassett, 2000). Indeed pedometers do not distinguish the magnitude of the acceleration sensed, so that a step is recorded whether the individual is walking or running (Trost, 2007). In addition they cannot discriminate between gradients, and loaded or unloaded walking. Some pedometer users however have determined cadence rates per time unit to be able to determine PA intensity time, for example Jago et al. (2006) found that 8000 steps in 60 min is equivalent to 60 min of MVPA (Jago et al., 2006). Additional metrics which pedometers provide include the distance walked by the end-user (distance = steps x stride length), which provides a useful measure of walking behaviour. However as stature, and therefore stride length varies in children, distance may vary between individuals at different maturational stages (Corder et al., 2008).

As outlined above, pedometers may have limitations but they do provide a valid and reliable objective measure of PA volume and can be used as a measure of PA intensity time over short periods. Their primary benefits are that they are relatively inexpensive, easy to use, require little data processing and thus are ideally suited for large-scale epidemiologic studies.

2.9.3.2 Accelerometers

Arguably the most significant advance made over the last two decades in the field of PA measurement is the development of ‘accelerometer-based’ activity monitors. As the vast majority of PA is dynamic, the measurement of body segment motion (i.e. trunk and limbs) is a logical method for indirectly capturing PA behaviour. Accelerometers are small electronic PA monitors that measure the acceleration and deceleration (Δ in velocity) of body segment movement along a reference axis (Chen & Bassett, 2005; Esliger et al., 2005). Usually built into a small plastic case, an accelerometer can be affixed to the hip, wrist or ankle dependent upon the device specifications and the practicality of the placement. The device itself consists of either a piezoelectric or piezoresistive element, which when subject to acceleration or deceleration deforms and emits a voltage signal directly proportional to the intensity of that movement (Chen & Bassett, 2005).

Accelerometers work on the principle that body acceleration is proportional to the mechanical force and thus energetic cost required to displace the body, therefore the intensity of human movement can be quantified (Chen & Bassett, 2005). Most accelerometers convert the raw acceleration signal into an activity count (proportionate in magnitude to the acceleration measured) which is integrated over a given time interval (e.g. 1 s, 10 s, 1 min), called an epoch. Therefore the greater the count output for a given epoch, the greater the intensity of the activity. However when the device is static it stores a count of zero, hence accelerometers can only be used to monitor dynamic (as opposed to static i.e. postures, isometric contractions) PA.

Accelerometer output is not solely the function of sensed acceleration; accelerometers also consider the frequency of the movement (Esliger & Tremblay, 2006). Human accelerations typically range in frequency from 0 to 25 Hz (Chen & Bassett, 2005) and therefore high and low band pass frequency filters are inbuilt to filter out or dampen accelerations that are not likely to have been generated by human movement. A more detailed discussion of the mechanical and technical properties of accelerometers can be found in the review of Chen and Bassett (2005).

The PA counts generated can be used in their raw form as a measure of PA volume (i.e. total counts) or rate (i.e. counts/min) (McClain & Tudor-Locke, 2009). However counts are essentially an arbitrary unit and therefore biological meaning must be obtained by the conversion of counts into physiologically meaningful units, such as EE or biomechanical loading. The majority of accelerometer end-users utilise intensity cut-points to assign the number of epochs (time) spent in a specific intensity bracket, which is anchored to a given level of EE (McClain & Tudor-Locke, 2009). For example if data were collected using 10-s epochs for 24 hr, and 40 epochs registered as above the MVPA count threshold (i.e. 3/4 METs) then an individual's accumulated time spent in MVPA would be 400 s (10 s x 40 = 400 s) or 6.6 min.

Accelerometers can therefore provide detailed information upon the volume, rate, frequency, intensity and duration of PA and sedentary behaviour, whilst remaining a small and relatively unobtrusive device (Esliger et al., 2005). There are numerous research grade accelerometer models currently in existence, all of which differ by size, weight and type. Some accelerometers can measure movement in single planes (uniaxial), two planes (bidirectional) and three planes (triaxial) or are sensitive to all planes of movement but more sensitive in one (omnidirectional). For a detailed review of the most frequently used commercial devices see Godfrey, Conway, Meagher, and OLaighin (2008). The following section will review (a) the validity and reliability of accelerometers in general, (b) the Actiwatch 4 device (which will be used in this thesis) and (c) issues related to accelerometer data collection and processing.

2.9.3.3 Accelerometer validity

The validity of a new accelerometer device (for research use) is typically judged by correlating proprietary activity counts against a criterion reference such as EE (e.g. VO_2), speed produced by treadmill locomotion, or directly observed structured physical activities (Freedson et al., 2005). Usually participants perform a range of locomotor and lifestyle-based activities (calibration source activities) that differ in type and intensity whilst wearing an accelerometer, with the criterion reference being measured concurrently. Based on the relationship between activity counts and EE (for example), validation can be taken a step further with the prediction of EE, known as value calibration. Value calibration studies derive estimates of EE from activity counts using regression modelling. Most accelerometer users solve derived equations for child-specific MET values to establish intensity count cut-points corresponding to intensity categories (Freedson et al., 2005). This approach has been somewhat superseded by the use of decision boundaries and receiver operating characteristic curve (ROC) analysis (Jago, Zakeri, Baranowski, & Watson, 2007) to generate intensity cut-points with minimal misclassification between intensity categories.

Typically value calibration studies use locomotor-based protocols, with the addition of lifestyle-based activities such as walking, sweeping or stepping (Esliger et al., 2011). An inherent problem with calibration is the unique relation between movement and EE between differing activities and the impact this has on linear regression prediction equations (altered slope/intercept) and cut-off points (Welk, 2005). It has been shown that equations derived on primarily locomotor activities tend to underestimate the energy expenditure of lifestyle activities (e.g. sweeping, washing windows), and that equations based on lifestyle source activities can overestimate the energy expenditure of locomotor activities (Crouter, Churilla, & Bassett, 2006a). This results from the lower count output at a given absolute work rate for lifestyle activities, compared to locomotor activities. Recent evidence in children however suggests that intensity count cut-points may not differ greatly when developed on locomotion only, as opposed to a combination of locomotion and intermittent play activities (Jimmy, Dossegger, Seiler, & Mader, 2012). Clearly no one single regression equation, or intensity, is able to accurately estimate the EE and therefore intensity categorisation of all activities (Bassett, 2000), however alternative calibration approaches such as two-regression methods which distinguish between activity types have improved explained variance of EE (Crouter, Clowers, & Bassett, 2006b).

As a result of differing calibration source activities, sample demographics and data analysis techniques employed between studies there is considerable heterogeneity in intensity cut-points for a given intensity, both within and between monitors (Bornstein et al., 2011), further discussion on this matter is provided below. Despite difficulties in monitor value calibration, the derivation of intensity cut-points allows the calculation of time spent at a given intensity, which allows more complex health-related questions to be answered, and importantly permits assessment of compliance to PA guidelines.

2.9.3.4 Accelerometer reliability

Due to the great number of studies devoted to accelerometer validation/calibration, there has been relatively less focus upon the reliability of certain accelerometer models (Krasnoff et al., 2008). To ensure that accelerometers are producing reliable outputs, studies have examined both the intra- and-inter-instrument reliability. The intra-instrument reliability is the difference in output within accelerometer over a number of trials and the inter-instrument reliability is the difference in output between like accelerometer models when exposed to a common movement stimulus. Studies that have previously investigated reliability of accelerometers have been either laboratory-based mechanical reliability studies or participant mounted field-based studies (Brage, Brage, Wedderkopp, & Froberg, 2003; Esliger & Tremblay, 2006; Fairweather, Reilly, Grant, Whittaker, & Paton, 1999; McClain, Sisson, & Tudor-Locke, 2007).

The investigations focusing upon mechanical laboratory experiments have used various apparatus to provide a constant motion stimulus. Examples of these devices comprise rotating wheels (Brage et al., 2003), turntables (Metcalf et al., 2002) and hydraulic shaker plates (Powell, Jones, & Rowlands, 2003; Esliger & Tremblay, 2006). Evidence has shown that variability is greater at very low and very high frequencies of movement (Brage et al., 2003), and may be inversely related to the frequency/acceleration of the test condition (Esliger & Tremblay, 2006; Powell et al., 2003). From the results of these studies investigators have been encouraged to routinely test their device's reliability (unit calibration) prior to use, and discard those with an inter-unit variability level of >5-10% to ensure that artificial changes in physical activity do not manifest due to poor unit reliability (Krasnoff et al., 2008; Esliger & Tremblay, 2006).

2.9.3.5 Actiwatch 4 accelerometer

The AW4 is a small (37 x 29 x 10 mm) primarily wrist-worn accelerometer which weighs 16g and has a random access memory (RAM) capacity of 64 kb. It consists of a rectangular piezoelectric bimorph plate and seismic mass. It is omnidirectional but is most sensitive in the vertical axis. This technology detects the peak amplitude of movement acceleration and generates a voltage signal proportional to amplitude of acceleration. The raw digital voltage strings are converted to activity counts, with the peak count being selected from each individual second. Peak activity counts are integrated (and recorded) during a user-specified time interval (epoch), which ranges from 2 s to 15 min. The device has a sampling frequency of 32 Hz, and collects motions in the frequency of 0.5-7.0 Hz (Chen & Bassett, 2005). A more detailed technical specification is given in Chapter Three of this thesis.

2.9.3.6 Actiwatch 4 validity

The first study to examine the validity of the AW (AW16) measured free-living PA in 40 3-4-year-old children over a 6-hr period in child care using concurrently measured direct observation (Children's Activity Rating Scale) as the criterion (Finn & Specker, 2000). The AW was worn on the waist positioned against the lumbar spine and set to measure at a 1-min epoch. The correlation coefficients for the total observation period ranged from r = 0.03 to 0.92 (median = 0.74). Further investigation between the children's correlation coefficient and the mean 3-min CARS (r = 0.37, p = 0.02) or mean sensor counts (r = 0.31, p = 0.05) showed that the relations were stronger in those children who were more active. Despite observing a relatively strong median correlation (r = 0.74), this study is limited as total PA was the outcome, therefore no inference can be gained as to the AW's ability to measure the intensity or EE of PA. This study is restricted to highlighting that the AW provides a valid measure of the total volume of PA.

In contrast a more comprehensive validation study was conducted by Puyau et al. (2002). They evaluated the association between activity counts from the Actigraph and Actiwatch (AW16, MiniMitter Co., Inc., Bend, OR) and EE as measured by room respiration calorimetry in 26 boys and girls (6-16 years). The children were fitted with an Actiwatch on the right hip and on the lower right leg, with the units set to collect data using a 1-min epoch. Over a 6-hr period in the room calorimeter the children performed a variety of activities, ranging from playing videogames to jumping ropes. The activities were designed to represent sedentary, light, moderate and vigorous intensity. The correlation between EE/AEE for the 6-hr period was r = 0.78 for the hip, and r = 0.80 for the leg, with stronger coefficients observed for AW data (r = 0.78-0.80) than Actigraph data (r = 0.66-0.73). In addition intensity cut-points to delineate sedentary, light, moderate and vigorous activity were derived. However the standard error was high for the regression of AEE on activity counts, suggesting that the equation is not accurate for individual level prediction of PAEE. Withstanding this issue, this study showed that the AW is a valid device for the measurement of both the intensity level and volume of PA (Puyau et al., 2002).

Following this initial validation, Puyau et al. (2004) value calibrated the AW16 using 32 children (7-18 years) with room respiration and indirect calorimetry as the criteria. Three sedentary, two light, three moderate and six vigorous activities were performed in controlled laboratory or free-living settings (room calorimeter). The AW was worn on the right hip and set to record at 1-min epochs. AW and Actical counts were found to be correlated with one another (r = 0.93). AW and Actical counts were strongly correlated with EE (r = 0.79, 0.83), AEE (r = 0.82, 0.85), PA ratio (r = 0.85, 0.87) and heart rate (r = 0.63, 0.60). Alongside total EE and AEE prediction equations, intensity cut-points were derived. Improvements were made to the previous intensity thresholds (Puyau et al., 2002) as ROC analysis was used to ascertain the trade-off in sensitivity and specificity of the thresholds to obtain optimal cut-off points. Showing good performance in the ROC analysis the intensity thresholds for the Actiwatch indicate acceptable rates of misclassification of raw AW activity counts between intensity categories. Ideally intensity thresholds would be cross-validated in a separate sample for validity confirmation, however this was not performed. In addition the fairly wide 95% confidence intervals for the EE prediction equation again suggest that further work is needed before the Actiwatch can be used to accurately predict the PAEE of individuals (Puyau et al., 2004).

The validation studies cited above used the AW16, an earlier generation of the AW range. There are no known hardware differences (e.g. raw sampling rate, sensor type) between the AW4 and AW16 other than memory size (personal communication, Cambridge Neurotechnology Ltd), therefore the validity coefficients and intensity cut-points reported above are deemed applicable for both versions (AW4 and AW16) of the AW.

2.9.3.7 Actiwatch 4 reliability

To date there has only been a single published study that has included a brief test of the AW's reliability. Alongside the comparison of the AW to direct observation Finn and Specker (2000) investigated intra- and inter-accelerometer variation. Six AWs were attached to the lower back of an adult whilst walking and running at various speeds (3.22, 6.44, 9.65 and 12.78 $km.h^{-1}$) for 4 min. The coefficient of variation (CV) was calculated as a means of assessing variability; the intra-accelerometer CV ranged from 1.8 to 5.6%. The CV between accelerometers at the speeds of 3.22, 6.44, 9.65 and 12.78 $km.h^{-1}$ was 17%, 11%, 8% and 10% respectively (Finn & Specker, 2000).

The values observed by Finn and Specker (2000) are comparable to other mechanical and participant-mounted reliability trials, with within unit CV values of <10%, and between units values having been reported to range between 1-18% (Esliger & Tremblay, 2006; Fairweather et al., 1999; McClain et al., 2007; Powell & Rowlands, 2004). However the CV reported by Finn and Specker (2000) may not be a true estimate of unit variance as this was a single subject-mounted trial and as such included variance due to accelerometer positioning, gait mechanics etc. Mechanical set-ups allow precise control over the experimental conditions, and permit identification of variability solely attributed to the accelerometer without the influence of accelerometer position or individual movement patterns (Esliger & Tremblay, 2006). There is therefore limited knowledge on the reliability of the Actiwatch 4 under mechanical conditions and a distinct requirement for this to be characterised before use in the field.

2.9.3.8 Accelerometer measurement issues

2.9.3.9 Placement site

Accelerometer placement site is of concern when measuring PA in children as the acceleration signal may vary between sites due to differential body segment movement patterns (Westerterp, 1999) and orientation of the unit's sensor (Welk, 2005). Typically accelerometers are worn over or around the hip, as value calibration studies have used this placement site on the basis that the trunk is the largest muscle mass active during movement and therefore output should logically be most closely related to EE. Despite this well-established practice of hip placement, interest in the interchangeability of sites has led to investigation of differences between contralateral hip placements and between the hip and lower back.

Several studies have examined differences in output between the lower back and hip in children (Nilsson et al., 2002) and adults (Yngve et al., 2003), finding hip placement leads to greater moderate activity in the field (Nilsson et al., 2002), and a greater count output during moderate laboratory activities (Yngve et al., 2003). However these differences are not seen under free living in adults (Yngve et al., 2003). Some studies have investigated contralateral hip placements by comparing accelerometer output between the left and right hip during free living (Cook & Lambert, 2009; Fairweather et al., 1999; McClain et al., 2007). Fairweather et al. (1999) found significant differences of ~5% of the daily mean in pre-school children who wore the CSA activity monitor (first-generation Actigraph) over two free-living days. However, Cook and Lambert (2009) did not observe any differences in count output between the left and right hip in adult males. Similarly McClain et al. (2007) found no differences in count output between hip placements, but observed small differences in MVPA.

On the basis of the studies presented above, it appears variance in accelerometer output due to placement site is unclear; in addition there are no data on differences between contralateral wrist placements. Further, these studies did not consider body side dominance. By comparing the left versus right hip these studies have negated the potential influence of side dominance. If data were re-categorised as the dominant versus non-dominant body side, it is possible that differences in accelerometer output may be present.

The AW validation studies placed the accelerometer either on the hip or the lower right leg (Puyau et al., 2002; Puyau et al., 2004). The Actiwatch however, was designed to be worn at the wrist and as an issue of practicality is worn more easily at the wrist. Indeed it has been shown in adults that wearing an accelerometer at the wrist is acceptable and may promote high compliance to extended data collection protocols (van Hees et al., 2011). The placement of the AW4 at the wrist raises a validity issue as PA intensity cut-points developed for the AW from a hip placement would not be valid for use with wrist-derived data due to the unique relationship between movement pattern and EE (Crouter et al., 2006a). Determination of differences in output between hip and wrist placement sites is important as this would provide (a) descriptive data upon accelerometer output at hip and wrist placement sites and (b) an opportunity to derive regression-predicted hip-derived PA intensity time from wrist-derived data, precluding full-value calibration at the wrist.

2.9.3.10 Epoch

The consideration of accelerometer sampling rate (epoch) is important as children's PA patterns are intermittent with few bouts lasting >80 s, and the majority of MVPA <2 s (Baquet et al., 2007). Logically a conventional 1-min epoch may underestimate higher intensity PA due to time smoothing - which occurs when bouts of low intensity and high intensity are integrated within the same epoch. Rowlands, Powell, Humphries, and Eston (2006) showed that relative to a 1-s epoch, using a 60-s epoch overestimated moderate and vigorous activity, but underestimated time in very hard activity. Similarly Nilsson et al. (2002) reported significant differences in vigorous and very vigorous activity between 5-, 10-, 20-, 40- and 60-s epochs, suggesting a more detailed picture of the PA pattern is gained by using shorter epochs.

Despite evidence that a short epoch is preferential, Reilly et al. (2008) contest that there is little empirical evidence to suggest that short epochs are essential, unless the outcome of interest is vigorous activity. They observed small significant differences in MVPA between 15-, 30-, 45- and 60-s epochs in 32 5-6 year olds (Reilly et al., 2008). However McClain, Abraham, Brusseau, and Tudor-Locke (2008) found that compared to direct observation, 5-s epochs minimised error between observed and accelerometer-predicted MVPA, with longer epochs introducing more error. In the largest sample (*n* = 311) epoch investigation to date Edwardson and Gorely (2010) found that in children aged 7-11 years a significant epoch effect was present for all intensities (comparing epochs of 5, 15, 30 and 60 s). A shorter epoch was associated with greater vigorous PA and decreased moderate PA. Interestingly average MVPA ranged from 122.67 to 139.92 comparing a 5- to 60-s epoch, which the authors suggest renders the choice of epoch when measuring MVPA less important (Edwardson & Gorely, 2010). However as an epoch of 15 s was shown to minimise bias compared to a 5-s sampling rate (Edwardson & Gorely, 2010) which McClain et al. (2008) showed to be the most accurate compared to a criterion, an epoch of ≤15 s seems appropriate.

Generally accelerometer end-users will set their activity monitors to record at the maximum resolution possible, with the latest generation of accelerometers having sufficient memory and battery life to record for at least 7 days using a short epoch. In light of the equivocal results regarding the epoch effect it appears wise to heed the recommendation of Corder et al. (2008) who suggest that the sampling rate should be as frequent as possible as data can always be re-integrated into a longer time frame, but not vice versa (Corder et al., 2008).

2.9.3.11 Wear time

Wear time refers to the sampling period required to produce a reliable estimate of habitual or 'usual' PA. Due to intra-individual variation in daily PA, multiple days are required to capture habitual PA. Trost, McIver, and Pate (2005a) reviewed youth studies examining monitoring time and concluded that 4-9 days are required to achieve reliability coefficients of 0.80. Mattocks et al. (2008a) found 2.9 days (600 min daily wear) produced r = 0.7 and 4.9 days produced r = 0.8 in 11-year-old children, suggesting 3 days of monitoring (2 weekdays and 1 weekend day) produce acceptable reliability and will reduce participant exclusion, whilst maintaining statistical power (Mattocks et al., 2008a). In younger children (5-6 years) Penpraze et al. (2006) reported 80% reliability for 3 days of monitoring and 84% when using 4 days, with greatest reliability achieved when ≥7 days were used. Clearly monitoring periods of ≥7 days will provide the most reliable estimates of habitual PA; however there must be a trade-off between accuracy and participant burden. Thus 3 days of monitoring with at least 1 weekend day appears an appropriate monitoring protocol for use in children.

2.9.3.12 Intensity cut-points

Despite much interest in value calibration there is little agreement between published intensity cut-off points - this has been termed 'cut-point non-equivalence' (Bornstein et al., 2011). The importance of this issue is highlighted by its appearance in numerous reviews (Freedson et al., 2005; Reilly et al., 2008; Ridgers & Fairclough, 2011). Cut-point non-equivalence between accelerometer models is expected as counts are tallied in manufacturer-specific units, yet cut-points also differ vastly within like models. Studies have shown the extent to which estimates of MVPA are dependent upon the cut-points applied, but the impact on PA-health relations is unclear (Stone, Rowlands, & Eston, 2009).

Intensity cut-points are dependent upon the type of calibration source activity used, age and maturational stage, gender, fitness level and body composition of the sample (Freedson et al., 2005; Welk, 2005) all of which can contribute to inter-individual variation in the relationship between counts and energy expenditure. Indeed Ekelund, Aman, and Westerterp (2003) found Actigraph counts ranged from ~400 to 2600 counts per minute for children walking at 4 $km.h^{-1}$. Efforts have been made to equate cut-points using regression models (Bornstein et al., 2011; Guinhouya et al., 2009). Guinhouya et al. (2009) modelled the bias in minutes of MVPA between cut-points at 100 $cts.min^{-1}$ increments, with the increment as the predictor (x) and mean bias as the predictive variable (y), however this was focused on hypothetical cut-points (3000-3900 $cts.min^{-1}$), not published cut-points. Bornstein et al. (2011) offered a direct conversion to predict MVPA estimates between five Actigraph thresholds. Despite this, there is a need for future work to present data using multiple cut-points and an on-going need for calibration studies to be conducted in large samples using standardised protocols.

2.10 Physical activity measurement summary

This review has summarised the strengths and weaknesses of the common field-based methods to assess PA in children. It is evident there are differences in the accuracy, feasibility and cost of each method, which are largely determined by the scale of the study in question. Evidently there remains no single method that can capture a perfect account of PA. The nature of children's PA, the multitude of activities engaged in and inherent limitations of each measurement method provide an upper limit to the accuracy of these tools (Welk et al., 2000). The choice of which tool to use should be informed by the construct (i.e. frequency, intensity, duration, type) that one is wishing to measure, its accuracy at measuring this, the richness of data that can be obtained and the feasibility and cost of deployment in the target population. It is however clear that accelerometers provide a balanced ratio between accuracy, richness of data and feasibility, and therefore considerable focus on these wearable motion sensors is warranted.

2.11 Promoting physical activity in children

Due to evidence for the health benefits of PA (Strong et al., 2005) and low levels of health-enhancing PA and cardiorespiratory fitness in contemporary children (Colley et al., 2011; Stratton et al., 2007) the development of interventions to promote PA in young people is a priority. Interventions are typically designed to influence a number of modifiable correlates of PA on the premise that altering these will induce behavioural change. This may include targeting psychological, social and environmental and policy level variables. Interventions to increase PA can be targeted at different levels and locations. These include individuals (e.g. individual counselling), clinical patients, groups (e.g. educational classes), communities (e.g. schools) and populations (e.g. policy change) (Biddle, Brehm, Verheijden, & Hopman-Rock, 2011). The majority of interventions target groups and are delivered in school settings as this is where children spend the majority of their waking hours and it lowers the risk of stigmatising low-active or overweight children (Dobbins, De Corby, Robeson, Husson, & Tirilis, 2009).

The growing appreciation of PA as a complex and multi-factorial behaviour has led to a shift from traditional structured exercise programmes (e.g. curriculum PE intervention) to more lifestyle-embedded PA interventions. As a result this has led to the use of body worn technologies that support behaviour change being used (Lutes & Steinbaugh, 2010). One device that has received increasing attention from 'exercise scientists' and health practitioners is the pedometer.

2.12 Pedometers as behavioural tools

Pedometers have conventionally been used to measure PA in children and adults; however they have now become used as a motivational tool to promote PA. In 2007 the first published meta-analysis of pedometer interventions in adults reported positive pedometer intervention effects in the order of +2491 (95% CI 1098-3885) steps per day compared to controls (Bravata et al., 2007), lending empirical evidence to the use of pedometers (in various guises) to promote PA. Similarly in a meta-analysis of adult and child data, Kang et al. (2009a) concluded that pedometer use is associated with a moderate and positive effect on PA. So what makes a pedometer an effective tool one may ask? Pedometers offer a real-time ambulatory measure of PA, providing feedback on steps taken. They permit an opportunity to self-regulate behaviour through the recording of step counts and setting of daily step goals (Tudor-Locke & Lutes, 2009). Pedometers themselves are simply step-counting devices, yet when utilised within a self-monitoring, feedback and goal-setting process, the pedometer provides a constant real-time barometer of PA behaviour that can motivate and hone behavioural choices (Tudor-Locke & Lutes, 2009).

2.13 Behavioural theories and pedometer programming

Behavioural models/frameworks should be used to guide intervention development so that specific correlates of behaviour are targeted, thereby increasing the likelihood of intervention success (Brug, Oenema, & Ferreira, 2005). In the case of PA there are numerous models that have been constructed in an attempt to explain PA behaviour. Currently there appear to be few pedometer interventions that have clearly acknowledged the use of a behavioural theory to guide intervention development. Tudor-Locke and Lutes (2009) found that of 27 pedometer studies reviewed only four outlined use of a theoretical model. Further, in an extensive Medline search, Lutes and Steinbaugh (2010) found only 31 pedometer studies in all populations that were theory driven. Particular theories identified by Lutes & Steinbaugh (2010) include social cognitive theory, the theory of planned behaviour, the transtheoretical model and self-regulation theory.

There is a growing argument that behavioural science is too focused on single-theory approaches as opposed to a problem-driven approach, whereby interventionists try to influence determinants (on multiple levels i.e. demographic, psychological, social and physical environment) of PA derived from assorted health behaviour theories (Brug et al., 2005). Coupled with a growing interest in the development of theory-driven interventions, there is interest in the identification of active intervention components as many interventions are poorly described and it is deemed essential to ensure that clarity is gained as to effective strategies for modifying correlates of PA (Michie, Fixsen, Grimshaw, & Eccles, 2009b). This is important for a cumulative science of behaviour *change* to be built, as presently there is a reliance on trying to explain/predict PA behaviour as opposed to understanding determinants of PA and what practical methods best modify these.

On this basis Michie, Abraham, Whittington, McAteer, and Gupta (2009a) sought to determine the most effective PA and dietary behaviour change techniques (BCTs) in a meta-analysis of experimental or quasi-experimental studies in adults. They reported that participants in interventions employing self-regulatory BCTs reported significantly better outcomes than controls. Combining self-monitoring with one or more of four other hypothesised self-regulation techniques, namely prompting intention formation, goal setting, feedback on performance and reviewing previously set goals showed a greater effect size (29 PA interventions, n = 5108) of d = 0.38 (95% CI = 0.27-0.49) compared to interventions that did not (40 PA interventions, n = 13,222), d = 0.27 (95% CI = 0.21-0.34). Thus including self-monitoring of behaviour and at least one other self-regulatory technique appears to be a powerful tool for behaviour change (Michie et al., 2009a).

As noted earlier pedometers permit the end-user the opportunity to self-regulate PA through self-monitoring of daily steps, the setting of daily step goals and receipt of immediate behavioural feedback from the step counter. The small changes model (shown in Figure 2.3 below) proposed by Lutes & Steinbaugh (2010) parsimoniously highlights the process by which pedometers can be used to change behaviour. Following baseline assessment of behaviour and ascertaining one's typical daily step total, an individual can begin to self-monitor. Daily step goals can be set on a self-selected basis (personalised/tailored), as increments over individual baseline values (standardised), as a percentage increment over individual or group baseline values (individual-standardised, group-standardised) or as a pre-set uniform goal (e.g. 10,000 steps.day^{-1}). One's accrual of steps towards a daily step goal can therefore be monitored through step-count feedback, and further with the recording of steps in a step diary (self-monitoring). From cognisance of real-time behaviour, lifestyle choices can be made to engineer opportunities to engage in more PA (Tudor-Locke & Lutes, 2009), much like a traditional negative feedback loop.

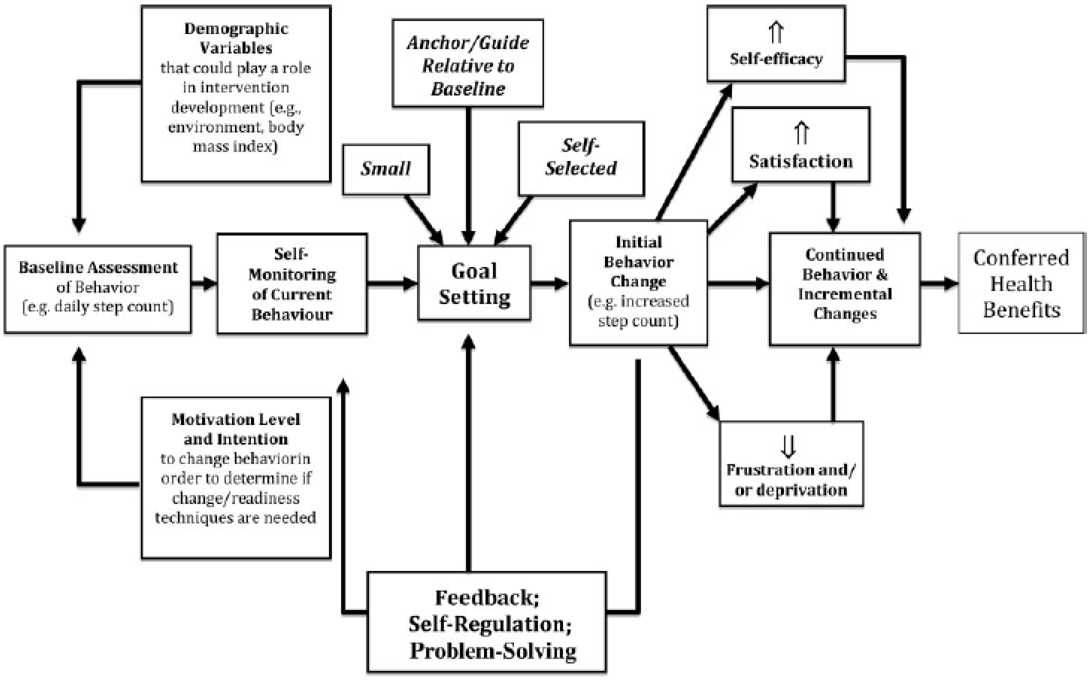


Figure 2.3 The Small changes model in reference to pedometer based self-regulation (Lutes & Steinbaugh, 2010).

The self-regulatory power of pedometry makes it an attractive tool for intervention designers. However if the optimal deployment of pedometers (i.e. step-goal-type, method of self-monitoring step counts etc.) are to be determined they must be tested separately from interventions that are poly-theoretical and target multi-level determinants using multiple BCTs. Indeed assessment of specific behavioural components as advocated by Michie et al. (2009b) requires the implementation of pedometer strategies alone. The following section will review extant pedometer studies in order to illuminate upon the current evidence base for pedometer-based programming in children.

2.14 Pedometer-based interventions in children

To date there have been almost 20 published pedometer interventions in children and adolescents (Berry et al., 2007; Butcher, Fairclough, Stratton, & Richardson, 2007; Goldfield, Kalakanis, Ernst, & Epstein, 2000; Goldfield et al., 2006; Hardman et al., 2011; Horne et al., 2009; Kang & Brinthaupt, 2009; Lubans & Morgan, 2008; Lubans, Morgan, Callister, & Collins, 2009a; Oliver, Schofield, & McEvoy, 2006; Rodearmel et al., 2007; Rooney, Gritt, Havens, Mathiason, & Clough, 2005; Schofield, Mummery, & Schofield, 2005; Shimon & Petlichkoff, 2009; Southard & Southard, 2006; Tsiros et al., 2008; Zizzi et al., 2006). To identify pedometer interventions relevant to the target age range a systematic search of the online electronic databases PubMed and SPORTDiscus was conducted. Keywords used were 'Pedometer' AND 'Study' AND 'Child'. To denote primary school aged-children an age range of 6 to 12 years was used. Publications that included study participants outside of this age range were excluded. All articles were limited to those available in English and conducted in humans. No limits were placed upon the publication date. The search process retrieved nine relevant studies. In addition hand searches of identified studies' reference lists were conducted. The studies are presented under three themes which were identified following this literature search: curriculum-based studies, goal-setting-/self-monitoring-based studies and reinforcement-based studies.

2.14.1 Curriculum based

One study evaluated the impact of integrating pedometer strategies into the school curriculum. Oliver et al. (2006) enrolled 78 children (aged 8-10 years) into a school programme which aimed to implement a curriculum unit based on accumulated pedometer steps. Lessons required participation in PA linked to the subject discipline, during which participants wore an unsealed pedometer every day of the 4-week intervention. PA was measured by 3 days of sealed pedometer monitoring at baseline and end-point. Overall the intervention showed no change in pedometer step counts for the entire sample, however when analysing sub-groups, those who were low active (<15,000 steps.day^{-1}) were seen to increase daily steps. Boys who achieved < 15,000 steps increased their average step counts from 12,793 ± 1453 to 14,498 ± 2762 steps.day^{-1} (p = 0.04) and girls increased from 10,399 ± 4554 to 13,668 ± 2960 steps.day^{-1} (p = 0.01). Despite evidence of being effective in low-active children and verifying the feasibility of pedometer use in a school setting, there are some limitations. Firstly no control group was used, thus any change in daily step counts cannot be solely attributed to the intervention. Further no follow-up assessment was conducted. As such no indication of the long term-behaviour change is given which is a key recommendation of the systematic review by Lubans et al. (2009b).

2.14.2 Goal setting/self-monitoring based

Butcher et al. (2007) randomly allocated 177 children (aged 9.1 ± 1.1 years) to one of three conditions. A sealed pedometer control, an open pedometer group (participants could view step counts, recorded them daily and were asked to increase steps the next day) or a pedometer plus PA information group (participants given ideas on how to increase PA). The 1-week school-based intervention showed that the feedback and information group (17.7 ± 4.87 steps.min^{-1}) achieved more steps.min^{-1} than the open pedometer (13.77 ± 4.06 steps.min^{-1}) or control (12.41 ± 3.12 steps.min^{-1}) group, but no differences were seen between the open pedometer and control group (p = 0.03). This suggests that feedback in the form of pedometer step counts alone may not be enough to promote changes in children's PA. Conclusions from this study are tempered by the short intervention period and the study only being implemented and evaluated within the school day. Further research is needed to confirm the effectiveness of receiving activity information alongside wearing a pedometer using a more sustained intervention including outside of school time.

Five studies have examined the effect of various combinations of pedometer feedback, goal-setting and self-monitoring strategies in children. Horne et al. (2009) tested a peer-modelling and rewards intervention in 100 primary school children (aged 9-11 years), based on the ‘Fit ‘n’ Fun Dude’ characters from a previously successful dietetic intervention. Forty-seven children were assigned to the 8-day intervention, and 53 to a control group. In the intervention group children were instructed to increase daily pedometer steps by 1500 per day over their baseline count, receiving a tangible reward (e.g. Frisbee) and verbal praise if they met or exceeded this target. Participants also received a ‘Fit ‘n’ Fun Dude’ home pack with peer-modelling ‘active’ character-based materials. The intervention was followed by a maintenance phase (11 weeks), during which no reward was given and step counts were self-monitored in a daily step diary. Physical activity in both groups was measured over 8 school days using pedometers, at baseline, intervention endpoint (baseline 2 for controls) and 12 weeks after baseline 2. Notably during the intervention both experimental boys and girls had significantly greater step counts than controls (Boys: +2785 steps over baseline, p = 0.001; Girls: +3822 steps over baseline, p = 0.001). However at 12-week follow-up girls in the experimental group maintained their increase in steps by 2873 steps.day^{-1} (26%) compared to baseline and had significantly higher step counts than control girls (p = 0.001), whilst boys displayed no difference over baseline (p = 0.468). Clearly this intervention shows promise for the maintenance of PA increases (particularly in girls) over a 3-month period. A stronger effect may have been gained by a longer intervention, as this was only a 2-week trial.

A further version of the ‘Fit ‘n’ Fun Dude’ study was examined by (Hardman et al., 2011) in a larger sample of 386 7-11-year-old primary school children. The children were randomised to a full intervention condition (n = 118) where they received ‘Fit ‘n’ Fun Dude’ peer-modelling materials and tangible rewards, and were given daily pedometer goals to attain in order to obtain rewards (i.e. +1500 of baseline steps); a no-rewards intervention group (n = 67) where children received only peer-modelling materials and the same pedometer goals; or an open pedometer control group (n = 51). Both intervention groups received praise contingent upon meeting daily step targets. The inclusion of a no-rewards group aimed to delineate the independent impact of tangible rewards combined with peer-modelling and pedometer strategies. The intervention phase lasted for 12 school days. This was followed by a maintenance phase (14 weeks) where participants self-recorded their daily step counts and received praise if they met or exceeded their step targets. Steps per day were significantly higher in the full intervention group (15,912 ± 3068 $steps.day^{-1}$) than in both the no-rewards group (13,995 ± 3719 $steps.day^{-1}$) ($p < 0.01$) and the control group (11,824 ± 2994 $steps.day^{-1}$) ($p < 0.01$) at the intervention phase (final week). However following the maintenance phase only the no-rewards group displayed a difference compared to the control group (15,307 ± 3550 vs. 13,403 ± 4238 $steps.day^{-1}$). There was no significant gender interaction, showing that the outcome did not differ between boys and girls. As with the previous 'Fit 'n' Fun Dude' study (Horne et al., 2009) an increase in PA (approximately +2000 steps) was observed 3 months post intervention, suggesting that pedometer self-regulation and goal setting combined with peer modelling without tangible rewards were sufficient to motivate children to maintain increased PA over time; the independent contribution of the pedometer components are however masked by the implementation of additional peer-modelling and contingent reward (praise/encouragement) strategies.

As part of a family-based coping skills intervention Berry et al. (2007) took a total of 80 parent and child dyads (aged 11-12 years) and randomised them to a nutrition and exercise group ($n = 40$) or a control group ($n = 40$). Children in both groups received nutrition education, exercise classes and behavioural modification education for a period of 24 weeks. Children in both groups were provided with a pedometer and step diary and encouraged to walk between 30-60 min a day. Self-report step count data were collected at baseline, 3 months (completion of the 12-week programme) and 6 months. Children in the experimental group increased daily steps by around 3000 steps, and those in the control group by around 2500 $steps.day^{-1}$, however no differences were seen between groups at the 6-month follow-up ($p = 0.3$). This shows that increases in pedometer steps can be attained without the use of structured goal setting, feedback and self-monitoring of steps.

In another family-based study Rooney et al. (2005) randomised 87 families (children aged 9-10 years) to a pedometer plus education (PPE), pedometer (P) or control group.

Those in the PPE and P groups were given pedometers to wear and were encouraged to walk 10,000 steps per day, with those in the PPE group also receiving six education sessions on healthy eating and exercise. Families recorded daily step totals and reported these bi-monthly. Results showed overall participants achieved the 10,000 steps.day^{-1} goal 48% of the time, with those in the PPE group displaying greater goal achievement compared to the pedometer only group (P) (53 vs. 43%, p = 0.012). Results for this intervention are poorly reported and it is unclear whether daily steps increased as a result of the intervention. However these data show that a generic step goal such as the 10,000 steps.day^{-1} concept may be an effective strategy.

Kang and Brinthaupt (2009) have provided the only examination of differing step goal conditions in children to date. They investigated the effect of a 6-week within-school pedometer intervention, comparing step counts between group-standardised and individual-standardised step goal conditions and the number of goal attainments between conditions and by PA level. Ninety nine boys and girls (aged 9-10 years) were cluster randomised by class. Those in the individual goal group were instructed to increase their baseline steps by 5%, after 2 weeks this was increased by 5% based on the previous 2-week average daily step count. Group-based goals instead took the class average, and increased by 5%. Participants in both conditions recorded their daily within-school step counts, at the end of every day. This was repeated for three 2-week testing periods. Prior to the intervention a 2-week baseline PA evaluation via pedometry was conducted. An average step count increase of ~19% from baseline (5454 ± 1432 steps.day^{-1}) to post intervention (6478 ± 2053 steps.day^{-1}) (p < 0.001) was observed. There was however no significant interactive effect between conditions and time (p = 0.845). Interestingly, for the least active children the number of goal attainments was higher for those with individual-based step goals (16.4 ± 3.1 goals attained) than those with group-based step goals (6.2 ± 4.9) (p < 0.001). However this difference in goal attainment between groups did not lead to differences in step counts between groups at end-point. A limitation of this study is that no follow-up was included, so the long-term sustainability of this intervention is unknown. Furthermore, a control group was not used, meaning that other factors may be responsible for the intervention effect as opposed to the pedometer intervention *per se*.

2.14.3 Reinforcement based

Two studies have examined the impact of an open-loop feedback-based intervention. Goldfield et al. (2000) randomised 34 obese children (aged 8-12 years) to one of three groups in a laboratory session which consisted of a 20-min activity phase followed by a 10-min sedentary reinforcement session. The groups were a contingent 1500 group (n = 12) which required participants to accumulate 1500 pedometer steps, with the rewards being 10 min of television activity time; a contingent 750 group (n = 13) which required an accumulation of 750 pedometer steps; or a control group wherein participants could access sedentary activities throughout the whole laboratory session. Physical activity was measured via accelerometer (TriTrac accelerometer) with total vector magnitude counts found to be higher in the contingent 1500 group (2200 counts) compared to the contingent 750 (1650 counts) or control (1100 counts) group ($p < 0.05$). Whilst this is a small-scale laboratory experiment involving sedentary reinforcement strategies, these data suggest that assigning higher step goals may motivate greater increases in PA.

In an extension of this work Goldfield et al. (2006) randomised 30 overweight/obese children (aged 10-11 years) to an open-loop feedback plus reinforcement group (n = 14) or a control group (n = 16). Children in the intervention group were provided with a Biotrainer activity monitor and set goals of 400 counts to permit access to 1 hr of television/VCR/DVD time. Those in the control group were provided with activity monitors but were not set activity goals, and had free access to television. The intervention lasted for 8 weeks with the children attending biweekly sessions to download activity monitors. PA outcome measures were assessed at baseline and at weeks 1-2, 3-5 and 6-8. Participants in the intervention group exhibited greater increases in mean PA counts (+160 vs. +33 $cts.day^{-1}$, p = 0.19) and minutes of MVPA

(+9 vs. +0.3 min, $p = 0.50$), but not VPA (+6 vs. +3 min, $p = 0.57$) from baseline to 8 weeks. Whilst again this sample is small and involved additional sedentary reinforcement strategies, these results provide support for the effectiveness for a generic goal-setting strategy.

2.14.4 Summary of pedometer-based interventions

From the studies reviewed it is clear that pedometer use in children is associated with increases in PA ranging in magnitude from +300 to +3000 steps.day^{-1} over baseline. At the upper limit this would equate to an increase in at least moderate-intensity activity of approximately 22 min (on the basis of (Jago et al., 2006)), which is a biologically significant volume when considering that increases in MVPA of 15 min are sufficient to lower odds of obesity by 40% in girls and 50% in boys (Ness et al., 2007).

Most studies reviewed employed pedometer goal setting alongside self-monitoring of step counts in a diary, and reported positive effects (Berry et al., 2007; Goldfield et al., 2000; Goldfield et al., 2006; Hardman et al., 2011; Horne et al., 2009; Kang & Brinthaupt, 2009). This suggests that chief strategies for increasing PA are the setting of a step goal, viewing step feedback and recording daily step totals in a step diary. These observations are in congruence with the conclusions of Bravata et al. (2007) whose systematic review showed that adult pedometer users who were given a goal (whether individualised or generic) and could view steps significantly increased steps over baseline, whilst those without a goal did not (Bravata et al., 2007).

An important observation is that low-active children may display the greatest increases in PA (Oliver et al., 2006). Limited evidence suggests that girls may be more responsive to pedometer use than boys, as in the Horne et al. (2009) study; only girls maintained significant increases in steps compared to baseline at 12-week follow-up - this may be a reflection of their lower baseline activity level. This is supported by the greater positive effect size for women (0.80) than reported for men (0.30) in a meta-analysis of adult intervention data (Kang et al., 2009a).

Clearly there is considerable heterogeneity in the deployment of pedometers in children, which somewhat clouds their use for current researchers. Whilst the use of a pedometer is associated with an increase in PA, many of the interventions employed additional behavioural strategies such as peer modelling and rewards (Hardman et al., 2011; Horne et al., 2009), behavioural re-enforcement (Goldfield et al., 2000; Goldfield et al., 2006) and education components (Oliver et al., 2006). The only study to employ goal setting, feedback and step diary use alone was Kang & Brinthaupt (2009). This study did not however include a control group. Kang & Brinthaupt (2009) also examined different goal types, showing that goal achievement was greater in low-active children when administered an individual-standardised goal (i.e. % increment of baseline values), but there was little difference between group- and individual-standardised goals in medium-, and high-active children. This suggests that individualised goal attainment is less influenced by baseline PA, whereas group-standardised goals may only be suitable for more active children. However no differences were reported in step counts between goal conditions.

The question is therefore to what extent is the type of step goal administered important? In low-active or overweight populations the setting of universal or group-standardised step goals may be unrealistic or unattainable, or conversely be too low for high-active children. Indeed, Tudor-Locke (2002) suggests that step goals should be personalised according to baseline values, specific health goals and sustainability. Whilst the impact of different pedometer step goals has been tested in children the relative effectiveness of goal type remains unclear. A recent review stated that further research is necessary to investigate the effectiveness of different types of pedometer step goals for enhancing habitual PA (Tudor-Locke & Lutes, 2009). Universal step goals (e.g. 10,000 steps.day^{-1}) are not in any way tailored to a population under study, and full personalised (i.e. self-set goals) goal setting may not be appropriate for primary school children as it requires advanced self-regulatory skills and researcher-led training. Of particular interest therefore is the relative impact of group-standardised and individual-standardised goals which are set as percentage increments over individual or group baseline values respectively.

There are a number of limitations in the studies reviewed that are worth highlighting. Many of the reviewed studies used small samples, and thus are at risk of being underpowered and at greater risk of type 2 error. Sample sizes ranged from 30 to 386 participants. Similarly, many of the experimental conditions were of short duration, ranging from 1 week to 12 weeks, which may not be sufficient to induce behaviour change. In a systematic review of pedometer interventions Kang et al. (2009) observed the largest effect size (0.76) for interventions that lasted >15 weeks. One must be cognisant however that there is a distinction between short-duration studies which aim to test optimal pedometer programming (Butcher et al., 2007) as opposed to those which are long-term complex behaviour change interventions (Hardman et al., 2011). Finally, few studies examined moderators of the main intervention effect such as gender, baseline PA and weight status. The determination of which sub-groups respond best to pedometer programming is of interest as many interventions may have failed to be successful due to the use of a blanket approach designed to maximise reach, as opposed to being specifically targeted. Indeed a recommendation of the review by Tudor-Locke & Lutes (2009) was that more research should be conducted to determine who benefits most from pedometer programming in order to target interventions appropriately.

One issue that is given little attention by prior pedometer studies is that pedometer steps only provide a volumetric index of PA. They offer no insight into PA intensity or pattern. Only one study used an accelerometer to measure change in PA (Goldfield et al., 2000). Pedometers are not designed to measure PA intensity and thus far only preliminary steps/min cut-points have been developed for interpreting children's time in MVPA (Jago et al., 2006). Whilst pedometer step counts have been related to some health outcomes such as body composition (Duncan, Schofield, & Duncan, 2006), time spent in activity of at least moderate intensity has been consistently linked to specific health benefits in children (Strong et al., 2005). It is therefore of interest to objectively measure change in MVPA within future pedometer-based interventions using accelerometry.

The use of multiple intervention components within the studies reviewed prohibits conclusions as to the optimal pedometer-based strategy. Bravata et al. (2007) suggested that because many adult interventions have used two or more pedometer strategies (e.g. pedometers, step goals, diaries, counselling), the independent contribution of these components is difficult to establish. Therefore Bravata et al. (2007) suggested future examinations should determine the impact of (1) pedometer use in which participants can see their daily step counts vs. pedometer use in which they are blinded to their daily step counts, (2) pedometer use with vs. without a step goal and (3) pedometer use with vs. without the use of step diaries. However, pedometer reactivity is known not to exist in children (Ozdoba et al., 2004) and there is evidence to suggest that adults assigned step goals increase daily steps, whilst those without goals, do not (Bravata et al., 2007). Bravata et al. (2007) also reported that participants who record daily step totals increase PA over baseline by a greater magnitude than those who do not. Whilst interesting theoretical questions *per se*, the merits of examining these issues alone are questionable. The 'power' of pedometer programming lies in viewing and recording of daily steps in pursuit of achieving a daily step goal. Without these components, behavioural change is unlikely to occur. From the current review of child-based interventions, questions of interest that this thesis will aim to investigate are:

> (1) What are the relative merits of wearing a pedometer, viewing daily step counts and self-monitoring with a step diary alone, versus simply wearing an open pedometer (control)?

And

> (2) What is the relative impact of differing step goal types (i.e. individual-standardised vs. group-standardised)?

2.15 Intervention moderating variables

As mentioned above few of the pedometer interventions examined potential moderating variables of the main intervention effect. The strength of the relationship between a pedometer intervention and programme outcome may vary according to a third variable (Bauman, Sallis, Dzewaltowski, & Owen, 2002). This is often referred to as an 'effect modifier' or more commonly a moderating variable (Bauman et al., 2002). The outcome of an intervention may differ between sub-groups of a study population, for example whilst it may lead to a positive change in women, no difference may be seen in men, despite the total sample showing a positive effect. In traditional epidemiologic studies researchers stratified data by sub-groups (e.g. smokers vs. non-smokers) to determine intervention moderation; now with complex multi-level data sets interaction terms are usually entered into statistical models (Bauman et al., 2002).

Knowledge of effect modifiers allows researchers to target sub-groups that respond to a particular intervention, so that maximum use is made of limited resources. Variables that might be of interest in children's PA interventions are gender, ethnicity, socio-economic status, baseline PA level and weight status. In particular there have been calls for future pedometer programmes to determine the effects of the intervention on children of differing weight status i.e. *healthy weight*, *overweight* and *obese* (Tudor-Locke & Lutes, 2009). This is deemed important as recent data (2009-2010) show that 14.6% and 18.7% of 10-11 year olds in England are overweight and obese respectively (NHS Information Centre, 2010). The following section will discuss the classification of weight status in children by proxy field methods.

2.16 Weight status classification

Several methods exist for the quantification of body fatness and weight status in children. Most of these methods are not frequently used due to their impracticality in the field and financial cost. These methods range from density-based methods (i.e.

hydrodensitometry, air displacement plethysmography), to scanning methods (i.e. magnetic resonance imaging, dual-energy x-ray absorptiometry), impedance methods (i.e. bioelectrical impedance) and anthropometric methods (i.e. skinfold, waist circumference, waist-hip ratio) (Sweeting, 2007).

2.16.1 Body mass index

Outside of the laboratory environment (and in large-scale studies) density or scanning-based methods are not practicable, therefore the use of impedance and anthropometric methods predominate. Generally proxy-based indices of body fatness are used. The most extensively used index is a measure of body weight relative to height referred to as body mass index (BMI). BMI is a surrogate measure of body fatness as it does not provide a direct measure of adiposity. BMI has been correlated with both fat mass index (FM) and fat free mass index (FFM) in children and adolescents (5-18 years), with the accuracy of BMI varying according to the degree of fatness (Freedman et al., 2005). That is, BMI is more strongly correlated with fat mass in heavier children (e.g. BMI-for-age ≥75th percentile) and the difference in fat free mass between the lowest and the middle BMI-for-age is greater than differences in fat mass (Freedman et al., 2005).

Therefore a primary limitation of BMI is that it cannot distinguish between fat mass and fat free mass. For a given absolute BMI value individuals may differ in the relative distribution of fat mass and fat free mass (Freedman et al., 2005). BMI varies between sex, age and ethnicity and is not static during childhood; indeed there is usually a sharp rise in BMI during infancy, declining until pre-school age, followed by a progressive increase through the remainder of childhood and adolescence (Cole, Freeman, & Preece, 1995). This increase in BMI may not be due to an increase in fat mass however as Freedman et al. (2005) reported that age differences in BMI among boys from age 12 were largely due to an increase in lean mass (even among the heaviest boys) (Freedman et al., 2005).

In adults cut-points of 25 and 30 kg.m^2 are used to define '*overweight*' and '*obesity*', and are related to an increased risk of chronic disease and mortality (Calle, Thun, Petrelli, Rodriguez, & Heath, 1999). As BMI differs between boys and girls and increases with age, the assessment of absolute values that represent a high or excess BMI is problematic. In children, overweight and obesity is defined relative to children of the same age or gender using a reference population, where BMI is expressed as age- and gender-specific *z*-scores (SD from mean BMI value) and percentiles, permitting the comparison of a child's BMI relative to a reference data set. In the UK, the most widely used reference population is the British 1990 growth reference dataset (UK90) (Cole et al., 1995), which was created from a sample of 30,535 (age 0-23 years) collected between 1978 and 1990 (Cole et al., 1995).

Typically '*overweight*' is defined as a BMI ≥85th centile and *obese* as ≥95th centile (Barlow & Dietz, 1998), with the 91st and 98th percentile recommended for clinical assessment. These cut-points were not based on associations with disease outcomes, but a review by Reilly et al. (2003) showed that a high BMI for age increases risk of a variety of morbidities in childhood and adolescence. Alternatively the cut-points provided by the International Obesity Task Force (IOTF) have been used in the UK (Cole, Bellizzi, Flegal, & Dietz, 2000). On behalf of the IOTF Cole et al. (2000) produced absolute cut-points for *overweight* and *obesity* that are anchored to BMI values of 25 and 30 kg.m^2 at age 18, based upon an international dataset (Brazil, Great Britain, Hong Kong, Netherlands, Singapore and United States) of 192,727 participants (aged 0-25 years). These absolute values have been shown to be related to an increased risk of mortality in adults (Calle et al., 1999).

The performance of the 95th percentile (relative to UK90 data) and IOTF cut-points to diagnose obesity accurately has been assessed by quantifying the sensitivity (ability to correctly identify obese children) and specificity (ability to correctly identify non-obese children) against a biological definition of excess body fatness (>95th percentile % body fat distribution) (Reilly, Dorosty, & Emmett, 2000). Importantly (Reilly et al., 2000) found that diagnostic accuracy did not differ between boys and girls and sensitivity was greater when using the UK90 cut-point compared to the IOTF criteria (88% vs. 46% (boys) or 72% (girls); i.e. 46/100 true positives). This suggests that using the IOTF cut-points in UK children may underestimate the true prevalence of obesity, and serves to highlight the fact that the prevalence of obesity will vary with the method used to define obesity.

Many studies that use BMI however make no reference to the undependability (physiologic variation (see (Ulijaszek & Kerr, 1999)) of this index due to the influence of daily variation in weight and height. This may well be attributed to the relative simplicity of weight and height measurement and the belief that any error resultant from imprecision (observer error) and undependability will be insignificant. Nevertheless it has long been observed that height fluctuates throughout the day hours (De Puky, 1935) due to changes in the intervertebral discs (Adams, Dolan, & Hutton, 1987).

Diurnal height loss has been reported in a number of studies of children and adolescents (Lampl, 1992; Siklar, Sanli, Dallar, & Tanyer, 2005; Voss & Bailey, 1997; Whitehouse, Tanner, & Healy, 1974). The average daily height loss from morning to afternoon/evening has been reported as being between 0.4-1.5 cm in child populations (Lampl, 1992; Siklar et al., 2005; Voss & Bailey, 1997; Whitehouse et al., 1974). In conjunction with height loss it well accepted that weight varies diurnally (Dittmar, Raschka, & Koch, 2002).

As BMI is calculated from height and weight it is entirely possible that a synergistic variation in these variables may result in the differential classification of weight status based upon the time of day when an individual is measured. To date no studies have investigated concurrent changes in weight and height and the effect of these variations on BMI-determined weight status when children are measured at different times of the day. Therefore it is uncertain as to whether a child classified as normal weight (and on the cusp of overweight e.g. 84th percentile) when measured in the morning would still be classified as normal weight when re-measured in the afternoon. Notionally an increase in weight coupled with a decrease in height would result in a greater BMI percentile score in the afternoon than the morning. If the timing of measurement is not standardised by researchers and there is sufficient variation in BMI percentile to affect weight status classification, this may bias intervention effects within BMI-determined weight status groups. Unreliable weight status classification would also make fair comparison of moderation effects between studies or differing intervention conditions difficult. There is therefore need to determine the potential impact of diurnal variation in height and weight upon the reliability of BMI-determined weight categorisation.

3 Chapter Three - Generic Methods

3.1 Introduction

The aim of this chapter is to describe the generic methods used for the measurement and processing of anthropometric and PA data in this thesis. Further, a brief technical specification is given for the Actiwatch 4 accelerometer (AW4) and Omron pedometers (HJ-104 & HJ-109-E) used.

3.2 Anthropometric measures

3.2.1 Height

Height was measured to the nearest 0.1 cm using a freestanding portable stadiometer (Seca 214, Seca Ltd, Leicester, UK). Participants removed their shoes and stood fully upright on the footboard with their back touching the vertical structure of the stadiometer and their feet slightly splayed and heels flush against the base. The head was positioned in the Frankfort plane and the headboard pushed down onto the crown of the head compressing the participant's hair. All hair clips and associated jewellery that inhibited the headboard were removed prior to measurement. Participants then relaxed their posture, again stood fully upright and a repeat measurement was taken. Repeat values were required to be within 0.4 cm of the first as per Mirwald, Baxter-Jones, Bailey, and Beunen (2002). From the two measurements the mean value was calculated as per Mirwald et al. (2002).

3.2.2 Weight

Weight was measured to the nearest 0.1 kg using electronic weighing scales (Tanita HD 352, Tanita Corporation, Tokyo, Japan). Participants removed their shoes and jumper/blazer and any items in their pockets, stood in the centre of the scales without support and were asked to distribute their weight evenly across both feet. Participants then dismounted from the scales and were then asked to step onto the scales again so that repeat measurements could be taken. Repeat values were required to be within 0.1 kg of the first. Where two measurements were taken the mean value was calculated as per Mirwald et al. (2002).

3.2.3 Body mass index

Body mass index (kg.m^2) was calculated by dividing the mean weight value in kilograms for each participant by the square of the mean height in metres. Absolute BMI values were then expressed relative to decimal age (at time of measurement) and gender as a centile of the British 1990 growth reference dataset (Cole et al., 1995) using the 'LMS Growth' calculator for Microsoft Excel (Pan & Cole, 2009).

3.2.4 Body mass index-determined weight status

Following the calculation of BMI-for-age (BMI percentile), participants were categorised as *underweight* (BMI ≤2nd centile), *healthyweight* (>2nd centile to <85th centile), *overweight* (≥85th centile to <95th centile) or *obese* (≥95th centile). These BMI centile cut-points to define weight status are known as 'population monitoring' cut-points and are the most frequently used in epidemiologic literature. The use of ≥95th centile to define obesity has been recommended by an expert committee in the United States (Barlow & Dietz, 1998).

3.3 Physical activity measures

For Chapters Five, Six and Seven the Actiwatch 4 accelerometer was used to objectively measure PA. The following section will detail the technical specifications of the Actiwatch 4 and Omron Pedometers (used in Chapter Seven), and the data collection and processing procedure for each relevant chapter.

3.3.1 Actiwatch 4 accelerometer

The Actiwatch 4 (AW4, Cambridge Neurotechnology, Cambridge, UK) is a small (37 x 29 x 10 mm) waterproof (1 m) accelerometer-based PA monitor (see Figure 3.1 below) which is 16 g in weight and has a random access memory (RAM) capacity of 64 kb. The accelerometer contains a rectangular piezoelectric plate, with a seismic mass attached to one end. When the device undergoes acceleration the seismic mass causes the piezoelectric beam to deform (bend), which creates an electric charge to gather on the plate which is proportional to the magnitude of the applied acceleration. This technology is sensitive to movement in all directions (vertical, medio-lateral and anterior-posterior) but is most sensitive to movement in the vertical axis, perpendicular to the longest dimension of the casing (Cambridge Neurotechnology, 2007) as shown in Figure 3.2.

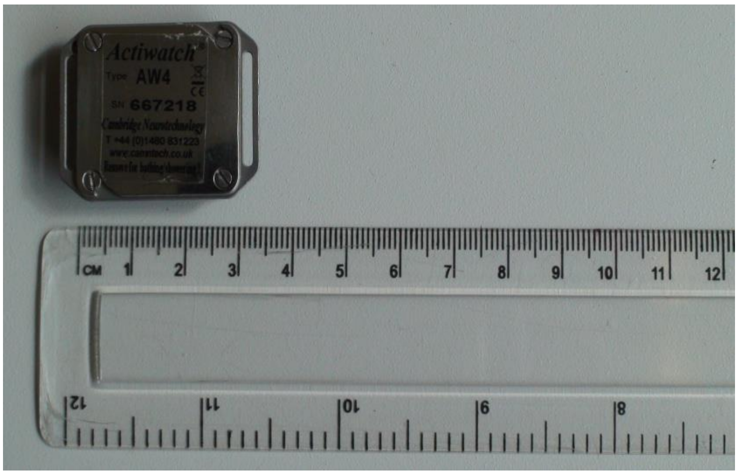


Figure 3.1 Actiwatch 4 accelerometer orientated with attachment surface facing upwards.

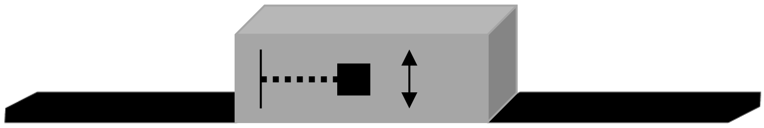


Figure 3.2 Schematic diagram of the Actiwatch 4 illustrating the orientation of the piezoelectric plate (dashed line) and attached seismic mass (small black square), with the double-headed arrow depicting the most sensitive axis, redrawn and adapted from Cambridge Neurotechnology (2007).

The raw voltage signal generated by the sensor is filtered and amplified and then passed to an analog-to-digital converter (A/D conversion), which converts the raw continuous signal to a digital string of numbers (discrete steps) which are called 'raw counts' (Chen & Bassett, 2005). Each raw count amplitude can range from -128 to +128; these values undergo full wave rectification so that all negative counts are converted to positive counts (Chen & Bassett, 2005). The AW4 repeats this process at a rate of 32 Hz, with the peak count carried forward and recorded from all 32 samples for each second of sampling. The peak counts for each second are then integrated over the user-defined time window (epoch), which can range from 2 to 15 min in the AW4, dependent upon the desired monitoring period (Cambridge Neurotechnology, 2007). The device has a frequency range of 0.5-7.0 Hz, but it is unknown as to whether the acceleration signal is fully weighted across this frequency range.

3.3.2 Unit initialisation

To initialise the AW4 each unit was placed on a proprietary telemetric USB reader (Actiwatch Reader, Cambridge Neurotechnology, Cambridge, UK) to permit communication with a personal computer. Upon connection to the personal computer, the proprietary Actiwatch software (Actiwatch Activity & Sleep Analysis Version 7.2.7, Cambridge Neurotechnology, Cambridge, UK) was loaded and the battery checked for sufficient life. For each unit a unique user identity was entered, along with the start date, start time (to enable delayed initialisation) and epoch length required for the given monitoring period. Following initialisation the AW4 unit was removed from the USB reader and was ready for use.

3.3.3 Epoch choice

As discussed in Chapter Two a short epoch is preferable when measuring PA in children, with an epoch of ≤15 s being optimal to minimise the underestimation of MVPA. For all studies of this thesis that involved accelerometer data collection, an epoch of 10 s was used as this is the maximum resolution that can be used over a 7-day monitoring period due to the memory capacity of the Actiwatch 4 (64 kb).

3.3.4 Data download

Data captured by the AW4 was stored within the internal non-volatile memory and then downloaded to a personal computer using CamNtech's proprietary telemetric USB reader (Actiwatch Reader, Cambridge Neurotechnology, Cambridge, UK). All captured data files were saved in a raw format (.AWF) which preserves the data according to the specifications that were entered upon initialisation (e.g. sampling rate). For Chapters Five and Six the raw files were then exported into Microsoft Excel for further analysis, whereas for Chapter Seven the raw AWF files were converted to a .DAT format to permit the assessment of habitual PA using a dedicated accelerometer data processing program (MAHUffe; available from www.mrc-epid.cam.ac.uk). The data processing procedures unique to each chapter are described in further detail below.

3.3.5 Data processing (Chapters Five and Six)

All AWF files were exported to Microsoft Excel for further analysis. Opened in Microsoft Excel, AWF files are a single continuous column of activity count data points (see Figure 3.3 below), with a header detailing the file name, date and time of initialisation, sampling rate and AW4 serial number.

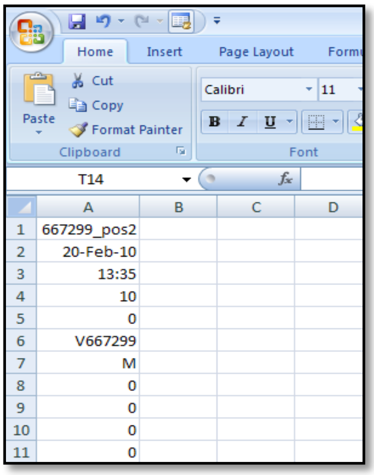


Figure 3.3 Raw Actiwatch 4 AWF file opened in Microsoft Excel.

In this format total activity counts were calculated by summing the individual data points, and mean activity counts were calculated by dividing total activity counts by the number of data points. Time above a given activity intensity (i.e. minutes of MVPA) was calculated by counting the number of data points above the specified cut-point (e.g. >116 to <417 counts) and multiplying this number by 10 to determine seconds, the sum of which was divided by 60 to obtain minutes. All files were checked for marker 'M' values which appear when the marker button on the AW4 is pressed. These values were deleted and checks for spurious data >2500 counts per epoch (Esliger et al., 2005) were made.

3.3.6 Data processing (Chapter Seven)

Prior to conversion (to .DAT files) all AWF files were opened in Microsoft Excel and checked for marker 'M' values. These values were deleted and checks for spurious counts (as per above) were conducted. Then all raw AWF files were converted to a .DAT file format using a custom software program (AWF2DAT, © JW Bevins, University of Worcester). This converted the single column of data points in the AWF format into 17 separate columns, with the data now being read in rows from left to right. By manually entering the initialisation time, date, epoch, download time and date, the converted .DAT files could be imported into the MAHUffe software for further data reduction. This software package allows the user to enter data reduction rules, such as intensity-cut-point values, wear time requirements, epoch integration and motionless data filters. The data reduction choices are outlined in more detail in Chapter Seven. Following input of the data reduction rules, files that met the validation criteria were batch processed. The batch processing command produced a summary file with outcome variables of total activity counts, mean activity counts per minute, minutes in each intensity category and registered wear time for each participant.

3.3.7 Omron HJ-104 pedometer

The Omron HJ-104 pedometer (64.0 x 30.2 x 21.0 mm) is a hair-spring lever arm electronic pedometer that weighs 25 g (including batteries) and is powered by a 1.5 v lithium battery (see Figure 3.4 below). The device tallies step counts and displays them on an electronic liquid crystal screen. A step is recorded when the lever arm is deflected vertically by a movement of sufficient magnitude to exceed the force sensitivity threshold, which is unknown in this model. The HJ-104 has the capacity to store total daily steps for a 7-day period and can measure distance walked in km if the participants stride length is entered. The HJ-104 has a sensitivity switch which enables the end-user to adjust to a ±5% accuracy rate (against 100 steps), to permit individual calibration. When deployed in the data collection of Chapter Seven, all switches were set to the middle setting (i.e. zero sensitivity adjustment).

3.3.8 Omron HJ-109-E pedometer

The Omron HJ-109-E pedometer (63.5 x 36.3 x 23.0 mm) is a hair-spring lever arm electronic pedometer that weighs 24 g (including batteries) and is powered by a 1.5 v lithium battery (Omron, 2003). The HJ-109-E (see Figure 3.4 below) is identical to the HJ-104 in its technical specification and memory capacity, except that it also has an 'aerobic steps' function which measures cumulative steps following 10 min of interrupted step counts. Further an estimate of total energy expenditure (kcal) and body fat oxidation (g) is provided for the period of accumulated 'aerobic steps', calculated from body weight, stride length, step counts and step frequency using a proprietary algorithm (Omron, 2003). As with the HJ-104 When deployed in the data collection of Chapter Seven, all switches were set to the middle (i.e. zero sensitivity adjustment) setting.

Figure 3.4 Omron HJ-104 (left) and Omron HJ-109-E pedometer (right).

3.3.9 Pedometer data processing (Chapter Seven)

Pedometer data in the form of daily step count totals were recorded by the pedometers on-board memory function for the 5-day monitoring period. Further data-reduction decisions are presented in Chapter Seven.

4 Chapter Four - Study One: The impact of school-day variation in weight and height upon body mass index-determined weight category in 10-11-year-old children

4.1 Introduction

There is growing concern over increasing levels of childhood obesity in the UK (NHS Information Centre, 2011). This interest in the childhood obesity epidemic is founded upon the deleterious health effects believed to be associated with excess body fatness. A systematic review of epidemiologic studies found excess body fatness to be positively associated with an increase in psychological morbidity, inflammation and cardiometabolic risk during childhood, and greater risk of premature mortality in adulthood (Reilly et al., 2003).

Higher levels of PA, particularly of at least moderate intensity, are negatively associated with fat mass in children and adolescents (Ness et al., 2007; Riddoch et al., 2009), however reverse causality has also been shown by some research groups (Metcalf et al., 2011). PA-based preventive efforts often target a reduction in body fatness as a primary outcome (Harris, Kuramoto, Schulzer, & Retallack, 2009). As obese children typically display lower PA levels than non-obese children (Cooper, Page, Fox, & Misson, 2000; Page et al., 2005) and may respond differently to PA-based interventions (Lubans, Morgan, & Callister, 2012) the use of weight-status categorisation as an intervention effect modifier (moderating variable) is of interest. Indeed recent pedometer-based literature has recommended that future interventions should examine main intervention effects by weight status group (i.e. overweight, obese) (Horne et al., 2009; Tudor-Locke & Lutes, 2009).

For the assessment of weight status and obesity the most widely employed field measure is body mass index (BMI), which is the ratio of body weight relative to the square of height. BMI is not a stable during childhood and varies by age and sex (Freedman et al., 2005). To determine 'overweight' and 'obesity' a children's BMI value is expressed as an age- and gender-specific percentile of a reference data set. In the

UK, the most used reference population is the British 1990 growth reference dataset (UK90) (Cole et al., 1995).

For the purpose of reporting prevalence estimates epidemiologists typically categorise the weight status of children according to ‘population monitoring’ BMI cut-offs using the 1990 UK growth reference dataset (Cole et al., 1995): *Underweight* = BMI ≤2nd centile; *Healthy weight* = >2nd centile to <85th centile; *Overweight* = ≥85th centile to <95th centile; *Obese* = ≥95th centile. Alternatively clinicians and surveillance initiatives such as the National Child Measurement Programme will use ‘clinical’ BMI cut-offs: *Underweight* = BMI ≤2nd centile; *Healthy weight* = >2nd centile to ≤91st centile; *Overweight* = >91st centile to ≤98th centile and *Obese* = >98th centile (Dinsdale, Ridler, & Ells, 2011).

The reliability and accuracy of BMI-determined weight status is a function of measurement error related to imprecision (intra-/inter-observer error) and undependability (biological variation) (Ulijaszek & Kerr, 1999). Past anthropometric research has focused upon sources of error attributable to the measurement instrument, the observer, the measurement protocol and the participant (Voss & Bailey, 1997; Voss, Bailey, Cumming, Wilkin, & Betts, 1990). A source of bias that will impact on the calculation of BMI is the diurnal variation of height. It has long been observed that height alters throughout the day, lengthening during sleep and shortening during waking hours (De Puky, 1935) which can be attributed to changes in the water content of the inter-vertebral discs, compression of inter-vertebral cartilage and loading of the inter-vertebral discs (Adams et al., 1987; Park, 1997).

Diurnal height loss has been reported in a number of studies of children and adolescents (Lampl, 1992; Siklar et al., 2005; Voss & Bailey, 1997; Whitehouse et al., 1974). One of the early studies by Whitehouse et al. (1974) measured 11 boys (aged 12-14 years) at 10:00 and 17:00 on four occasions, 3-4 months apart. A further 19 boys (aged 12-14 years) were measured at 09:30 and 14:00 on a single occasion. A significant difference of 0.2 cm (95% range: -0.1 to 0.6 cm) was observed in height between subjects over the half-day measures (09:30-14:00); however the mean change of 0.46 cm (0 to 1 cm) across the whole day (10:00-17:00) was not significant ($p > 0.05$). As the standard error of measurement was reported as ±0.18 cm, it was concluded that diurnal variation in height was likely due to measurement error (i.e. positioning of the child), as opposed to height loss *per se*. In a larger sample, Voss and Bailey (1997) measured 53 children (aged 3-11 years) in a single day. Each subject was measured on four occasions (09:00, 11:00, 13:00 and 15:00) by two observers, in a random order, using both a stretched and an unstretched technique. As per previous studies, the greatest height loss reported was in the morning, and the largest decrement was between the hours of 09:00 and 11:00 (Whitehouse et al., 1974). Compared to the unstretched measurements, the stretched technique increased the height measurement by ~0.3 cm. However mean height loss over the whole 6-hr period was similar for both stretched (0.551 cm, range: −0.6 to 1.8 cm) and unstretched (0.555 cm, range: −0.4 to 1.9 cm) measurements.

More recently a large sample of Turkish children (n = 478, aged 9.9 ± 2.3 years) were measured twice in one day between 09:00 and 10:00, and between 15:00 and 1600, by one observer (Siklar et al., 2005). A moderate height loss was found for the whole sample (0.47 ± 0.05 cm; $p < 0.05$) (for girls 0.48 ± 0.04 cm and for boys 0.47 ± 0.08 cm). There is no consensus however as to the expected height loss that may occur in a given individual. Nonetheless the average daily height loss from morning to afternoon/evening has been reported as being between 0.4 cm and 1.5 cm in child populations (Lampl, 1992; Siklar et al., 2005; Voss & Bailey, 1997; Whitehouse et al., 1974). More solid conclusions can be made regarding the time of greatest height decrement, which has been consistently reported as during the first three waking hours (Tillmann & Clayton, 2001; Voss & Bailey, 1997; Whitehouse et al., 1974).

In conjunction with body height, it is accepted that body weight varies within and between days (Cheuvront, Carter, Montain, & Sawka, 2004; Dittmar et al., 2002). Acute variation in body weight occurs hourly due to changes in hydration status, dietary intake, and participation in daily PA. If one considers the additive effect of equipment-related measurement error, biological variation of both height and body weight and intra-/inter-observer error, there is potential for a child's BMI category to be differentially classified dependent upon the time of day when measurements are taken. It is consequently unknown as to whether a child who is at the extreme centiles within a weight status group may shift weight category between measurements that are conducted in the morning and afternoon.

4.2 Study aim

To date no studies have investigated concurrent changes in weight and height in direct relation to their impact on BMI weight category. The aims of the current study therefore, were to determine the impact of school-day variation in weight and height on BMI-determined weight category in 10-11 year old children and to investigate if the magnitude of any such variation differed between BMI-determined weight categories.

4.3 Methods

4.3.1 Participants

Seventy-four Year 6 children (Boys (n = 34; height: 141.16 ± 7.45 cm; weight: 36.48 ± 9.46 kg, BMI: 18.19 ± 3.98 kg.m^2) and girls (n = 40; height: 144.58 ± 7.66 cm; weight: 42.25 ± 11.29 kg; BMI: 19.97 ± 3.98 kg.m^2)) aged 11 ± 0.3 years were recruited from three primary schools in the West Midlands region of England. In loco parentis consent from the school head teacher and/or assumed parental consent were obtained and the study was approved through institutional ethics procedures.

4.3.2 Study design

Data were collected across four separate school days with standing height (cm) and body weight (kg) measured between 09:00 and 10:45 in the morning school session (HEIGHT-am, WEIGHT-am) and again between 13:20 and 14:50 in the afternoon school session (HEIGHT-pm, WEIGHT-pm). Intra-observer-/equipment-related measurement error was estimated through repeating the morning measurements within 1 hr of the original morning measurement in a sub-sample of 35 children.

4.3.3 Procedures

All anthropometric measurements were taken by one researcher and the weighing scale and stadiometer were calibrated on each test day. Participants wore light clothing and removed shoes, hair ornaments and any objects from their pockets.

Height and weight were measured according to the measurement protocols outlined in Chapter Three. Repeat measurements were performed for each variable at each time of measurement with the second measurement required to be within 0.4 cm or 0.1 kg of the first for height and weight respectively. Mean values were used for all subsequent analyses (Mirwald et al., 2002). Participants were measured in a randomised order at each session and were blind to the aims of the study, only being told they would be measured multiple times. The researcher was blind to any previous measurements.

4.3.4 Body mass index

Body mass index ($kg.m^2$) was calculated for each measurement session (BMI-am and BMI-pm) according to the protocol outlined in Chapter Three. These data were then used to express BMI as a centile based on the British 1990 growth reference data, which was calculated as per Chapter Three (Cole et al., 1995). Both the population monitoring and clinical BMI cut-offs, as described above, were used to determine participant's BMI weight category. These were expressed as BMI morning session category (BMICATClin-am; BMICATPop-am) and BMI afternoon session category (BMICATClin-pm; BMICATPop-pm).

4.3.5 Measurement error

To investigate intra-observer measurement error the technical error of measurement (TEM) was calculated as the square root of the sum of the squared differences between the two repeat measures (sub-sample of 35 children) divided by two times the sample size according to Ulijasek & Kerr (1999). The coefficient of reliability, as a unitless measure of anthropometric precision was calculated as 1 minus the TEM squared divided by the overall SD of the study, squared, according to Goto and Mascie-Taylor (2007). Ulijaszek and Kerr (1999) suggest a cut-off of 0.95 be used (i.e. human error of up to 5%).

4.3.6 Data treatment and statistical analyses

All statistical analyses were performed using SPSS 17.0 (SPSS, Chicago, IL, USA). To investigate intra-observer measurement error between the morning repeat measurements in the sub-sample of 35 children, intra-class correlation coefficients (ICC) with a two-way random effects model for absolute agreement were used to examine the strength of association between replicate measurements. In the main study, all mean measurement data and calculated BMI-am and BMI-pm variables were checked for normality of distribution using the Kolmogorov-Smirnov test. WEIGHT-am, WEIGHT-pm, BMI-am, BMI-pm, BMICATClin-am, BMICATPop-am, BMICATClin-pm, BMICATPop-pm, BMICentile-am and BMICentile-pm all violated the normality assumption. To investigate the difference between HEIGHT-am and HEIGHT-pm a paired samples *t*-test was used. To investigate differences between WEIGHT-am and WEIGHT-pm, BMI-am and BMI-pm, and BMI percentile-am and BMI percentile-pm Wilcoxon signed-rank tests were used. Change in BMICATClin-am to BMICATClin-pm and BMICATPop-am to BMICATPop-pm was investigated using Wilcoxon signed-rank tests, and individual level analysis. To investigate change from HEIGHT-am to HEIGHT-pm, WEIGHT-am to WEIGHT-pm, BMI-am to BMI-pm, and BMICentile-am to BMICentile-pm between BMI-am weight category groups using both BMI cut-offs, Kruskal-Wallis tests were used with Mann-Whitney *U*-tests as *post-hoc* tests where necessary. The alpha level for statistical significance was set at $p < 0.05$ for all tests. If a difference was found, the effect size was calculated. For parametric tests partial eta squared (small = 0.01, medium = 0.06, large = 0.14) was calculated. For non-parametric tests the approximate r statistic was calculated (small = 0.1, medium = 0.3, large = 0.5, Cohen, 1988).

4.4 Results

4.4.1 Intra-observer measurement error

Results from the intra-observer measurement error sub-sample analyses can be found in Table 4.I.

Table 4.1 Intra-observer reliability statistics for height and weight measurements (n = 35).

Reliability statistic	**Height (cm)**	**Weight (kg)**
SD of differences	0.48	0.20
TEM	0.33	0.14
R coefficient (*r*)	0.98	1.00
ICC (*r*)	0.99*	1.00*

*Significant at the $p < 0.05$ level.

The TEM was primarily used to quantify measurement error. Ulijaszek and Kerr (1999) suggest that for 10-11-year-old children the acceptable TEM threshold for height is 0.60-1.30 cm and 0.24-1.61 kg for weight. Moreover coefficients of reliability above 0.95 are indicative of good quality measurement control (Ulijaszek & Kerr, 1999). The TEM for height was 0.3 cm, and 0.1 kg for weight. The coefficient of reliability (r = 0.98) and the ICC (r = 0.99, $p < 0.05$) were above 0.95, therefore the measurement error observed in this study was deemed acceptable for the subsequent analyses of the main study measurements to be meaningful.

4.4.2 Height

Table 4.2 shows descriptive statistics for height, weight and BMI in boys, girls and the total sample and *p* values from inferential statistics.

Height decreased in 62 of the participants (83.8%), increased in 10 (13.5%) and did not change in two (2.7%), between am and pm measurements. Height decreased from am to pm measurements in the whole group ($t(73) = 8.2$, $p = 0.01$, partial eta squared = 0.5), in boys ($t(33) = 5.8$, $p = 0.01$, partial eta squared = 0.5) and in girls ($t(39) = 5.7$, $p = 0.01$, partial eta squared = 0.5).

4.4.3 Weight

Weight increased in 39 participants (52.7%), decreased in 29 (39.2%), and did not change in six (8.1%) between am and pm measurements. Weight did not change between morning and afternoon in the whole sample ($Z = -1.6$, $p = 0.09$) or in boys ($Z = -8.2$, $p = 0.41$) but increased in girls ($Z = -2.8$, $p = 0.01$, $r = 0.4$).

4.4.4 Body mass index

BMI increased in 60 participants (81.1%), decreased in 11 (14.9%), and did not change in three (4.1%) between the am and pm measurements. BMI increased in the whole sample from BMI morning to BMI afternoon ($Z = -5.8$, $p = 0.01$, $r = 0.6$) and in both boys ($Z = -3.2$, $p = 0.01$, $r = 0.5$) and girls ($Z = -4.8$, $p = 0.01$, $r = 0.7$).

4.4.5 Body mass index percentile

BMI percentile increased in 46 participants (62.2%), decreased in 7 (9.5%) and did not change in 21 participants (28.4%). Change in BMI percentile ranged from -11 to +11 centiles. BMI percentile increased in the whole sample (Mdn = 57.0th to 59.5th, Z = -4.4, p = 0.01, r = 0.5), in boys (Mdn = 45.5th to 47.0th, Z = -2.2, p = 0.02, r = 0.3) and in girls (Mdn = 77.0th to 78.5th, Z = -3.9, p = 0.01, r = 0.6).

4.4.6 Clinical cut-off body mass index category (91st and 98th centile)

BMI weight classification for boys, girls and total sample from the morning measurements (BMICATClin-am) and afternoon measurements (BMICATClin-pm) are shown in Table 4.3. BMI weight classification between the morning measurements (BMICATClin-am) and afternoon measurements (BMICATClin-pm) did not change (Z = -1.0, p = 0.32) in the whole sample. Change in BMICATClin from am to pm measurements was observed in one girl. The participant moved from the *healthy weight* to the *overweight* category with a BMI increase of 0.28 kg.m^2, a BMI-percentile increase of 1.2 centiles, a height decrease of 0.75 cm, and a weight increase of 0.20 kg.

There was no difference in the magnitude of height, weight or BMI change from am to pm measurements between BMICATClin-am groups (H(3) = 2.8, p = 0.41; H(3) = 6.6, p = 0.86; H(3) = 2.2, p = 0.53). The magnitude of BMI-percentile change was different between BMICATClin-am groups (H(3) = 19.3, p = 0.01). *Post-hoc* Mann-Whitney *U*-tests (controlling for type I error across tests using the Bonferroni-corrected alpha (0.05/3), $p \leq 0.02$) identified a difference between the *healthy weight* and *overweight* groups (U = 127.5, p = 0.01, r = 0.4), and *healthy weight* and *obese* groups (U = 61.0, p = 0.01, r = 0.3) with *healthy weight* children increasing in BMI centile (Mdn diff pm-am = +2 centiles) and both *overweight* and *obese* children displaying no change (Mdn diff pm-am = 0 centiles).

Table 4.2 School-day variation in height, weight and body mass index (BMI).

Variable	Statistic	Boys (*n* = 34)	Girls (*n* = 40)	Total (*n* = 74)
Morning height (cm)	*Mean ± SD* *95% CI* *range*	141.16 ± 7.45 (138.56 to 143.76) 125.55 to 163.50	144.58 ± 7.66 (142.13 to 147.03) 126.00 to 165.50	**143.01 ± 7.71** **(141.23 to 144.80)** **125.55 to 165.50**
Afternoon height (cm)	*Mean ± SD,* *95% CI* *range*	140.63 ± 7.34 (138.07 to 143.19) 124.95 to 161.45	144.08 ± 7.61 (141.65 to 146.51) 125.60 to 165.35	**142.50 ± 7.64** **(140.73 to 144.26)** **124.95 to 165.35**
Height change (cm)	*Mean ±SD* *95% CI* *range* *p**	-0.53 ± 0.53 (-0.35 to -0.72) -2.05 to 0.45 0.01	-0.50 ± 0.55 (-0.32 to -0.67) -2.50 to 0.75 0.01	**-0.51 ± 0.50** **(-0.39 to -0.64)** **-2.50 to 0.75** **0.01**
Morning weight (kg)	*Median* *(IQR)* *range*	32.70 (13.30) 26.60 to 71.70	41.40 (12.08) 23.70 to 75.40	**36.40 (13.63)** **23.70 to 75.40**
Afternoon weight (kg)	*Median* *(IQR)* *range*	32.60 (13.40) 26.50 to 71.60	41.60 (12.15) 23.90 to 75.90	**36.35 (13.54)** **23.90 to 75.90**
Weight change (kg)	*Median* *(IQR)* *range* *p**	-0.08 (0.24) -1.00 to 1.05 0.41 ns	+0.10 (0.30) -0.45 to 0.60 0.01	**+0.07 (0.30)** **-1.00 to 1.05** **0.09ns**
Morning BMI ($kg.m^2$)	*Median* *(IQR)* *range*	16.76 (4.21) 13.54 to 34.58	19.22 (6.02) 13.90 to 27.86	**18.04 (4.79)** **13.54 to 34.58**
Afternoon BMI ($kg.m^2$)	*Median* *(IQR)* *range*	16.88 (4.20) 13.65 to 34.87	19.42 (6.43) 13.90 to 27.94	**18.13 (5.03)** **13.65 to 34.87**
BMI change ($kg.m^2$)	*Median* *(IQR)* *range* *p**	+0.12 (0.24) -0.50 to 0.51 0.01	+0.18 (0.17) -0.27 to 0.63 0.01	**+ 0.16 (0.23)** **-0.50 to 0.63** **0.01**

*Time difference ($p < 0.05$).

4.4.7 Population monitoring cut-off body mass index category (85th and 95th centile)

BMI weight classification from the morning measurements (BMICATPop-am) and afternoon measurements (BMICATPop-pm) did not change (Z = -1.7, p = 0.08) in the whole sample. BMI category change from am to pm measurements was also only observed in girls. Three of the girls (7.5%) increased BMI weight category between morning and afternoon measurements with two moving from the *healthy weight* to *overweight* category and one moving from the *overweight* to *obese* category with BMI increases of only 0.30, 0.55 and 0.26 $kg.m^{-2}$ respectively.

There was no difference in the magnitude of height, or BMI change from am to pm measurements between BMICATPop-am groups (H(3) = 3.1, p = 0.38; H(3) = 2.9, p = 0.40). The magnitude of weight change was different between BMICATPop-am groups (H(3) = 9.4, p = 0.02). *Post-hoc* Mann-Whitney U-tests (controlling for type I error across tests using the Bonferroni-corrected alpha (0.05/3), $p \leq 0.02$) identified a difference between the *healthy weight* and *obese* groups (U = 265.0, p = 0.02, r = 0.3) with *obese* children increasing in weight, and *healthy weight* displaying no change (Mdn diff pm-am = +0.20 vs. 0.00 kg). The magnitude of BMI-percentile change was different between BMICATPop-am groups (H(3) = 19.3, p = 0.01). *Post-hoc* Mann-Whitney U-tests identified a difference between *healthy weight* and *obese* groups (U = 161.0, p = 0.01, r = 0.4), and *overweight* and *obese* groups (U = 12.5, p = 0.01, r = 0.4). *Healthy weight* children displayed a median increase in BMI percentile (pm-am) of 2 centiles, overweight an increase of 1 centile, and *obese* no change.

Table 4.3 Frequency and percentage of BMI weight categories (population monitoring cut-offs - 85th and 95th centile) in the morning and afternoon for boys, girls and total sample.

	Morning				Afternoon			
	UW	HW	OW	OB	UW	HW	OW	OB
Boys	1(2.9)	24(70.6)	2(5.9)	7(20.6)	1(2.9)	24(70.6)	2(5.9)	7(20.6)
Girls	1(2.5)	23(57.5)	5(12.5)	11(27.5)	1(2.5)	21(52.5)	6(15.0)	12(30.0)
Total	**2(2.7)**	**47(63.5)**	**7(9.5)**	**18(24.3)**	**2(2.7)**	**45(60.8)**	**8(10.8)**	**19(25.7)**

*UW = underweight, HW = healthy weight, OW = overweight, OB = obese.

4.5 Discussion

This is the first study to investigate the impact of the additive interactions between biological variation in height and weight upon BMI-determined weight category in children aged 10-11 years. As shown in Table 4.2, mean height decreased from morning to afternoon. The effect size was large, and the magnitude of diurnal variation in height observed parallels that found by Siklar et al. (2005) who reported a total diurnal height loss of -0.47 cm, and Voss and Bailey (1997) who also found a -0.555 cm decrease. In the present study boys' height decreased by 0.53 cm and girls' height decreased by 0.50 cm, which was comparable to the -0.47 cm for boys and -0.48 cm girls found by Siklar et al. (2005) and -0.46 cm for boys by Whitehouse and colleagues (1974).

Whilst no changes in weight were found for the whole sample or for boys, weight increased during the school day in the girls with the data showing a median difference of +0.10 kg (0.30 IQR). This may be an indication of greater propensity for weight variation in girls compared to boys, although the reasons for such a gender-dependent increase are unknown as food and fluid intake and energy expenditure between the morning and afternoon measurements were not assessed as part of the current study.

These measurements were considered but were not included due to the possible impact this obvious 'monitoring' of food and physical activity may have had upon the children either consciously or subconsciously altering their normal school-day food and activity habits.

The additive effect of school-day variation in height and weight was reflected in the increase of BMI observed in 81% of the sample. At the group level the effect size showed a large increase in BMI (median of the increase: +0.16 $kg.m^2$, r = 0.5) for the whole sample, for boys (median of the increase: +0.12 $kg.m^2$, r = 0.5) and for girls (median of the increase: +0.18 $kg.m^2$, r = 0.7). Therefore there was a systematic increase in BMI over the school day. The previously identified systematic group changes in actual BMI from the morning to afternoon measurements, at the individual level, resulted in one BMI category change when applying the clinical cut-offs. In one girl, an increase in BMI of 0.28 $kg.m^2$ was sufficient to increase her categorisation from just within the *healthy weight* category (measured at the 90th centile) in the morning to just within the *overweight* category (measured at 91.2 centile) in the afternoon.

When applying the population cut-offs, differential classification was magnified as three girls increased their BMI categorisation, with an increase in BMI as small as 0.26 $kg.m^2$ sufficient to alter a girl's BMI category. It appears therefore that whilst no significant changes were observed in categorisation using either cut-offs, at the individual level girls are more susceptible to differential classification when categorised using the more widely used population monitoring cut-offs. Despite this observation the population monitoring cut-offs have been shown to have high sensitivity for identifying obese children (Reilly et al., 2000) and are commonly used in studies of PA behaviour in children and therefore more readily permit comparison with previous studies.

There was also a systematic increase in BMI percentile in the whole sample (median of the increase = 2.5 centiles) and whilst children in the *healthy weight* groups as categorised using both BMI cut-offs were significantly more susceptible to increasing BMI percentile, compared to the *overweight* and *obese* groups respectively, this is most likely due to the presence of a greater range of available centiles within the *healthy weight* range compared to the *overweight* and *obese* categories rather than any category-dependent pattern of diurnal change in BMI. That is the centiles on the UK 90 BMI curve are particularly skewed so that the distance between the upper centiles is far greater than the lower centiles, in the order of four times as wide as the bottom centile channel at all ages (Cole et al., 1995).

The findings in the current study are all resultant from measurements performed by one trained researcher, using the mean of repeat measures within protocol-defined accuracy limits (±0.4 cm, ±0.1 kg), and undertaken with the explicit intention of measuring the child accurately in a clearly defined time period within the school day. Arguably therefore the time of day in which the measurements are performed (morning or afternoon) should be standardised. Standardisation is of particular importance if a given study group has a greater number of *healthy weight* or *overweight* individuals (girls particularly) measured during either the morning or afternoon, as the proportion of *overweight* and *obese* individuals may differ between study groups, and importantly so may the characteristics (e.g. activity level, attitude towards physical activity) of these individuals. This would also cloud the 'true' intervention effect of a given weight status group. The occurrence of such a phenomenon could therefore bias intervention effects of a particular weight status group, and make comparisons of weight status moderation between multiple intervention groups unfair. Reliable categorisation of individuals into BMI-determined weight-status groups, as stated above, can however be maintained by simply standardising the time of day when measurements are taken. As there is no 'true' measure of height and weight due to diurnal fluctuation in these measurements, the choice of timing of measurement should logically be determined by logistical considerations.

One of the main limitations of this study is the relatively small sample size. In a larger sample, with a greater number of individuals at the extreme percentiles for the *healthy weight* and *overweight* category it is conceivable that significant changes in categorisation may be observed. Whilst the sample in the current study is limited in size, it did include all of the pupils from three primary schools all of whom had previously been included in collection of data for the National Child Measurement Programme (NCMP) (a nationwide obesity surveillance programme) in each school. As such, the sample represents the children measured by the NCMP in 2009/10 in this district and represents similar proportions in each of the BMI categories compared with the national data. The number of children classified into each BMICATPop category from the morning and afternoon measurements in the sample used in this study (Table 4.3) is comparable to recent nationwide NCMP data (NHS Information Centre, 2009). Further the sample size is also suitable for the analysis of reliability in anthropometric variables.

4.6 Conclusion

This is the first study to examine the effect of the additive interactions between school-day biological variation in height and weight upon BMI-determined weight category in children aged 10-11 years. The summative contribution of intra-observer measurement error and biological height loss from morning to afternoon has been shown to systematically increase BMI, and BMI percentile, in both boys and girls. In combination with a systematic increase in weight in the afternoon in girls, these data have illustrated that at the individual level, BMI category may potentially be classified differently and unfavourably in a small number of girls when measured in the afternoon compared to the morning. To ensure reliable categorisation of weight status the time of day when measurements of height and weight are conducted should be standardised.

5 Chapter Five - Study Two: Intra- and inter-instrument reliability of the Actiwatch 4 accelerometer in a mechanical laboratory setting

5.1 Introduction

Accelerometry has become an increasingly popular method to objectively assess PA in children (Rowlands, 2007). Evidence from studies employing accelerometers has been used to better identify relationships between PA and hypokinetic conditions such as obesity (Ness et al., 2007). This is in part due to the increased measurement accuracy and precision afforded through the use of accelerometers compared to subjective measures (Corder et al., 2008). Despite the widespread use of accelerometers and the increase in monitor technology, information on certain aspects of these devices is still limited (Esliger & Tremblay, 2006). That is the majority of research using accelerometers has focused upon the development of EE prediction equations and intensity cut-points. Consequently research on the reliability of some accelerometer models remains limited. This is despite the recommendations from expert panels that ascertaining unit reliability (i.e. unit calibration) should be a priority (Ward, Evenson, Vaughn, Rodgers, & Troiano, 2005; Welk, 2005).

Studies that have previously investigated accelerometer unit reliability can be divided into participant-mounted (either laboratory-based or free-living protocols) or laboratory-based mechanical reliability studies (Brage et al., 2003; Esliger & Tremblay, 2006; Fairweather et al., 1999; McClain et al., 2007). The mechanical laboratory experiments have used various apparatus to expose the units to a common stimulus. These include rotating wheels (Brage et al., 2003), turntables (Metcalf et al., 2002) and hydraulic shaker/vibration plates (Esliger & Tremblay, 2006; Powell et al., 2003). In comparison to human experiments mechanical devices have several advantages, such as the large number of accelerations that can be generated, the ability to record data from multiple instruments simultaneously and the high reproducibility of oscillations between trials (Esliger & Tremblay, 2006).

Fairweather et al. (1999) developed a mechanical shaker to determine the variability between four Actigraph accelerometers (Model 7164, Actigraph, Fort Walton Beach, FL), during a single-speed condition (2 Hz). Inter-instrument correlations between pairs of accelerometers tested on the mechanical shaker were high (r = 0.98-0.99, p < 0.01), with the percent mean difference between devices approximately 3%. However, this study was limited both by its small sample size and single test speed. Metcalf et al. (2002) developed a motorised turntable that captured the response of 23 Actigraph accelerometers (7164) undergoing sinusoidal oscillations at two speeds. It was found that the Actigraph demonstrated good intra-instrument reliability, displaying an average coefficient of variation (CV = SD/mean) of 1.4% at medium speed and <1% at high speed. These experiments demonstrate that the Actigraph (7164) shows good reliability both within and between monitors when tested over a small range of movement speeds.

Brage et al. (2003) devised a rotational wheel experiment to explore the intra- and inter-instrument variability of six Actigraph accelerometers (7164) subjected to 51 different accelerations (0.1 to 19.7 $m.s^2$), produced by altering the frequency and radius of oscillation. Low intra-instrument variability (CVintra = 4.4%) was found, but inter-instrument variability was large by comparison (CVinter = 5 - >100%), leading to the suggestion that between-device calibration factors should be derived. Similarly, Powell et al. (2003) evaluated the technical reliability of the RT3 triaxial accelerometer (Stayhealthy, Inc., Monrovia, CA) over a range of motion frequencies. Each RT3 unit was mounted on a motorised vibration table, which delivered vibrations at 2.1, 5.1 and 10 Hz. The ICC for the vector sum and the three individual axes was r = 0.99. The inter-instrument CV (CVinter) decreased as movement frequency increased, ranging from 21.9-26.7% at 2.1 Hz to 4.2-7.2% at 10.0 Hz.

Only one study has compared different models of accelerometer units. Esliger and Tremblay (2006) used a hydraulic shaker table to quantify the variability of the Actical, RT3 and Actigraph accelerometers using a hydraulic shaker table, on which 79 monitors were subjected to six differing conditions designed to modulate acceleration or frequency of movement independently. It was found that variability in the Actical was negatively related to the acceleration of the condition, whereas no relationship was found between acceleration and variability in the Actigraph, but instead with frequency of movement. In addition the RT3 was found to be highly variable (CVintra = 46%, mean CVinter = 43%), which was suggested in part to be due to its wide frequency range. The findings of this study show that reliability may not only vary between different models of accelerometer but also within the same model of accelerometer.

Accelerometer reliability studies have to date focused solely upon raw activity counts. However most published research investigating the relationship between accelerometer-measured PA and health outcomes is presented using derived activity variables, such as time in PA intensity categories e.g. minutes spent in MVPA (Andersen et al., 2006; Ness et al., 2007; Riddoch et al., 2009). Only a single study to date has investigated the reliability of derived activity variables from accelerometer output. McClain et al. (2007) examined the inter-instrument reliability of concurrently worn (left hip and right hip sites) Actigraph accelerometers (7164) during free-living conditions. They assessed both raw and derived variables and found that inter-instrument reliability of the Actigraph for classifying time spent in MVPA was acceptable (CV= 3.7%, APE= 4.9% and ICC r = 0.99), and better than when considering moderate activity *per se*. McClain et al. (2007) therefore concluded that MVPA may be the best derived PA intensity variable to use due to the reduced likelihood of count misclassification between the moderate and vigorous threshold as a consequence of using a composite variable; that is moderate + vigorous activity.

The Actiwatch 4 accelerometer (one of the only wrist-worn accelerometers currently available) has been validated against energy expenditure in children, with energy expenditure prediction equations and intensity cut-points also being derived (Puyau et al., 2002; Puyau et al., 2004). Despite the AW4's validation as an activity monitor there have been no published examinations explicitly examining either the intra- or inter-instrument reliability of the AW4, and therefore the reproducibility of this accelerometer-based PA monitor is unknown.

5.2 Study aim

The purpose of this study was therefore to quantify the intra- and inter-instrument reliability of the Actiwatch 4 when accelerated at speeds representative of moderate and vigorous intensity in a mechanical laboratory setting.

5.3 Methods

5.3.1 Instrumentation

5.3.2 Actiwatch 4

See Chapter Three for a detailed technical specification of the Actiwatch.

5.3.3 CSMi Isokinetic Dynamometer

All testing was completed using a CSMi Isokinetic Dynamometer (Computer Sports Medicine Inc., Stoughton, MA, US). The dynamometer was selected as it can produce constant motion at speeds ranging from 1 to 500 deg.s^{-1}, with a total range of motion of 360°.

5.3.4 Procedure

To select the test speeds that were representative of moderate and vigorous intensity, five Actiwatch 4 units were attached to the knee/hip arm adapter of an isokinetic dynamometer and were accelerated at six different test speeds (50, 100, 150, 200, 250, 350 deg.s$^{-1)}$. The units were set to record at 10-s epochs, as this is the maximum resolution for 7-day data capture in the AW4. The mean of the five units was compared to published 1-min intensity cut-points (Puyau et al., 2004), which were divided by six to provide a moderate intensity threshold of 117-416 cts.10 s^{-1}, and a vigorous intensity threshold of ≥417 cts.10 s^{-1} for the 10-s epoch data captured. The test speed of 50 deg.s^{-1} produced ~300 cts.10 s^{-1} and was therefore selected as the MPA representative Condition A. The test speed of 200 deg.s^{-1} produced ~600 cts.10 s^{-1} and was therefore selected as the VPA representative Condition B. Twenty-seven Actiwatch accelerometers were intialised to collect data using 10-s epochs. Five accelerometer units at a time were mounted to the knee/hip adapter of the isokinetic dynamometer. They were positioned perpendicular to the floor, maximising time spent in the vertical axis. The dynamometer was set to move through a 90° range of motion, and each unit was accelerated for 30 min at 50 deg.s^{-1} (Condition A), and 30 min at 200 deg.s^{-1} (Condition B). An identical repeat trial was conducted in each condition (Trial 1, Trial 2). All study procedures were approved by the Ethical Advisory Committee of the Institute of Sport and Exercise Science, University of Worcester, UK.

5.3.5 Data treatment and statistical analyses

Data were first imported into Microsoft Excel and the recorded condition start and end times were identified. The first and last minute of each unit's data were deleted, to ensure that no spurious results were included in the dataset, leaving raw data for 28 min per condition (Esliger & Tremblay, 2006). The data were imported into SPSS for Windows Version 17.0 (SPSS Inc., Chicago, IL) for further analysis. Mean activity counts (cts.10 s^{-1}), and derived variables of time spent in MPA for condition A, and time in VPA for condition B, and time spent in MVPA in both conditions were calculated from the raw data in each accelerometer for Condition A, Trial 1: (ConA_Tr1), Condition A, Trial 2 (ConA_Tr2), Condition B, Trial 1 (ConB_Tr1) and Condition B, Trial 2 (ConB_Tr2).

5.3.6 Intra-instrument reliability

To explore the reliability within accelerometers five methods were used: (a) the standard deviation (SD) between trials; (b) the coefficient of variation (CVintra) for each condition between trials calculated by dividing the SD of the individual unit mean (between Trials 1 and 2) by the individual unit mean (Trial 1 mean + Trial 2 mean/2) multiplied by 100 [(SD/mean) x 100)] as per Esliger and Tremblay (2006); (c) the absolute percent error (APEintra) (without consideration of sign) calculated by subtracting the individual unit mean for Trial 2 (Trial 2 mean - Trial 1 mean) from Trial 1, the product of which was divided by the overall trial mean (Trial 1 mean + Trial 2 mean/2) multiplied by 100 [(Trial 2-Trial 1)/Trial means x 100]; (d) by paired samples *t*-tests on the differences in unit means between trials to determine systematic bias; and (e) by Pearson's correlation coefficients. The alpha level was set at $p < 0.05$ for all tests. If a difference was found Cohen's *d* was calculated (small = 0.2, medium = 0.5, large = 0.8, Cohen, 1988) as an estimate of effect size.

5.3.7 Inter-instrument reliability

Reliability between accelerometers was examined using the CVinter and APEinter. CVinter was calculated by dividing the SD between individual unit means (Trial 1 mean + Trial 2 mean/2) by the overall group mean (Trial 1 group mean + Trial 2 group mean/2) multiplied by 100 [SD/mean x 100]. APEinter was calculated by subtracting the individual unit mean (Trial 1 mean + Trial 2 mean/2) from the overall group mean (Trial 1 group mean + Trial 2 group mean/2), the product of which was divided by the overall group mean multiplied by 100 [(individual - group)/group x 100].

5.4 Results

Descriptive data for mean activity counts and reliability statistics for both conditions and trials are presented in Table 5.1. Descriptive data for mean time spent in PA intensity categories and reliability statistics for both conditions and trials are presented in Table 5.2.

From Table 5.1 the mean CVintra for mean activity counts was 4.6% for Condition A and 3.9% for Condition B, the combined mean being CVintra = 4.3%. Mean activity counts per epoch were greater in ConA_Tr1 compared to ConA_Tr2 (mean ± SD: 329 ± 24 vs. 310 ± 21 cts.10 s^{-1}, $t(26) = 5.2$, $p = 0.01$, $d = 0.8$), and were greater in ConB_Tr1 compared to ConB_Tr2 (mean ± SD: 621 ± 44 vs. 602 ± 38 cts.10 s^{-1}, $t(26) = 2.3$, $p = 0.03$, $d = 0.5$). The APEintra was 6.6% for Condition A and 5.6% for Condition B, the combined mean being APEintra = 6.1%.

From Table 5.2, the mean CVintra for time in MPA was 3.2% for Condition A, and for time in VPA in Condition B was 0.0%. The CVintra for time in MVPA in both conditions was 0.0%. Time spent in MPA was greater in ConA_Tr2 compared to ConA_Tr1 (27.6 ± 0.7 vs. 26.5 ± 1.8 min, $t(26) = -4.0$, $p = 0.01$, $d = 0.7$). There was no difference in time spent in VPA between ConB_Tr1 and ConB_Tr2 (28.0 ± 0.0 vs. 28.0 ± 0.0 min). No differences were found in time spent in MVPA between ConA_Tr1 and ConA_Tr2 or between ConB_Tr1 and ConB_Tr2. The APEintra was 4.5% for Condition A and 0.0% for Condition B, the combined mean being APEintra = 2.2%. The correlation for mean activity counts and minutes of MPA between ConA_Tr1 and ConA_Tr2 was $r = 0.67$ for both ($p = 0.01$ for both). The correlation for mean activity counts between ConB_Tr1 and ConB_Tr2 was $r = 0.51$ ($p = 0.01$).

Table 5.1 Mean values of raw Actiwatch output, and intra- and inter-instrument reliability statistics.

Conditions			Intra-Instrument			Inter-Instrument		
Trial	Condition	Counts	SD	CV	APE	SD	CV	APE
1	A	329*	14.8	4.6	6.6	20.4	6.4	5.2
2	A	310*						
1	B	621†	24.3	3.9	5.6	35.9	5.9	4.7
2	B	602†						
Overall mean		**465**	**19.6**	**4.3**	**6.1**	**28.2**	**6.1**	**5.0**

*Difference between trials (p = 0.01). †Difference between trials (p = 0.03). SD = standard deviation, CV = coefficient of variation, APE = absolute percent error.

Table 5.2 Mean values of derived PA category variables, and intra- and inter-instrument reliability statistics.

Conditions				Intra-Instrument			Inter-Instrument		
Trial	Condition	Intensity	Minutes	SD	CV	APE	SD	CV	APE
1	A	MOD	26.5*	0.8	3.2	4.5	1.2	4.3	3.2
2	A	MOD	27.6*						
1	A	MVPA	28.0	0.0	0.0	0.0	0.0	0.0	0.0
2	A	MVPA	28.0						
1	B	VIG	28.0	0.0	0.0	0.0	0.0	0.0	0.0
2	B	VIG	28.0						
1	B	MVPA	28.0	0.0	0.0	0.0	0.0	0.0	0.0
2	B	MVPA	28.0						

*Difference between trials (p = 0.01). SD = standard deviation, CV = coefficient of variation, APE = absolute percent error, MOD = moderate intensity, MVPA = moderate to vigorous intensity, VIG = vigorous intensity.

From Table 5.1 the mean CVinter for mean activity counts was 6.4% for Condition A and 5.9% for Condition B, the combined mean being CVinter = 6.1%. The mean APEinter for mean activity counts was 5.2% for Condition A and 4.7% for Condition B, the overall mean being 5.0%. From Table 5.2, the mean CVinter for time spent in MPA was 4.3% in Condition A, and for time spent in VPA in Condition B was 0.0%. The mean CVinter for MVPA in both conditions was 0.0%. APE for time spent in MPA was 3.2% for Condition A and for time spent in VPA in Condition B was 0.0%. The APE for MVPA in both conditions was 0.0%.

5.5 Discussion

This is the first study to fully evaluate the reliability of the Actiwatch 4 and only the second study to examine the reliability of derived activity variables in an accelerometer. These data demonstrate the AW4 to have acceptable intra-instrument reliability for raw mean activity counts according to the CV values observed. In Condition A the CVintra was 4.6%, which is of similar magnitude to the 4.1% observed in the Actigraph 7164 (Esliger & Tremblay, 2006), the 1.4% observed in the CSA (now known as Actigraph 7164) (Metcalf et al., 2002) and the 1.8% observed in the RT3 (Krasnoff et al., 2008). The lower CV values reported in previous studies may in part be due to smaller unit sample sizes and the use of a single production batch of units. The units used in this study were re-conditioned from various clinical trials which may account for the heterogeneity in between-trial variance within this sample. A number of units displayed CVintra of >5%, potentially inflating the mean CV values. In addition, past studies have used a 1-min epoch. Silva, Mota, Esliger, and Welk (2010) found improved reliability when integrating data from 15-s to 1-min epochs, suggesting that the higher CV values observed in this experiment may also in part be related to the sampling rate used.

A significant difference was found between mean activity counts between Trial 1 and Trial 2, with systematically lower activity counts produced in Trial 2 in both conditions. This may reflect both systematic unit bias and random trial-related error such as AW4 battery discharge, as well as resonance in the experimental setup between trials. Despite the significant difference between trials, there was only a difference of approximately 20 counts between trials over a 28-min test condition, which on average reflects approximately 5% of the combined mean of all trials, and is therefore not deemed practically 'meaningful'.

In Condition B the intra-instrument variance was reduced (CVintra = 3.9%), the implication being that raw counts show some variance within units, becoming less variable as the test speed increased. The intra-instrument reliability of raw activity data was greater than inter-instrument reliability, which is consistent with the findings of prior studies (Esliger & Tremblay, 2006; Krasnoff et al., 2008; Powell et al., 2003). The CVinter was observed as 6.4% in Condition A and 5.9% in Condition B, again higher in Condition A. Krasnoff et al. (2008) found fairly high CVinter (9.5-34.7%) among RT3 accelerometer units oscillated on a hydraulic shaker table. Similarly Esliger & Tremblay (2006) observed a mean CVinter of 8.6% between Actigraph units that were accelerated at varying speeds; therefore the findings of the present study in the Actiwatch 4 are aligned with previously published parameters of inter-instrument variability in other accelerometer models. Unfortunately direct comparisons cannot be drawn with past reliability studies due to differences in the study protocol and accelerometer model used.

The Pearson's correlation reported between trials ranged from 0.51 to 0.67. No previous accelerometer reliability studies have reported these statistics, instead using the intra-class correlation coefficient (ICC). However, the Pearson's correlation values can be used to inform the power analyses calculation in Study Five of this thesis. The coefficients reported are moderate to strong in size, and as with the aforementioned CV values, may be reduced in magnitude due to a number of units with high within-trial variance. Interestingly the intra- and inter-instrument variability in raw and derived variables was found to be greater in the moderate intensity condition (Condition A) compared to the vigorous intensity condition (Condition B). This is congruent with data from previous studies showing an inverse relationship between test speeds (intensity of work) and variability in raw activity counts produced by the Actical (Esliger & Tremblay, 2006), and between frequency/acceleration and variability in raw activity counts both within and between RT3 devices (Powell et al., 2003). A previous study has shown that variance in accelerometer output (within units) is greatest at the lower and upper limits of acceleration (reliability is decreased at speeds designed to mimic very slow movement and fast running for example) (Brage et al., 2003). In a participant-mounted study using the RT3 device in adolescents, Vanhelst, Theunynck, Gottrand, and Beghin (2010) reported the highest CVintra during sedentary activities (17.3 ± 2.3%), with the lowest observed during vigorous intensity activities (6.6 ± 0.1%). Data from the present study show that the magnitude of error in the AW4 differed between test speeds, such that measurement error in the AW4 may depend upon the magnitude of the acceleration measured.

As a calibration check Esliger & Tremblay (2006) suggest an example a-priori calibration variability limit of an APEinter of <5% may be set for the selection of reliable units. If this had been applied to the units used in the current study, 14 of the 27 (52%) units would have been rejected as unreliable from the outset. Individually, these 14 units displayed bias (both under and overestimation) in mean activity counts, when compared to the mean value of the entire sample. However by assessing inter-instrument reliability between the separate derived variables of time spent in PA intensity categories as opposed to using mean activity counts, the discrepancy between the individual units and the sample mean expressed as APE was reduced (mean APE raw vs. derived (mean of MOD and MVPA)) Con A: 5.2% vs. 1.6%, Con B: 4.7% vs. 0.0%) to under the suggested 5% inclusion threshold. Further, when examining the intra-instrument reliability, CVintra reduced from 4.6% for raw variables to 1.6% for derived variables in Condition A, and from 3.9% to 0% in Condition B. Therefore when applying the 5% reliability threshold to the separate intensity-derived variables in both Condition A and B, all 27 units were acceptable for research use.

As noted previously there was a significant difference in activity count output between trials in Condition A, resulting in a difference of 1.1 (decimal) min of MPA. Whilst significant in this mechanical laboratory setting, in the field the clinical significance of this bias may be small. Further, 1.5 (decimal) min of MPA were misclassified as VPA in ConA_Tr1, with 0.4 min misclassified as VPA in ConA_Tr2. Similarly the APE between instruments was 3.2% for minutes of MPA alone, for both trials combined. Collapsing data from the separate moderate- and vigorous-intensity categories resulted in improved intra- and inter-instrument reliability in both conditions (CV = 0.0%, APE = 0.0%) in agreement with the findings of McClain et al. (2007). This suggests that the use of separate intensity categories may result in the misclassification of activity counts, as counts are placed into discrete categories i.e. moderate or vigorous activity (McClain et al., 2007).

On this basis a pragmatic applied research decision should be made. Whilst it is clear that there are discrepancies in mean activity counts, both intra- and inter-unit reliability is clearly improved by using derived variables. Those units that may have been excluded on the basis of an APEinter >5% in mean activity counts (Esliger & Tremblay, 2006), were deemed acceptable for the purposes of research when running a calibration check using derived activity data. The reporting of MVPA outcomes by accelerometer users has increased in regularity. This is largely attributable to the popularity of lifestyle-embedded PA interventions and focus upon participation in at least moderate intensity PA as a means for meeting current PA guidelines (O'Donovan et al., 2010). While the choice of intensity category to report in future studies will be dictated by the precise research question (McClain et al., 2007) the current study would suggest that in future, AW4 users should present data as combined MVPA to ensure optimal data reliability, if separate intensity categories are not required.

The observed variation between AW4 units suggests that when employing these devices longitudinally participants should wear the same device to ensure that artificial differences between time-points do not manifest. On this note it is important that researchers should test the precision of all wearable motion sensors prior to deployment. Where possible this should be conducted using a mechanical device which can replicate test speeds that are physiologically relevant, ensuring that identification of (in)variance can solely be attributed to the monitor, and not to within-subject biological variation (i.e. gait biomechanics, monitor positioning associated with body composition and clothing.

The primary limitation of this study is the restricted range of test speeds employed. In retrospect testing the AW4 performance across a greater range of speeds would have given a fuller picture of the AW4 performance, in particular upon misclassification between the moderate and vigorous thresholds. The optimal test speed would be closer to the lower end of the vigorous category, and closer to the upper end of the moderate intensity category. Similarly the potential for misclassification between the low intensity and moderate intensity category is important, however test speeds lower than 50 $deg.s^{-1}$ (moderate representative speed) could not be generated using the current testing equipment. Such test speeds could be achieved by the use of a mechanical shaker, which allows the production of a wide range of accelerations by manipulating the frequency or displacement amplitude of vibrations (Esliger & Tremblay, 2006). The strengths of this study are the use of a mechanical device to expose the accelerometers to a common movement stimulus, and therefore variability solely attributable to the AW4 has been identified. Future accelerometer users should perform unit calibration (test reliability) on their devices before deployment in the field using mechanically produced movement, so that inherent variance components such as placement site, accelerometer tilt angle and fidelity to a laboratory protocol are excluded.

5.6 Conclusion

In conclusion, in a mechanical laboratory setting both intra- and inter-instrument reliability of raw activity counts were acceptable, with greatest variance observed in the moderate representative condition. When derived variables of time spent in MPA, VPA and MVPA were used, greater reliability was observed in both conditions. It is apparent that the AW4 can reliably categorise raw activity counts into the health-enhancing intensity category of MVPA when accelerated at speeds producing a count output of at least moderate intensity in a mechanical laboratory setting. Therefore, dependent upon the research question (and if separate intensity categories are not of interest) future research using the AW4 should report the combined category of time spent in MVPA as opposed to separate categories of moderate and vigorous activity to increase data reliability.

6 Chapter Six - Study Three: The impact of placement site and body side dominance upon accelerometer-measured physical activity in 10-11-year-old children

6.1 Introduction

Accelerometry has become the most frequently employed method for the objective assessment of PA in children (Trost, 2007). Accelerometers measure and record acceleration produced by displacement of the body segment the device is attached to. On the basis that the magnitude of acceleration is proportional to the muscular force required to overcome inertia, the energetic cost of body movement can be estimated. Accelerometer validation studies have therefore used protocols including a combination of treadmill-based locomotor activities and lifestyle-based activities to determine the relationship between accelerometer output and measured EE (Eston et al., 1998; Puyau et al., 2002; Puyau et al., 2004).

For validation, researchers typically place accelerometer devices on the hip, close to the centre of mass (Freedson et al., 2005). The site of accelerometer attachment is important when measuring body motion as depending on the behaviour engaged in, body sites are differentially active, and movements of one body segment may not be correlated with movements of another (Westerterp, 1999). No prior research, in children, has examined other potential body placement sites during validation such as the wrist, upper arm, or ankle, or indeed compared the output between the hip and any of these sites in the field. Therefore the suitability of hip-derived validation coefficients and intensity cut-points, for use at these alternative placement sites is unknown.

Despite the aforementioned paucity in research, there has been some interest in the interchangeability of placement sites, which has led to comparison of the output between contralateral hip placements, and between hip and lower back placements within the field-based environment. Nilsson et al. (2002) investigated the effect of placement of the Actigraph accelerometer (model 7164) on the right hip and lower back in 16 7-year-old children over a 4-day period (age: 7.5 ± 0.3 years; height: 128 ± 4.5 cm; weight: 28.2 ± 5.9 kg). Results showed no differences in mean counts per minute between the two placement sites (hip = 751 ± 100 $cts.min^{-1}$ vs. back = 729 ± 112 $cts.min^{-1}$, $p = 0.20$). However when counts were classified into moderate or vigorous activity, hip data resulted in higher levels of moderate activity (hip = 320 ± 51.8 min vs. back = 294 ± 45.4 min, $p < 0.01$), but no differences were seen for time spent in vigorous or very vigorous activity (Nilsson et al., 2002). Notably the difference in moderate activity was only detected when 5-s epochs were used, suggesting that short sampling periods are needed to detect placement site differences.

Similarly Yngve et al. (2003) compared Actigraph placement (model 7164) on the lower back and the right hip in 28 healthy adults (age: 23.4 ± 2.6 years; height: 174 ± 4.5 cm; weight: 68.7 ± 4.6 kg). Activity counts from the hip placement were higher during normal walking (2822 vs. 2578 $cts.min^{-1}$) and fast walking (4926 vs. 4594 $cts.min^{-1}$), but lower during jogging (10,069 vs. 10,548 $cts.min^{-1}$) ($p < 0.01$). As with the results of Nilsson et al. (2002) the agreement between placement sites was greater during higher intensity movement (jogging compared to normal walking and fast walking), suggesting that positional differences are perhaps most apparent during moderate intensity movement. However in a free-living sub-study of 34 adults (age: 43.9 ± 10.6 years; height: 175.5 ± 7 cm; weight: 72.7 ± 10.1 kg) over 7 days no differences were observed for total activity counts, signifying that placement effects observed under laboratory conditions may diminish during extended free living (Yngve et al., 2003).

Some studies have investigated contralateral hip placements by comparing the accelerometer output between the left and right hip during free living in children and adults (Cook & Lambert, 2009; Fairweather et al., 1999; McClain et al., 2007). Fairweather et al. (1999) found significant differences between the left and right hip (left: 629 ± 254 vs. right: 598 ± 225 $cts.min^{-1}$) in 10 pre-school children (Age: 3.7 ± 0.5 years) who wore the CSA activity monitor (first-generation Actigraph) over two free-living days. However Cook and Lambert (2009) did not observe any differences in count output between the left and right hip (right: 999 ± 215 vs. left: 1008 ± 177 $cts.min^{-1}$) in a convenience sample of seven adults (Age: 29-46 years). Interestingly Cook and Lambert (2009) examined individual variance components and showed that monitor position did not contribute to the variability in output between hip placements, whereas the variance due to inter-instrument differences was 1.7% for total activity counts.

McClain et al. (2007) found no differences in raw accelerometer variables between hip placements, but observed differences in MVPA (right: 75 ± 29 vs. left: 73 ± 29 min, p = 0.037) in 10 adults (30.1 ± 3.8 years). The greatest percent error between placement sites was found for moderate intensity (18.8 %), which was reduced when collapsed into the composite of MVPA (4.9%), suggesting misclassification between placement sites and monitors occurs between the moderate and vigorous range as opposed to between moderate and light activity. These findings further highlight the issue of placement differences in the moderate intensity range. It appears small differences in accelerometer output may manifest between trunk placements during moderate activities performed in the laboratory and in the field (Nilsson et al., 2002; Yngve et al., 2003). This is not unequivocal however as differences in MVPA have been found to be small (<5%) (McClain et al., 2007) and may diminish over extended monitoring periods (Yngve et al., 2003). A possible variance component that none of these studies have considered is the effect of body side dominance. By comparing the left versus right hip these studies have negated the issue of side dominance; if data were re-categorised as dominant versus non-dominant body side, it is conceivable that differences in accelerometer output may be apparent (Fairweather et al., 1999).

Few studies have explicitly examined accelerometer output between the trunk and upper limb body segments, and to the best of the author's knowledge none have been conducted in children. Kumahara, Tanaka, and Schutz (2004) examined the relationship between upper limb movements and whole body trunk movements using the Lifecorder accelerometer in adults during 24 hr in a respiratory chamber. During this period participants completed two walking sessions on a treadmill, as well as a bout of self-paced walking, and recorded remaining unrestricted activities (bar 8 hr controlled sleep time) on a time-table at a frequency of 1 min. The accelerometer score of the waist was found to be slightly greater than the wrist during treadmill walking (ratio of wrist/waist: 0.8, range 0.7 to 1.1), but lower during self-care tasks and sedentary activities (2.9, range 2.6 to 3.1) ($p < 0.001$). Whilst this was a uniaxial device it is unlikely the limitations of uniaxial accelerometry contributed to these findings, as a recent calibration study using the tri-axial Genea accelerometer (Esliger et al., 2011) showed similar concordance between the wrist and hip output during ambulatory activities.

There is notably a lack of evidence concerning wrist and hip comparisons during free living. One of the only accelerometers designed specifically to be wrist -worn is the Actiwatch 4 (AW4) (Cambridge Neurotechnology, 2007). Currently no studies have examined differences in AW4 output between wrist and hip placements in children. As a wrist placement is intuitively less obtrusive than the waist, placement at the wrist may enhance participant compliance and allow longer monitoring periods to be used (Esliger et al., 2011). Indeed a recent study in adult men and women found acceptability of wear scores (1-10) to be significantly higher at the wrist (compared to hip) in men (+0.51), but not women (+0.43) following 7 days of wearing a single tri-axial accelerometer at the hip and at the wrist (van Hees et al., 2011). In addition the percent of wear time (of total possible) was very high for the wrist placement (Mdn: 99%), suggesting that participants find a 24-hr monitoring protocol acceptable when using a wrist placement (van Hees et al., 2011). In addition, as mentioned previously, the issue of side dominance is under-researched, with prior studies only comparing 'left vs. right' body side placements as opposed to 'dominant vs. non-dominant' body sides. The manufacturer's instructions (Cambridge Neurotechnology, 2008) stipulate that the AW4 should be worn on the non-dominant wrist, although no empirical evidence is available to support this requirement.

For the wrist placement to be used in the ensuing studies of this thesis, characterisation of the difference, and relationship between hip and wrist measured

AW4 output is required to preclude the necessity for full-value calibration at the wrist in the laboratory. By the application of published AW intensity cut-points (Puyau et al., 2004) to data derived from hip- and wrist-placement sites it is hypothesised that systematic differences in output will exist, and therefore wrist-measured intensity data (i.e. mins of moderate-to-vigorous activity (MVPA)) may be used to predict hip-measured minutes of MVPA, which is arguably the most important intensity category to assess in relation to children's health (Janssen & Leblanc, 2010). It is acknowledged that the original cut-points (Puyau et al., 2004) were developed with a degree of error, and the prediction of hip-measured MVPA from wrist data may contribute additional error. However if accurate at an individual level this would provide an efficient field-based alternative to full laboratory-based value calibration. Similar '*doubly indirect*' regression approaches have been proposed to equate accelerometer estimates of MVPA between different sets of MVPA cut-points in children (Bornstein et al., 2011).

6.2 Study aim

This study therefore aimed to investigate whether AW4-measured PA counts and minutes of moderate-to-vigorous PA (MVPA) differed between the hip and wrist, worn on the dominant and non-dominant sides in 10-11-year-old children and to investigate the predictive ability of wrist-measured minutes of MVPA and demographic variables against hip-measured minutes of MVPA.

6.3 Methods

6.3.1 Participants

Twenty-four Year 6 children (boys: *n* = 15; height: 141.9 ± 8.1 cm; weight: 37.8 ± 9.5 kg; BMI: 18.6 ± 3.3 kg.m^2; girls: *n* = 9; height: 141.6 ± 6.5 cm; weight: 36.9 ± 7.6 kg; BMI: 18.4 ± 3.5 kg.m^2) aged 11.2 ± 0.5 years were recruited from three primary schools in the West Midlands region of England. The experimental protocol received institutional ethics committee approval and written parental consent and child assent was obtained.

6.3.2 Instrumentation

See Chapter Three for a detailed description of the AW4 technical specification.

6.3.3 Study design

All data were collected across three separate school days within the hours of 09:30 and 15:00, with standing height (cm) and body mass (kg) measured between 09:30 and 10:00 to minimise diurnal variation in BMI-determined weight category (as per the findings of Study One). AW4 units were worn at each placement site between ~09:30 and ~14:30, on each school day.

6.3.4 Procedure

All anthropometric measurements were taken according to the protocols outlined in Chapter Three. Height to the nearest 0.1 cm was measured using a freestanding portable stadiometer (Seca 214, Seca ltd, Leicester, UK). Weight to the nearest 0.1 kg was measured using electronic weighing scales (HD 352, Tanita Corporation, Tokyo, Japan). Body mass index (kg.m^2) was then calculated by dividing mean weight in kg by the square of mean height in metres. Weight status was categorised according to BMI cut-offs from the 1990 UK growth reference dataset (Cole et al., 1995): *Underweight* = BMI ≤2nd centile; *Healthy weight* = >2nd centile to <85th centile; *Overweight* = ≥85th centile to <95th centile; *Obese* = ≥95th centile. Participants were fitted with a single AW4 (Cambridge Neurotechnology Ltd., Cambridge, UK) superior to the iliac crest on both hips using an elasticated waistband and a single AW4 on both wrists using a wrist strap proximal to the head of the ulna. The AW4s were set to record at 10-s epochs to ensure that high-intensity activity would not be underestimated by time smoothing (Edwardson & Gorely, 2010). Side dominance was assessed through a single item asking about writing and kicking preference. Placement sites were therefore categorised as dominant hip (DH), non-dominant hip (NDH) and dominant wrist (DW) and non-dominant wrist (NDW).

6.3.5 Data treatment and statistical analyses

To determine time spent in moderate-to-vigorous activity, activity counts were compared to 1-min intensity cut-points (Puyau et al., 2004), which were divided by six to provide an epoch-adjusted MVPA threshold of ≥117 cts.10 s^{-1}. The outcomes were therefore total activity counts (TC) (cts.4 hr^{-1}), mean activity counts (MC) (cts.10 s^{-1}), and total time spent in MVPA (min.4 hr^{-1}) during the measurement period.

Data were first imported into Microsoft Excel and the recorded condition start and end times were identified. The first and last 30 min of each unit's data were deleted as AW4 fitting and removal times were staggered across participants, leaving raw data for 4 hr (10:00-14:00). The data were imported into SPSS for Windows Version 17.0 (SPSS Inc., Chicago, IL) for further analysis**.** Normality of all output variables were assessed using the Shapiro-Wilks test, none of the variables violated the normality assumption. Paired samples *t*-tests were used to determine differences in output variables between the four placement sites. The between-epoch coefficient of variation (CV) was calculated for each placement site as a measure of variance (SD/mean*100). Pearson's correlations were used to determine the relationship between PA variables derived from the wrist and hip. Stepwise multiple linear regression analysis was used to predict minutes of hip-derived MVPA from non-dominant wrist MVPA (NDW_MVPA), gender, and BMI. Based on changes in *R* and a decrease in SEE, variables were accepted if they made a significant contribution to the predictive model. Collinearity values indicated multicollinearity was not an issue. The variance inflation factor (VIF) values were all ~1 and tolerance values were all >0.2 (Field, 2009). A paired *t*-test was used to determine differences between predicted non-dominant hip MVPA (PredNDH_MVPA) and measured minutes of hip-derived MVPA (NDH_MVPA). In addition Bland-Altman plots (Bland & Altman, 1986) were used to determine agreement between PredNDH_MVPA and NDH_MVPA. The alpha level was set at p <0.05 for all tests.

6.4 Results

6.4.1 Participant characteristics

Table 6.1 Participant characteristics.

Variable	**Boys** *n* **= 15**	**Girls** *n* **= 9**	**Total** *n* **= 24**
Age (years)	11.2 ± 0.5	11.2 ± 0.5	11.2 ± 0.5
Height (cm)	141.9 ± 8.1	141.6 ± 6.5	141.8 ± 7.4
Weight (kg)	37.8 ± 9.5	36.9 ± 7.6	37.5 ± 8.7
BMI (kg.m^2)	18.6 ± 3.3	18.4 ± 3.5	18.5 ± 3.3
Underweight (%)	0 (0)	0(0)	0 (0)
Healthy weight (%)	9 (60)	7 (78)	16 (67)
Overweight (%)	2 (13)	0 (0)	2 (8)
Obese (%)	4 (27)	2 (22)	6 (25)

6.4.2 Differences and associations between placement sites

From Figure 6.1, TC were lower at the NDH compared to the NDW ($t(23) = -15.4$, $p = 0.01$, mean diff: -97,729.7 ± 30,985.5 cts.4 hr^{-1}, 95% CI: -110,813.7 to -84,645.6 cts.4 hr^{-1}) and lower at the DH compared to the dominant wrist DW ($t(23) = -12.0$, $p = 0.01$, mean diff: -104,374.9 ± 42,545.3 cts.4 hr^{-1}, 95% CI: -122,340.2 to -86, 409.6 cts.4 hr^{-1}). There were no differences in TC between the NDH and DH ($p = 0.76$) or between the NDW and DW ($p = 0.45$). MC were lower at the NDH compared to the NDW ($t(23) = -15.4$, $p = 0.01$, mean diff: -67.9 ± 21.6 cts.10 s^{-1}, 95% CI: -77.1 to -58.8 cts.10 s^{-1}) and lower at the DH compared to DW ($t(23) = -12.3$, $p = 0.01$, mean diff: -74.3 ± 29.5 cts.10 s^{-1}, 95% CI: -86.7 to -61.8 cts.10 s^{-1}). No differences were found in MC between the NDH and DH ($p = 0.75$) or NDW and DW ($p = 0.25$).

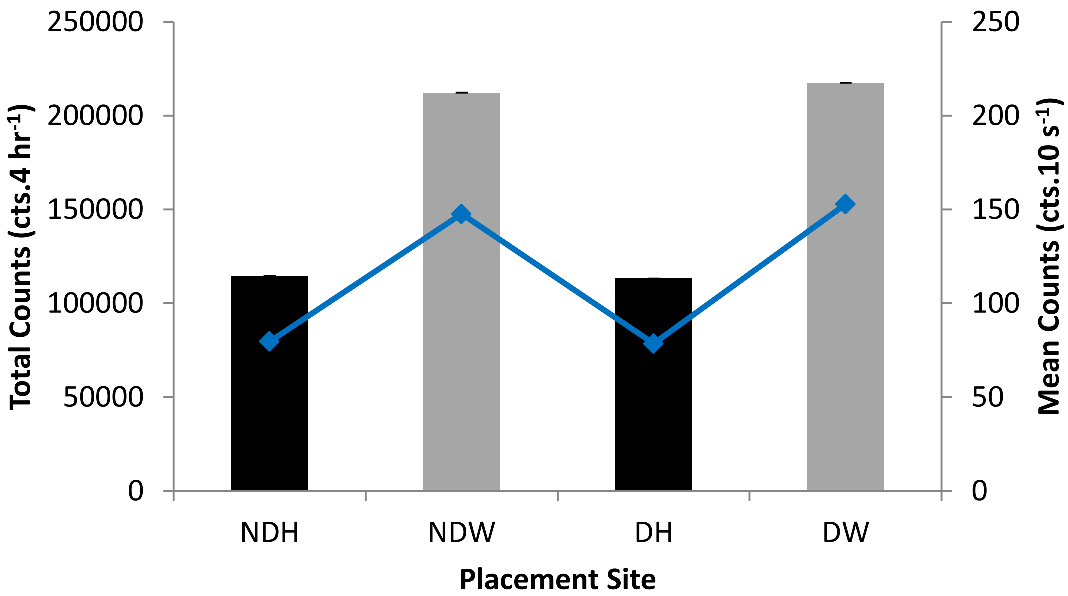


Figure 6.1 Total activity counts (bars) and mean activity counts by placement site (blue line). Black bars denote hip placements, grey bars denote wrist placements. Error bars (± 1SD) are not shown for ease of interpretation.

Minutes of MVPA were lower at the NDH compared to the NDW ($t(23) = -15.6$, $p = 0.01$, mean diff: -46.2 ± 14.4 min.4 hr^{-1}, 95% CI: -52.3 to -40.1 min) and lower at the DH compared to the DW ($t(23) = -12.6$, mean diff: -46.3 ± 17.9 min.4hr^{-1}, 95% CI: -53.8 to -38.7 min.4 hr^{-1}). No differences in MVPA were found between the NDH and DH ($p = 0.78$) or NDW and DW ($p = 0.86$).

TC from the NDH were related to those derived from the NDW, ($r = 0.80$, $p = 0.01$) and between the DH and DW ($r = 0.71$, $p = 0.01$). Similar relations were observed in MC between the NDH and NDW ($r = 0.80$, $p = 0.01$) and between the DH and DW ($r = 0.71$, $p = 0.01$).

6.4.3 Variance in AW4 output by placement site

From Table 6.2 variance as represented by the CV was observed to be greater at the NDH and DH compared to the NDW and DW.

Table 6.2 Mean coefficient of variation values by placement site.

Placement site	CV (%)	SD	Minimum	Maximum
NDH	179.8	30.0	119.7	232.6
DH	181.4	32.0	109.5	242.4
NDW	117.8	13.2	83.2	147.9
DW	120.0	22.3	85.5	208.5

6.4.4 Derivation of prediction equation

The relationship between NDH_MVPA and NDW_MVPA ($r = 0.79$, $p = 0.01$) was stronger than between DH_MVPA and DW_MVPA ($r = 0.75$, $p = 0.01$), therefore NDW_MVPA was used as the predictor. The equation developed from NDW_MVPA alone explained 62% of the variance in NDH_MVPA ($R^2 = 0.62$, SEE = 13.51); including gender accounted for an additional 16.0% ($R^2 = 0.77$, SEE = 10.55), including BMI accounted for a further 4.9% (R^2 = 0.83, SEE = 9.54). The final model to predict minutes of NDH_MVPA from minutes of NDW_MVPA was:

NDH_MVPA = 17.871 + 0.693 NDW_MVPA (min) - 17.652 gender (male = 0, female = 1) - 1.444 BMI (kg.m^2).

R^2 = 0.83 SEE = 9.54 min

There was no systematic bias between predicted time spent in MVPA (PredNDH_MVPA) and measured time spent in MVPA (NDH_MVPA) (Mean diff: 0.005 ± 8.90 min.4 hr^{-1}, 95% CI: -3.75 to 3.76 min.4 hr^{-1}, p = 0.99). The degree of agreement between the two values is illustrated by a Bland and Altman plot in Figure 6.2. The mean difference was 0.005 min (p = 0.98), and the 95% limits of agreement (±1.96 SD) were -17.44 and +17.45 min.4 hr^{-1}. The fit of the prediction equation was checked (normal probability plot of residuals and plot of the residuals against the predicted values) and deemed acceptable for group level estimates, but a large SEE was observed reducing accuracy at the individual level.

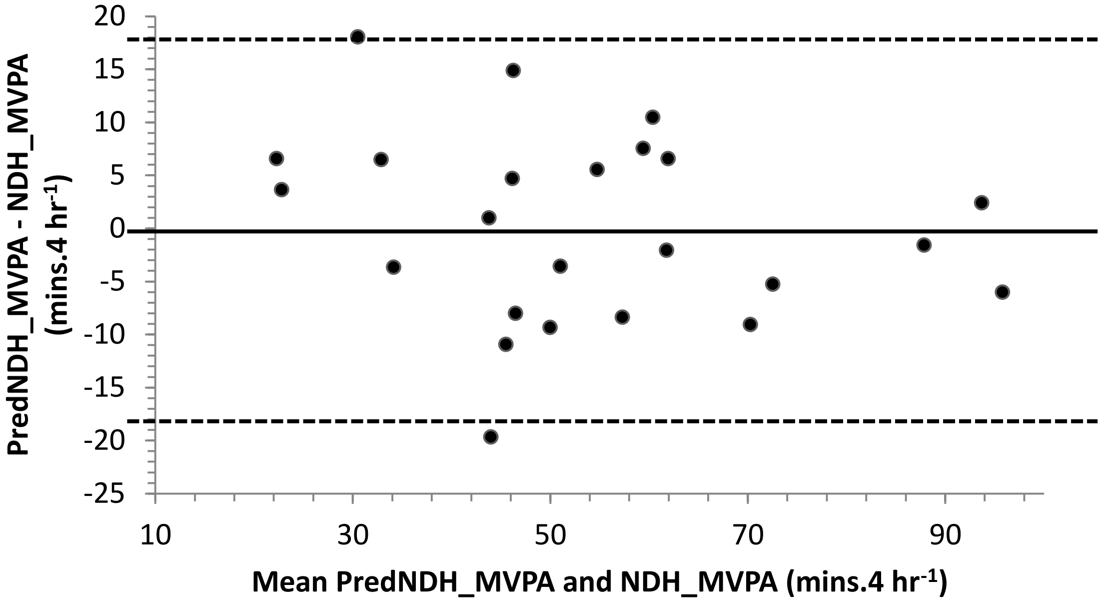


Figure 6.2 Bland-Altman plot showing mean difference (0.005 ± 8.9 min.4 hr^{-1}) and 95% limits of agreement (-17.44, +17.45 min.4 hr^{-1}) between minutes of PredNDH_MVPA and NDH_MVPA.

6.5 Discussion

This is the first study to investigate whether AW4-measured PA counts and time spent in MVPA differ between the hip and wrist, and between the dominant and non-dominant body side in children. The present data show that total activity counts, mean activity counts, and time spent in MVPA derived from hip-placed AW4 units were systematically lower than those derived from wrist-placed AW4 units, regardless of which side of the body they were placed. The difference in raw output variables between hip- and wrist-derived activity data was large, showing comparisons between data derived from these placement sites should not be made.

The variance in AW4 output between the hip and wrist is contributed to by numerous factors. Firstly there is likely a greater acceleration signal at the wrist during mixed static/dynamic movements such as skipping or playing catch. This assertion is supported by previous observations that the acceleration signal is stronger at upper limb sites compared to trunk placements during mixed static/dynamic lifestyle activities in adults (Kumahara et al., 2004). Further, when seated during class time which would be expected to involve little trunk movement, wrist movements would have been detected (Kumahara et al., 2004) as a consequence of fidgeting, or writing for example. This is reflected by the lower variability in activity counts observed between epochs (CVhip = ~180% vs. CVwrist = ~119%) at both wrist sites compared to hip sites i.e. fewer transitions from low to high counts and vice versa.

Further to this, there is the issue of sensor orientation, as accelerometer output is partly dependent upon the orientation of the piezoelectric sensor(s) (Welk, 2005). That is, if the device is not fully triaxial it should logically be aligned to the axis of movement it is designed to be most sensitive in. The AW4 is an omnidirectional device, which although affected by motion in all planes, is most sensitive to vertical movement. When placed at the wrist in the present study the piezoelectric sensor was oriented in the vertical plane (most sensitive axis), however when placed at the hip, the sensor was aligned in the anterior-posterior plane, likely resulting in an attenuated acceleration signal.

In contrast to the limb and trunk comparisons, no differences were apparent in mean activity counts or time spent in MVPA between the DH and NDH and between the DW and the NDW. These findings are congruent with previous data showing little or no variation between contralateral hip placements measured using the Actigraph in adults (Cook & Lambert, 2009; McClain et al., 2007). However these studies examined differences between the right vs. left hip, masking the issue of side dominance. At present no studies have investigated side dominance effects at the wrist or hip. According to the current results and converse to the manufacturer's instructions (Cambridge Neurotechnology, 2008), side dominance is not a monitoring issue in this population and thus the AW4 may be placed on either side of the body without risk of biasing data.

As systematic bias was present between wrist- and hip-derived data (but not between side dominance), a regression equation was created to predict NDH_ MVPA from

NDW_MVPA. The final model showed good predictability accounting for 83% of non-dominant hip-measured MVPA. The mean difference (0.005 min.4 hr^{-1}) and limits of agreement (-17.44, +17.45 min.4 hr^{-1}) between actual and predicted minutes of hip-measured MVPA, and the size of the SEE (9.54 min.4 hr^{-1}) suggest that the derived equation is suitable for a general group level prediction, but is imprecise at the individual level. Bland and Altman plots showed that for '*inactive*' children (<60 min of MVPA per day) the equation both over and underestimated NDH_MVPA, whilst for '*active*' children (≥60 min of MVPA per day) the equation generally underestimated. The observed large SEE could reduce the ability to detect between group differences in PredNDH_MVPA due to the introduction of additional variance to the data set, as statistical tests such as ANOVA calculate the test statistic based upon the ratio of systematic variation (experimental effect) to unsystematic variation (individual differences and random error) (Field, 2009).

Beside the poor performance of the derived regression model at the individual level, it was evident that hip-estimated MVPA was noticeably high over this monitoring period. Participants accumulated ~53 min of MVPA over 4 hr (10:00-14:00), which is elevated when considering that population-level studies of habitual daily MVPA in UK primary school children have ranged from 20 to 74 min. Whilst the monitoring period used in the present study included discretionary time such as recess and lunchtime, it did not include '*school related*' (i.e. 06:00-0900 and 15:00-16:00) or '*after school*' (i.e. 16:00-) periods which contribute to ~40-50% of weekday MVPA (Gidlow et al., 2008).

Table 6.3 Accelerometer-derived estimates of MVPA in UK primary school children.

Study	Sample size	Age of sample (years)	MVPA (min.day^{-1})	Meeting PA guidelines (%)
Present study†	*24*	*11.2 ± 0.5*	*53.8 ± 21.4*	*29.2*
Riddoch et al. (2007)	5595	11.8 (11.6-11.9)*	20 (12-31)*	2.5
van Sluijs et al. (2008)	2064	10.3 ± 0.3	74.1 ± 24.9	69.1
Owen et al. (2009)	2071	9.9 (9.2-10.7)*	66 (49-87)*	64

*Values are median and interquartile range (IQR).†MVPA (min.4 hr^{-1}) estimate from the NDH.

These relatively high levels of hip-measured MVPA may be reflective of a small 'high active' sample, however when compared with weekly estimates of MVPA from UK children of a similar age (Table 6.3) it is possible that this results from the application of published intensity cut-points which may not be accurate in this population. Whilst this cannot be 'proven' without laboratory 'cross-validation', contextualisation against data from similar UK populations (albeit using different accelerometers and cut-points) is at the least evidence rationalised. Disparity in MVPA data between studies is a popular issue in current accelerometer literature due to dissonance in the computation of MVPA in youth between studies, believed to be a result of the use of different intensity cut-points (known as the '*cut-point conundrum*', (Trost, 2007)) as opposed to a distinct behavioural difference (Bornstein et al., 2011). The heterogeneity in cut-points to define at least moderate intensity results from differing sample characteristics (i.e. age, body composition, cardiovascular fitness, biomechanical efficiency (Freedson et al., 2005)), the intensity and type of calibration source activities (Crouter et al., 2006a; Guinhouya et al., 2011) and data analysis techniques employed (Jago et al., 2007).

When calibrating accelerometers, developed cut-points/calibration equations are accurate only to the extent that the sample is representative of the population it will be used in (Welk, 2005). The cut-points created by Puyau et al. (2004) were derived on a sample of 32 children aged 7-18 years, of which it is not known how many were aged 10-11 years (age of present sample). Similarly, researchers should use a range of source activities characteristic of the population under study, in particular it is recommended that intermittent activities such as jumping/skipping are included alongside locomotor activities, as activity patterns are rarely steady state in children (Welk, 2005). Puyau et al. (2004) used walking (3.5 mph), aerobics and ball toss games, which may not be typical of the types of activities the children in the present sample engaged in.

Due to the large variance in individual level prediction and high SEE of the derived regression equation, and the relatively high MVPA estimates observed when using the Puyau et al. (2004) cut-points, AW4 data will be expressed as raw count data (i.e. total activity counts, mean activity counts) in the ensuing studies of this thesis. As strong positive correlations were observed in raw data between wrist and hip placements, it appears the wrist placement can provide a valid reflection of gross bodily movement during free living. As wrist-worn accelerometers are likely more participant friendly than hip worn (Esliger et al., 2011; van Hees et al., 2011), to increase participant compliance the AW4 will be placed on either the dominant or non-dominant wrist (at the participant's discretion) in the ensuing studies of this thesis. Future studies should calibrate wrist-placed AW4 against EE to derive PAEE prediction equations and activity intensity cut-points.

The particular strengths of this study are that all findings are resultant from AW4 units that have been tested for intra- and inter-instrument reliability in a mechanical laboratory setting, displaying both CVintra and CVinter of ~5% (see Study Two). The attachment of AW4 units were randomised and thus the differences seen in PA data between the hip and wrist placements can be attributed to kinematic and sensor orientation differences and not to artificial error intrinsic to the AW4 units themselves. In addition a short epoch (10 s) was used to capture the data, which ensured that any acute differences between placements were not masked by the time smoothing that could occur when using longer epochs.

The short monitoring period (4 hr) is a limitation of this study. However as this period included classroom time (capturing sedentary behaviour and light activity) and recess time (capturing moderate and vigorous activity), and data were registered in each intensity category, it is unlikely relative placement-site difference results from a longer monitoring period would differ. The sample used in this study is relatively small and homogenous, yet these are the first data to highlight the identified measurement issues using the Actiwatch, and is similar in sample size to other accelerometer measurement issue studies (Nilsson et al., 2002; Puyau et al., 2002). Future investigations into placement site differences should therefore use larger and more heterogeneous samples. Further, the validity of epoch-adjusted cut-points is unknown, however their use is supported in previous published accelerometer work (Edwardson & Gorely, 2010; McClain et al., 2008) and has been shown not to influence cut-point derivation in a recent value calibration study (Jimmy et al., 2012).

6.6 Conclusion

In conclusion, the placement of The AW4 accelerometer has a significant effect on raw activity counts and estimates of time spent in MVPA, therefore comparisons between wrist-measured and hip-measured PA data are not valid. Contrary to the AW4 manufacturer recommendations the AW4 may be worn on either side of the body, as no differences in output were apparent between contralateral hip and contralateral wrist placements. The preliminary attempt to regress wrist- against hip-monitored MVPA appeared suitable for group level prediction. However the inaccurate individual level prediction and high SEE observed may reduce the ability to detect between-group differences, and therefore it will not be used further in the ensuing studies of this thesis. Further, as hip-derived MVPA was noticeably high over a short monitoring period, the Puyau et al. (2004) cut-points will not be used to reduce raw count data into intensity minutes in the ensuing studies of this thesis. For more accurate individual estimates of activity intensity future studies should calibrate both wrist- and hip-worn AW4 to determine the performance of derived PAEE prediction equations and activity intensity cut-points. This is of importance as a wrist placement is likely to be more participant friendly, increasing compliance to monitoring protocols and as a consequence quality of data.

7 Chapter Seven - Study Four: Development and evaluation of the impact of individual- and group-standardised pedometer step-goal interventions upon accelerometer-measured physical activity in 10-11-year-old children

7.1 Chapter Structure

The following chapter is split into two sections, 'Phase One' which details the rationale for the development of two pedometer-based interventions, the intervention mapping process used to inform the development of these interventions, and 'Phase Two' which reports on a controlled experimental evaluation of these interventions.

7.2 Phase One Introduction

Regular participation in PA is important for children's health as it is associated with a reduction in risk factors for numerous hypokinetic diseases (Janssen & Leblanc, 2010; Strong et al., 2005). Despite the potential benefits of regular engagement in PA, reviews of self-report data show evidence for secular decreases in active transport, school physical education and organised sport (Dollman et al., 2005; Knuth & Hallal, 2009). In conjunction, accelerometer-derived data show many children do not attain recommended volumes of health-enhancing MVPA (Colley et al., 2011; Riddoch et al., 2007). Subsequently, identifying approaches that will be effective in increasing and sustaining PA in children is important.

Since the shift in focus from purposeful high-intensity exercise to moderate lifestyle-embedded PA in public health recommendations of the 1990s (Pate et al., 1995) researchers have sought methods to support PA behaviour change (Lutes & Steinbaugh, 2010). Pedometers, simple body-worn step counters, are becoming increasingly used as a motivational tool within PA interventions. Pedometers themselves are simply step counters; however they provide real-time feedback on PA volume of the end-user. This continual feedback permits the wearer the ability to self-regulate behaviour by monitoring and recording daily step totals and comparing this behavioural feedback against a reference step goal (e.g. 10,000 steps.day^{-1}).

Evidence from two systematic reviews of adult pedometer literature shows pedometer use can lead to moderate increases in PA over baseline in the order of 2-3000 steps.day^{-1} more than controls (Bravata et al., 2007; Kang et al., 2009a). In children less evidence exists for the effectiveness of pedometer-based interventions. There have been almost 20 published interventions in children and adolescents. Data from these studies show that pedometer use is associated with increases in PA of up to 3000 steps.day^{-1} (Berry et al., 2007; Hardman et al., 2011; Horne et al., 2009; Kang & Brinthaupt, 2009 Oliver et al., 2006).

The majority of pedometer-based studies in children have used one of, or a combination of, self-monitoring (recording daily step totals), pedometer feedback and step-goal setting alongside additional behavioural components (Berry et al., 2007; 2006; Hardman et al., 2011; Horne et al., 2009; Oliver et al., 2006;). The use of additional intervention components and poor reporting of pedometer-based strategies clouds judgement on the active intervention ingredients that drive behaviour change, limiting understanding of *how* pedometer deployment relates to effectiveness. Few studies have tested pedometer strategies alone in children. Butcher et al. (2007) found a combination of step-count feedback and activity information to increase steps per minute compared to 'no feedback and information' controls. Similarly, Kang and Brinthaupt (2009) found that a combination of step feedback, self-monitoring and goal setting was sufficient to increase daily steps by ~ 19% over baseline, following a 6-week intervention, however a control group was not included within this study. To date, no studies have compared goal setting, step-count feedback and self-monitoring using a step diary against a control group in children.

In addition, little research has delineated the relative importance of step-goal type. A meta-analysis of adult data (Kang et al., 2009a), found interventions administering pre-set (i.e. 10,000 steps.day^{-1}) goals to elicit a marginally higher positive effect size than individual-standardised goals (i.e. incremental increases above baseline steps). However Tudor-Locke and Lutes (2009) suggest that due to the limited research comparing goal types, the setting of any goal above baseline values for an individual to work towards may be sufficient to increase PA. The types of step goals that could be administered include self-set (user sets own goal, also known as tailored or personalised), individual-standardised (i.e. increment above individual baseline), group-standardised (i.e. increment above group baseline) or pre-set universally administered (i.e. 10,000 steps.day^{-1}) goals (Tudor-Locke & Lutes, 2009). Only one study has examined differences between step-goal conditions in children (Kang & Brinthaupt, 2009), showing no differences in steps between individual-standardised (+5% of participant's previous 2-week average) and group-standardised (+5% of group's previous 2-week average) goal conditions. Children in both conditions were able to view pedometer steps and were instructed to record daily steps in a diary. No further studies have examined the relative impact of goal type in children.

Therefore there is a need for further pedometer interventions to be developed to test the relative impact of individual-standardised and group-standardised step goals. The development of such physical activity interventions would typically be guided by a single behavioural theory (e.g. social cognitive theory, self-regulation theory) on the premise that focusing on influencing key constructs of this theory will improve intervention effectiveness (Gardner & Campagna, 2011). Criticisms of relying upon behavioural theories are that they have been derived to understand (i.e. explain or predict) behaviour, not behaviour change *per se* (Brug et al., 2005). It has also been argued that the design of an effective intervention should be problem driven as opposed to theory driven; that is focus should be given to using constructs derived from multiple theories (and multiple levels e.g. socio-ecological framework) to solve the pre-defined health problem, not to the single- theory approach adopted by many interventionists (Brug et al., 2005). The integration of multiple correlates/determinants into an intervention package requires a guiding framework to ensure evidence-selected variables are linked to the intervention aim and are appropriately targeted by theoretical change methods. It is then advantageous to the scientific community to ensure that these active intervention ingredients are clearly described for the benefit of future work.

Seemingly the development of effective physical activity interventions should be informed by an 'intervention theory', which involves the identification and description of the planning and evaluation of the intervention to ensure researchers can explain what happens in an intervention and why (Tudor-Locke et al., 2001). A commonly used 'intervention theory' approach in health promotion studies is the intervention mapping (IM) protocol (Bartholomew et al., 1998). Intervention mapping is a theoretical framework designed to guide the development, implementation and evaluation of a health behaviour intervention. There are 6 main steps to the framework; these are (1) a needs assessment, (2) identification of outcomes and behaviour change objectives, (3) identification of theoretical models and practical strategies to induce behaviour change, (4) programme development and pilot testing, (5) development of an implementation and adoption plan and (6) development of an evaluation plan. The IM protocol is therefore both a problem-driven and a theory-driven approach that was designed to ensure an integrative (i.e. multi-theory/multi-level) systematic and evidence-based 'best practice' approach to intervention development (Bartholomew et al., 1998). A more comprehensive summary of the protocol can be found at http://interventionmapping.com.

Of the nine pedometer interventions presented in the literature in Chapter Two of this thesis, only three studies explicitly made reference to the use of behavioural or psychological theory to inform their development (Butcher et al., 2007; Hardman et al., 2011; Horne et al., 2009) This is consistent with the findings of Tudor-Locke and Lutes (2009) who upon reviewing 27 pedometer studies, found only four outlined a theoretical model that guided intervention design. In addition, none of the nine pedometer intervention studies presented in the literature review of this thesis used an intervention or programme theory approach to guide their development. Clearly there is a requirement for the proposed pedometer-based PA interventions of this thesis to be guided by an intervention mapping approach to make salient the behavioural techniques used in the interventions, to allow consideration of intervention design issues (e.g. setting, duration and sequence), and to pilot-test developed intervention materials before use in the evaluation trial of Phase Two of this study.

7.3 Phase One aim

Therefore the aim of Phase One of this study is to present the development process of two pedometer step-goal interventions using a framework of intervention development. The development of these pedometer step-goal interventions was guided by a revised version of the intervention mapping protocol so that the active intervention ingredients were outlined, intervention design was evidence informed, and intervention materials were appropriate for the study population.

7.4 Intervention mapping protocol

The IM protocol is outlined for illustrative purposes in Table 7.1. The following sections will discuss the use of steps 1, 2, 3 and 4 of the IM protocol in the development of the two pedometer-based interventions.

Table 7.1 Intervention mapping protocol (Bartholomew et al., 1998; Prins, van Empelen, Beenackers, Brug, & Oenema, 2010).

Step	Task
(1) *Needs assessment*	• Plan and conduct needs assessment • Establish intervention outcomes
(2) *Matrices*	• Specify performance objectives • Specify determinants of target behaviour • Create matrices of change objectives
(3) *Theory-based methods and practical strategies*	• Identify theoretical methods • Choose intervention methods • Select or design strategies • Match strategies to change objectives
(4) *Intervention plan*	• Consult with participants and intervention implementers • Create intervention scope, sequence, theme and materials • Develop intervention protocol • Pre-test intervention materials with target group and intervention implementers
(5) *Adoption and implementation plan*	• Identify intervention adopters and participants • Specify adoption and implementation performance objectives • Specify determinants of performance objectives and create matrix, as per step 2 • Identify theoretical methods and strategies • Develop implementation plan
(6) *Evaluation plan*	• Describe intervention outcomes and effectiveness questions • Identify process-related outcomes • Identify or develop outcome measures • Specify evaluation study design

7.4.1 Step one - Needs assessment

Carrying out a needs assessment is the first stage of intervention development (Prins et al., 2010). This section of the IM protocol typically involves the description of the condition and risk behaviour of interest and the rationale for choosing it, a literature review to determine which theoretical constructs best predict/explain the chosen behaviour, and finally it is suggested focus groups be conducted to identify any localised barriers to behaviour change and determine any other potential correlates of the identified risk behaviour (Bartholomew et al., 1998). The evidence cited in the literature review (Chapter Two and in the introduction to this chapter) serve as justification for the development of an intervention to increase PA in primary-school-aged children. That is many children do not take part in sufficient levels of moderate-to-vigorous PA (Riddoch et al., 2007), and therefore are at greater risk of deleterious health risk status due to physical inactivity in childhood (Andersen et al., 2006). As pedometers have been shown to be successful motivators of increased PA in both children and adults (Kang & Brinthaupt, 2009; Lubans et al., 2009b) a review of the pertinent literature was undertaken to identify which pedometer intervention strategies have been successful, and which require further research. The critical evaluation of research literature in Chapter Two revealed that a combination of behavioural methods including goal setting (i.e. daily step goals), self-monitoring (i.e. recording daily step goals) and feedback (i.e. electronic display of cumulative steps taken) appears effective in increasing PA in primary-school-age children. The independent contribution of these combined strategies alone is unknown. It was also clear from this review that there is no strong evidence to support one type of pedometer step goal over another.

On completion of a review of literature it is recommended for focus groups to be conducted among a selection of the target population (Bartholomew et al., 1998). In the present study, however, focus groups were not conducted as part of step one, as it was not the intention of the current study to develop an 'effective' intervention based on correlates/determinants of primary school children's PA, the rationale for which will be made explicit in step two. From the literature reviewed the overall aim of this intervention study was to determine the effectiveness of pedometer goal setting, step count feedback and self-monitoring. A specific secondary aim was to determine the relative impact of individual-standardised and group-standardised step-goal conditions for increasing PA in primary-school-aged children.

7.4.2 Step two - Intervention outcomes and performance objectives

Following identification of the overall intervention aim or outcome, the next stage of the intervention mapping process is to state the performance objectives (Bartholomew et al., 1998). PA is well known to be a complex behaviour which is influenced by numerous factors (Sallis & Saelens, 2000), and is consequently difficult to increase, and maintain any increased levels. As a result a series of proximal behavioural objectives called 'performance objectives' are set (e.g. begin to monitor current PA levels). Following identification of performance objectives relevant to the risk behaviour, empirically or theoretically derived determinants of the behaviour that are appropriate to achieve the objective are selected (e.g. intention). These are finally mapped against a series of 'change objectives' in a matrix to outline exactly what the interventions' participants must action in order to effect the determinant and therefore meet the performance objectives (e.g. increase intention to monitor PA levels) (Bartholomew et al., 1998). Interventions may require multiple matrices for individual-level and environmental-level performance objectives to be linked with suitable determinants and change objectives.

Stage 2 of intervention mapping lends well to the development of interventions by the identification of behavioural determinants (from multiple behavioural theories/domains of influence) using a deductive approach. In the case of the present study, an inductive approach was taken, with the behavioural techniques of goal setting, self-monitoring and feedback identified a priori from the review of literature. These specific self-regulatory behavioural techniques may be linked to a number of behavioural theories, and could be used to influence a number of determinants of PA, dependent on one's theoretical approach. It was the intention of the present study to determine the effectiveness of specific behavioural strategies (i.e. type of step goal) alone. Therefore the identification of performance objectives and the selection of theoretically important determinants and additional behavioural strategies were not deemed appropriate. Indeed the introduction of additional active intervention components would interact with the combined implementation of goal setting, self-monitoring and feedback.

7.4.3 Step three - Identifying change strategies

The third step of the intervention mapping protocol is the selection and design of theoretical methods (behaviour change techniques) that are used to modify the previously determined change objectives. These techniques or theoretical methods are then linked to practical strategies, which outline exactly how the method will actually be implemented (Bartholomew et al., 1998). In the present study no performance objectives or change objectives were determined during stage two, so behavioural change techniques and practical strategies *per se* are presented in Table 7.2. Because different labels or terms are often used to describe the same theoretical method (Michie et al., 2009b), to ensure a standardised description of these methods, definitions are guided by a previously published taxonomy of behaviour change techniques (Abraham & Michie, 2008).

Table 7.2 Behaviour change techniques.

Behaviour change technique	Practical strategy
Individual-based goal setting	Daily step goals set as a percentage of individual-average baseline step count
Group-based goal setting	Daily step goals set as a percentage of group-average baseline step count
Provide feedback on performance	Participants given a pedometer so that visual feedback on cumulative steps taken is available
Self-monitoring	Participants will be asked to self-record daily step totals in a step diary and/or step calendar

The two goal types to be implemented were selected as individual-based and group-based goals. The only previous study to examine individualised- versus group-based goals (Kang & Brinthaupt, 2009) set step goals by increasing the participant's or group's average step count from the previous 2 weeks by 5%. As the current intervention could only be delivered by a single researcher it was decided that step goals be set to increase as a percentage over average baseline values by 5% each week (i.e. +5%, +10%, +15% over baseline average). Past pedometer studies have set step goals between 5-10% (Croteau, 2004; Kang & Brinthaupt, 2009) and therefore the selection of a 5% increase over baseline for each week of the intervention was deemed an achievable and appropriate target. Performance feedback on the number of steps accumulated will be achieved by the provision of pedometers with a liquid crystal display so that intervention participants can see their cumulative step totals in real time. Self-monitoring of PA will be achieved by the provision of individual step diaries, so that participants can record and chart daily step totals. The daily step goals (either individual- or group-standardised) for each participant will be written in the participant's step diary.

The ensuing section details the intervention outline, production of the intervention materials and pre-testing of these materials using a short focus-group interview.

7.4.4 Step four - Intervention development

In this step of intervention mapping the behaviour-change strategies are combined within a programme plan so that a clear outline of the full intervention is created (Bartholomew et al., 1998). This includes outlining the intervention scope, setting, duration and sequence. In addition relevant intervention materials should be generated and pilot tested with the target population and revised as and when appropriate (Bartholomew et al., 1998).

7.4.4.1 Setting

As children spend the majority of their waking hours in the school environment, and school-based interventions appear to be the most effective delivery setting for children and adolescents (van Sluijs, McMinn, & Griffin, 2007) it was decided that the intervention should be delivered in the classrooms of the targeted Worcestershire primary schools. The particular schools were selected as they had expressed the greatest interest in the project during recruitment.

7.4.4.2 Duration

The duration of the current intervention is dictated by the length of the primary school term. To ensure that baseline measurements can be taken for three intervention groups (individualised-goal group, group-goal group and no-goal control group) data collection will be scheduled to stagger over a period of 3 weeks (interjected by 1 week's half-term break) due to the limited number of Actiwatch accelerometers. The finite length of the school term (7 weeks), and the requirement for intervention end-point measurements to be staggered over a 3-week period, necessitate that the intervention duration itself be limited to 3 weeks. Theoretically this duration should be sufficient to elicit short-term behaviour change as short-duration (1-4 weeks) pedometer self-monitoring interventions have reported significant increases in step counts over baseline (Butcher et al., 2007; Shimon & Petlichkoff, 2009).

7.4.4.3 Sequence

In the first week of the programme children in the intervention groups will be given a digital pedometer and a step diary (see Appendix Three) and assigned a tailored or group-based step goal dependent upon which group they will be randomised to. The use of pedometers will be demonstrated to participants; and they will be provided with an instruction sheet (see Appendix Two) and given feedback on how to correctly attach the units. At the start of each school week (Monday) the previous week's step diaries will be collected and the participants given new diaries and assigned an increased step goal, starting in week one with +5% over baseline steps, to +10% in week two and finally +15% in week three. This process is described in more detail in Phase Two of this study.

7.4.4.4 Intervention materials

Step diaries were designed as an A5 booklet. The diary included the daily step target for the week, a space for each day of the week for participants to record daily step totals and a space to record the achievement or non-achievement of daily step goals (Appendix Three). In addition step graph sheets were devised so that participants could chart daily step totals, thus providing a visual means of self-monitoring PA behaviour. Single instructional sheets were also devised for both the Omron HJ-109-E and Omron HJ-104 pedometers (Appendix Two).

7.4.4.5 Pilot testing

To determine the feasibility of recording and charting daily step totals, acceptability of the programme materials and use of pedometers in the target population a sub-sample of five primary school children aged 10-11 years (2 boys, 3 girls) from the pre-identified study schools were invited to pilot test the materials. Participants were given the pedometer step diaries and step charts and were assigned a generic step goal of 10,000 steps.day^{-1} for a period of two full days of wear time. Subsequent to this, a focus-group interview was conducted. The study protocol received institutional ethics committee approval and written parental consent and child assent was obtained prior to the pilot testing taking place. The focus group lasted approximately 20 min and was semi-structured in nature with key questions on the suitability of the programme materials and use of the pedometer. The interview was audio-recorded throughout, with the audio-recording transcribed verbatim post interview. Following this, the interview transcripts were read over a number of times, before key themes pertinent to the use of the intervention materials and pedometer were identified.

Results of the focus groups showed that the children expressed concern over the use of the step chart to visually record their daily step totals.

"I found the step diary easy, but the step chart hard" (Participant 5)

The majority opinion was that the use of the step diary alone was sufficient to self-monitor daily step totals.

"Yeah I don't think you need that (points to step chart)" (Participant 3)

A key practical issue raised was that of being able to remember to attach the pedometer each morning, from which it was salient a strategy to increase wearing compliance would need to be implemented.

"Yeah you forget to put it (the pedometer) on when you're doing things" (Participant 4)

"You need some kind of alarm so you know, that it, you know when to put it on" (Participant 3)

A number of the children articulated a degree of confusion over how to use both the pedometer and step diary, with one suggesting that clearer instructions on how to use key functions on the pedometer such as the reset function could be given using a schematic diagram of the device.

"How you accidently change the mode, how you change it back, but you don't know what buttons to press. You could take a little picture of it with arrows on it like a diagram" (Participant 3)

A final function concern identified was that of the comfort of wearing the pedometer and the suitability of clothing to attach the pedometer to. The children reported that the pedometers were uncomfortable to wear and could irritate the skin.

"I found it hard to wear sometimes. It was just rubbing, so I wore it over my school uniform sometimes" (Participant 4)

On the basis of these findings the step charts were removed from the intervention materials as they were not deemed a necessary self-monitoring tool alongside the step diary, and were found to be confusing by the pilot study participants. As some of the participants had difficulty in recording step totals in the diary, an explicit example of what should be recorded in the diary has been included, alongside which clear verbal instructions will be given at the start of the intervention. To improve the likelihood of the participants remembering to re-attach their pedometer in the morning and wear it to school a sticker chart has been created so that each child will receive a sticker to place next to their name upon successfully presenting their pedometer to the class teacher at the start of the school day, acting as an incentive to wear their pedometer. In addition class teachers will be prompted to remind the class each day to wear their pedometer, and report to them if it has been lost so that a replacement can be issued. Furthermore, the pedometer instruction sheets were re-designed to include a schematic diagram of the pedometer (see Appendix Two), so that the multiple pedometer functions will be clearer and more 'child friendly'. The issue of comfort during pedometer wear will be addressed by verbally instructing and demonstrating to the intervention participants how to attach the pedometer to their outer waist belt, so that it does not irritate the skin.

7.4.5 Steps five and six - Evaluation and implementation plan

The purpose of step 5 of is to specify the diffusion process for the intervention by determining how the intervention will be adopted and implemented to maintain fidelity to the intended programme upon delivery (Bartholomew et al., 1998). This requires the intervention designer to determine what programme facilitators need to do to ensure reliable and appropriate intervention delivery (i.e. fidelity), and what the possible determinants are that may influence intervention adoption (e.g. outcome expectations that the intervention will have a positive effect). As the current intervention was delivered by a single researcher and it was not the purpose to create a replicable health promotion programme this step of intervention mapping was excluded.

The purpose of step 6 of the intervention mapping protocol is to outline the evaluation plan to determine the effectiveness of the proposed intervention. The end product of this step is to create a plan for evaluating the impact of the intervention upon determinants of the risk behaviour, the behaviour itself, and any health-related outcomes (Bartholomew et al., 1998). In the present intervention determinants of PA behaviour in children were not identified in step two of the process, therefore outcome measures were solely focused on changes in PA. A small-scale controlled experimental study was used to evaluate the effectiveness of these pedometer step-goal interventions. This is described in full in Phase Two of this study.

7.5 Phase One Discussion

The aim of Phase One of this study was to develop a pedometer step-goal intervention using the intervention mapping protocol as a theoretical framework. No pedometer-based interventions in children have been developed using a systematic intervention-theory approach. In adult populations the 'First Step Program' (Tudor-Locke et al., 2001) was developed using programme theory (analogous to intervention mapping) (Tudor-Locke et al., 2001) and reported increases of >3000 steps.day^{-1} relative to controls at intervention end-point (Tudor-Locke et al., 2004), the success of which has in part been attributed to the consideration of pedometer characteristics, programme components and programme participants during the development phase (Tudor-Locke & Lutes, 2009). Whilst it is simply the intention of the developed intervention to delineate which step goal elicits the most favourable response in PA behaviour, it is important to increase the likelihood of a positive intervention effect by careful consideration of intervention components and suitability of the developed intervention materials to participants.

Due to the foreseen timescale, the requirements of the present thesis and the atheoretical (in reference to behavioural theory) approach taken, reduced fidelity to the full intervention mapping protocol was adopted. Specifically, the focus groups to determine barriers and correlates of the risk behaviour in step one were not deemed appropriate as it was not the intention of this intervention to modify previously identified correlates of PA. Secondly, the identification of performance objectives and change objectives was not conducted as it was the intention of this intervention to answer a specific research question regarding the effectiveness of a behaviour change technique, as opposed to developing an integrative polytheoretically driven intervention. Finally an adoption/implementation plan was not used as this step is more relevant for larger scale multi-site health-promotion initiatives, whereby the developed intervention is adopted and delivered on multiple sites.

Even in a reduced format, the intervention mapping protocol followed here provided a useful framework for the development of this pedometer intervention. By clearly outlining the behaviour change components of this pedometer step-goal intervention, it allows future researchers to replicate these strategies if they are found to be effective. This is critical for there to be a cumulative science of behaviour change, which as Michie et al. (2009b) state permits ineffective interventions (theoretical methods) to be avoided, and effective interventions to be adopted and refined in future studies. In addition, the development and pre-testing of programme materials has ensured that the step diary, pedometer instruction sheet and pedometer are made as ‘child-friendly’ as possible, limiting the possibility that reduced adoption to these materials may influence intervention effectiveness. The power of intervention mapping is depicted well by this iterative process, as the protocol serves as a record by which the intervention planner can determine possible causation for intervention ineffectiveness (Bartholomew et al., 1998).

The pilot findings from the focus-group interview conducted following the small-scale pilot of materials are of practical importance for the implementation of the developed interventions in Phase Two of this study. In particular, important issues regarding the correct attachment of the pedometer were brought to light. Pedometer attachment may appear tenuous; however pedometer output can be biased in children if the device is tilted away from the vertical direction (Duncan, Schofield, Duncan, & Hinckson, 2007), by improper attachment to clothing for example. This matter has been highlighted formerly by Gardner and Campagna (2011) who reported that participants (adult women) in a pilot intervention reported concerns over the capacity of the pedometer to remain in position, particularly when wearing clothing made of thin material. This information is vital for programme developers, and serves to highlight the value of qualitative investigations into the acceptability of pedometer use in target populations. Currently there have been no such investigations in children or adolescents, and therefore there is sufficient requirement to build further evidence on matters such as pedometer functionality/acceptability.

7.6 Phase Two Introduction

As highlighted in the introduction to Phase One there is a paucity of research examining the impact of pedometer goal setting, step-count feedback and self-monitoring interventions per se and further only one study which has directly compared differing goal types in children. The purpose of the intervention mapping work in Phase One was to develop such an intervention with consideration of intervention transparency issues (i.e. outlining active BCTs) and the many factors (i.e. duration, setting, sequence, materials etc.) which may influence the likelihood of a positive intervention effect.

Through pilot testing of step goal diaries and pedometer instruction sheets, sample specific feedback from a focus group was used to further refine these materials. This included revising the pedometer instruction sheets and removing the step goal charts from the materials as these were not deemed a necessary self-monitoring tool by the pilot participants. In addition practical issues such as wearing compliance and wearing comfort were highlighted, and solutions to these issues proposed i.e. class pedometer sticker charts, teacher reminders and verbal wear instruction. The final intervention materials and plan were then generated, and evaluated in a controlled experimental study in two West-Midlands primary schools. The following section, Phase Two, reports on the methods and results of this experimental field study.

7.7 Phase Two aim

The aim of Phase Two of this study was to determine the impact of 3-week school-based interventions employing pedometer goal setting, step-count feedback and self-monitoring on accelerometer-measured PA in 10-11-year old children. A specific aim was to determine the relative impact of individual-standardised and group-standardised step goal conditions.

7.8 Phase Two Methods

7.8.1 Participants

Sixty-eight Year 6 children (boys: n = 27; height: 146.9 ± 7.7 cm; weight: 40.0 ± 7.6 kg; BMI: 18.5 ± 2.7 kg.m^2; girls: n = 41; height: 145.6 ± 7.2 cm; weight: 41.5 ± 11.6 kg; BMI: 19.3 ± 4.1 kg.m^2) aged 11.2 ± 0.3 years were recruited from two primary schools in the West Midlands region of England. The experimental protocol received institutional ethics committee approval and written parental consent and child assent was obtained prior to participants enrolling in the study.

7.8.2 Outcome measures

Physical activity

Accelerometry

Actiwatch 4 (AW4, Cambridge Neurotechnology, Cambridge, UK) accelerometers were placed on the wrist (side at participant's discretion), and were set to record at 10-s epochs. Participants were instructed to wear the device for five consecutive weekdays. Weekend days could not be measured due to an unexpected alteration to the initial data collection schedule.

Pedometry

Omron HJ-109-E and HJ-104 pedometers (Omron Healthcare, Milton Keynes, UK) were affixed to the waistband of the participant's right hip (in line with the patella). Participants were instructed to wear the device for five consecutive weekdays. Weekend days could not be measured due to an unexpected alteration (i.e. school rescheduling of intervention start date) to the initial data collection schedule.

Anthropometrics

All anthropometric measurements were taken in the morning to ensure standardisation across participants as per the findings of Study One. Height and weight measurements were taken according to the protocols outlined in Chapter Three.

Body mass index

BMI was calculated using the methods described in Chapter Three. Following the calculation of BMI-for-age (BMI percentile), participants were categorised as *underweight* (BMI ≤2nd centile), *healthy weight* (>2nd centile to <85th centile), *overweight* (≥85th centile to <95th centile) or *obese* (≥95th centile). Due to the low frequency of cases in each group participants were split into *non-overweight* (N-OW) (<85th centile) and *at least overweight* groups (AL-OW) (≥85th centile) for analysis purposes.

7.8.3 Sample size

G-power software (Version 3.1.3) was used to calculate the required sample size for the primary outcome of time spent in MVPA. The statistical test was set at *F*-test repeated-measures ANOVA within-between group design, the alpha level was set at 0.05, the effect size (Cohen's f) was set at small (f = 0.10), the number of measurements was set at 2 and the correlation between measurements was set to r = 0.9 to give a power of 80% using a within-between group design. The minimum sample size required, calculated with these parameters, was n = 54, which equates to n = 18 per group, with a power of 0.82.

7.8.4 Study design

Three classes of children were randomly assigned to one of three conditions: (a) individual-standardised goal group (IS), (b) group-standardised goal group (GS), or (c) open-pedometer control group (CON). All outcome measurements were taken at baseline and end-point. For both IS and GS participants, baseline measurements were taken 2 weeks prior to the start of the intervention due to a 1-week half-term holiday period (see Table 8.1). For the CON participants baseline measurements were taken the week immediately preceding the intervention. The study ran over a 9-week period and consisted of three phases: a 1-week baseline evaluation, a 3-week intervention period and a 1-week endpoint evaluation.

Table 7.3 Study Five timeline.

Study phase	**Week**								
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
IS baseline									
IS intervention									
IS end-point									
GS baseline									
GS intervention									
GS end-point									
CON baseline									
CON phase									
CON end-point									

IS=individual-standardised, GS=group-standardised, CON=control. Grey squares denote baseline or end-point evaluation, horizontal-lined squares denote the half-term period and black squares denote the intervention/control period.

7.8.5 Procedure

Baseline and end-point evaluation

At baseline and end-point on the first day (Monday) anthropometric measures were taken in a quiet corner of the school classroom. Pedometers and AW4 accelerometers were distributed individually by the researcher between ~09:00 and 10:30. The researcher demonstrated the correct placement of the pedometer (in line with the right patella) and AW4 (tightly affixed to preferred wrist). Each pedometer was sealed securely with sellotape and marked with the individual's participant identification number. Children were instructed to wear both devices at all times except when engaging in water activities, and when asleep in the case of the pedometer. The researcher returned to collect both devices 5 days later on the Friday, between 14:00 and 1500. The researcher made frequent visits during the week and resealed any unsealed pedometers, and replaced broken or lost pedometer units.

7.8.6 Intervention

Following baseline measurement the researcher calculated daily step goals for the children in the intervention conditions. Children in the IS condition were set daily step goals of +5% over their individual average baseline step count, whereas children in the group-standardised condition were set at +5% over the group average baseline.

On the first morning of the intervention the researcher verbally addressed the intervention class, outlining the intervention procedure (which had been discussed during a prior visit) and clarified any specific questions. Particular mention was given to comfortable pedometer wear, which was one of the issues highlighted by the focus groups in Phase One. Following this, each child was seen individually where they were given a pedometer, a step diary with their daily step goal inside (IS or GS dependent on condition) and the appropriate pedometer instruction sheet (see Appendix Two). During this time the researcher reiterated the process of recording daily step totals and goal attainment within the step diary. On the first morning of the second intervention week the researcher returned to the respective intervention class, collected the previous week's step diaries and issued each child with a new diary (including revised step goal) for the ensuing week. Lost or broken pedometers were replaced by the researcher. This process was repeated exactly the same for the final and third intervention week. Step goals started at +5% over baseline for week 1 and increased to +10% at week 2 and +15% for week 3.

Children in the control group were given an open pedometer to wear for the duration of the 3-week intervention with no further contact or materials provided. The use of an open pedometer alone has been shown not to induce reactivity or behavioural change in children (Butcher et al., 2007; Ozdoba et al., 2004). All children were reminded to wear their pedometer by their class teacher and all classes were provided with a sticker chart. These were the aforementioned solutions to the pedometer wearing compliance issues raised in the focus groups in Phase One. To promote compliance to the intervention the children would receive a sticker to place on the chart each day of the week they presented their pedometer to the class teacher. Throughout the intervention/control phase the researcher made frequent visits to each condition group during each week, to replace lost step diaries or pedometer units and to encourage compliance with the intervention protocol.

7.8.7 Data treatment and statistical analyses

7.8.8 Pedometer data

Participants required 3 days of pedometer data to be included in further analyses, as 3 days of measurement has been shown to produce a reliable estimate of pedometer-determined PA in youth (ICC = 0.9) (Craig et al., 2010) . A mean imputation method based on an individual information-centred (I-C) approach (Kang, Rowe, Barreira, Robinson, & Mahar, 2009b) was used to replace missing data. Where no step data were recorded at baseline, mean imputation (based on group information-centred methods (Kang et al., 2009b)) was used for the purpose of creating step goals for the GS condition. There were no participants in the IS group who returned totally incomplete step data at baseline.

7.8.9 Accelerometer data

To distinguish between actual sedentary behaviour and periods when the AW4 were removed, non-wear time was defined as periods of ≥20 min of consecutive zero counts, as per Esliger et al. (2005). In addition, sleep time between 22:00 and 06:00 was removed to prevent dilution of mean counts per time unit (Esliger et al., 2005). This time period was selected from visual inspection of the data, and is in line with previous limits used in a similar sample (Corder et al., 2010). Participants required at least 600 min of data on at least three separate weekdays to be considered for further analysis (Mattocks et al., 2008a). Activity diary data (i.e. removal periods, activity undertaken during removal) were not collected as imputation of missing data from diary records can be ambiguous and subjective (Esliger et al., 2005). Following the high levels of MVPA found when applying the Puyau et al. (2004) cut-points to the data from Study Three, a decision was made (see page 125) not to use the Puyau cut-points to reduce AW4 data in phase Two of this study. However, following this decision, Ekblom, Nyberg, Bak, Ekelund, & Marcus, in early 2012, published new intensity cut-points for wrist-placed AW4. Therefore raw activity counts were converted into minutes of MVPA using the cut-points (≥175 cts.10 s^{-1}) of Ekblom et al. (2012).

Total activity counts and minutes of MVPA were calculated as the sum total from valid days/number of valid days and mean counts per minute were calculated as sum of counts from valid days/minutes of registered wear time. As registered wear time differed between baseline and end-point by ~33 min (p = 0.01) total counts and minutes of MVPA were adjusted for wear time (volume/hours of registered wear time) as per (Graves, Ridgers, Atkinson, & Stratton, 2010). PA outcomes were therefore total activity counts, TC (cts.hr^{-1}), minutes of MVPA per hour (min.hr^{-1}) of registered wear time and mean counts per minute, CPM (cts.min^{-1}) of registered wear time. Participants were then split into two groups using baseline TC (50th percentile split) and categorised as *less-active* (<39,852 cts.hr^{-1}) or *more-active* (≥39,852 cts.hr^{-1}).

Complete pedometer step data were only available for n = 36, therefore only accelerometer data were used as an outcome measure in the final analyses. Participants who did not provide valid accelerometer (primary outcome) data (n = 18) at baseline and/or end-point were excluded from analysis of baseline characteristics, and any further intervention analyses (see Figure 8.1).

All statistical analyses were conducted using SPSS 19.0 (Chicago, IL, USA). Baseline descriptive statistics were calculated for all outcome measures. Normality of these variables was assessed per group using the Shapiro-Wilks test. Only BMI violated the normality assumption. A one-way ANOVA (Kruskal-Wallis for BMI) was used to determine differences in baseline characteristics between step-goal conditions and weight status groups. An independent samples *t*-test (Mann-Whitney *U* for BMI) was used to determine differences in baseline characteristics between ‘excluded’ and ‘final sample’ participants. A three-factor repeated-measures ANOVA was used to determine the effect of time (two levels: baseline and end-point), condition (three levels: IS, GS and CON) and PA level (two levels: less-active or more-active) on PA outcomes. To investigate significant main effects either independent or paired samples *t*-tests with Bonferroni adjustments were used. Where significant interactions were found, independent samples *t*-tests with Bonferroni adjustments were conducted. Independent samples *t*-tests were also used to compare Δ in PA outcomes (time) between weight-status groups within conditions. The alpha level for statistical significance was set at $p < 0.05$ for all tests.

7.9 Phase Two Results

7.9.1 Participant flow (CONSORT format)

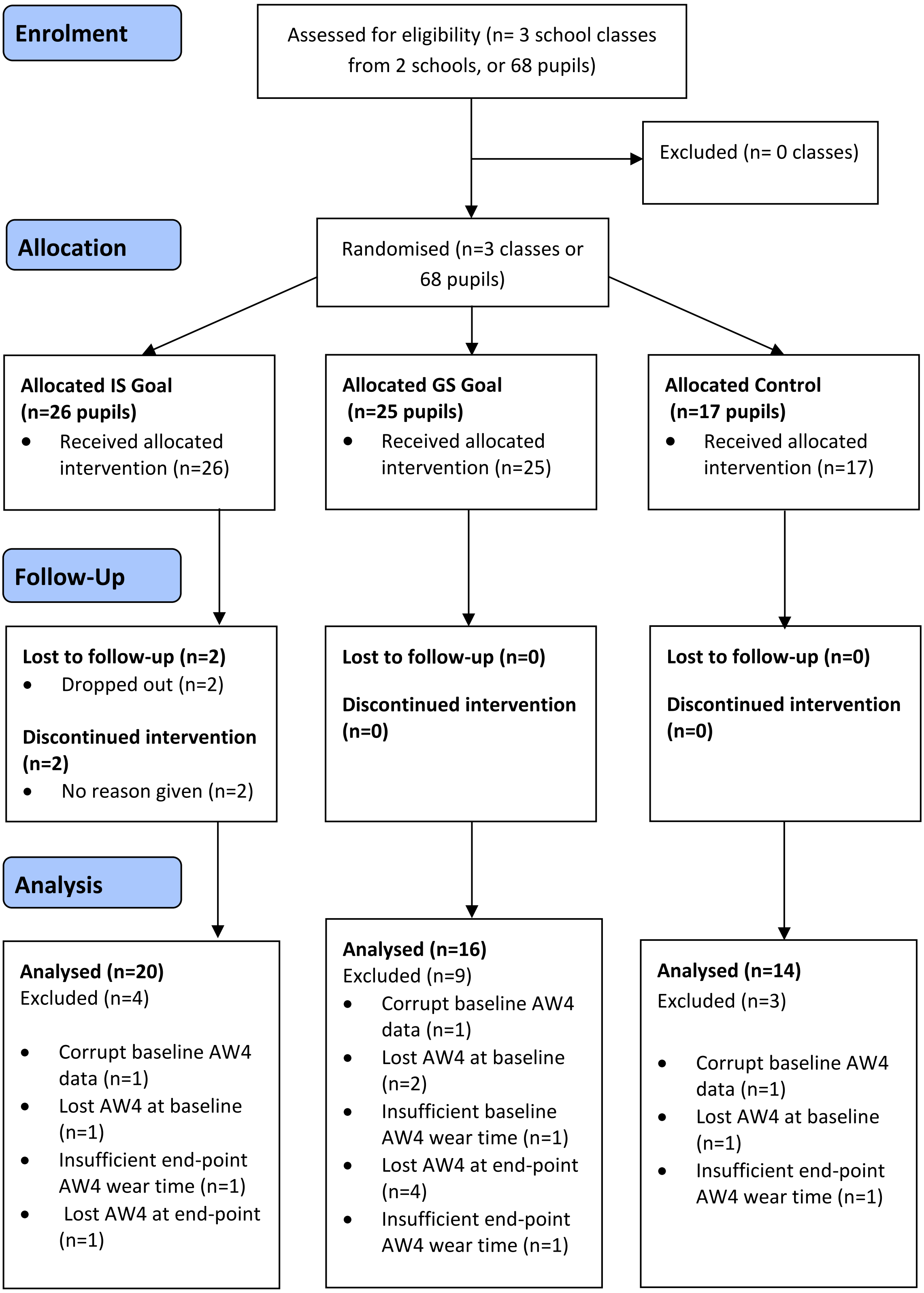


Figure 7.1 Participant progression through study phases.

Complete accelerometer data sets were available for 20 of the eligible 26 in the IS condition, 16 of the 26 in the GS, and 14 of the 17 eligible in the CON condition. Reasons for exclusion at either baseline or end-point included lost AW4 units (n = 9), corrupt data files (n = 3), insufficient wear time (n = 4) and drop-out (n = 2).

There were no differences in decimal age (p = 0.84), BMI (p = 0.28), steps (p = 0.15), TC (p = 0.98), CPM (p = 0.98) or MVPA (p = 0.71) between 'excluded' (n = 18) and 'final sample' (n = 50) participants at baseline.

7.9.2 Baseline characteristics

Baseline characteristics for the final sample (n = 50) are shown in Table 7.4 There were no differences in decimal age (p = 0.43), BMI (p = 0.96), steps (p = 0.40), TC (p = 0.35), CPM (p = 0.35) or MVPA (p = 0.60) between conditions at baseline. There were no differences in decimal age (p = 0.46), steps (p = 0.98), TC (p = 0.68), CPM (p= 0.68) or MVPA (p = 0.83) between N-OW and AL-LOW at baseline. There were no differences in age (p = 0.05) or BMI (p = 0.23) between *less-active* and *more-active* children.

Table 7.4 Baseline descriptive characteristics of the final sample (*n* = 50)

Variable	**Individual-standardised (IS)** ***n* = 20**	**Group-standardised (GS)** ***n* = 16**	**Control (CON)** ***n* = 14**
Age (years)	11.1 ± 0.3	11.3 ± 0.4	11.3 ± 0.3
Height (cm)	145.8 ± 6.7	146.0 ± 7.9	147.0 ± 8.6
Weight (kg)	41.7 ± 12.4	40.0 ± 9.3	40.0 ± 9.6
BMI ($kg.m^2$)	19.5 ± 4.8	18.6 ± 3.1	18.3 ± 3.2
Steps ($steps.day^{-1}$)	16,675 ± 5,133	18, 871 ± 3887	18,457 ± 4935
N-OW (%)	12 (60.0%)	10 (62.5%)	11 (78.6%)
AL-OW (%)	8 (40.0%)	6 (37.5%)	3 (21.4%)
Less-active (%)	9 (45.0%)	6 (37.5%)	10 (71.4%)
More-active (%)	11 (55.0%)	10 (62.5%)	4 (28.6%)

Table 7.5 presents means and standard deviations for all accelerometer-measured PA variables at baseline and end-point, as well as the difference between time points for all conditions.

Table 7.5 Physical activity variables for the whole sample (*n* = 50) for individual-standardised (IS), group-standardised (GS) and control (CON) conditions

Time point	**Condition**	**Total counts ($cts.hr^{-1}$)**	**Counts per minute ($cts.min^{-1}$)**	**MVPA ($min.hr^{-1}$)**
Baseline	IS	41,580.2 ±14,093.9	693.0 ± 234.9	13.4 ± 5.0
	GS	44,188.0 ± 9254.6	736.4 ± 154.2	14.0 ± 3.1
	CON	38,259.2 ± 7584.6	637.7 ± 126.4	12.5 ± 3.5
End-point	IS	42,238.9 ± 10,954.7	704.0 ± 182.6	14.1 ± 4.4
	GS	41,144.9 ± 7816.9	685.8 ± 130.3	13.0 ± 3.2
	CON	34,665.7 ± 7437.8	577.8 ± 124.0	11.0 ± 3.3
Time change (95 % CI)	IS	658.7 (-3105.6 to 4423.0) *NS*	11.0 (-51.8 to 73.7) *NS*	0.8 (-0.5 to 2.1) *NS**
	GS	-3043.1 (-5988.7 to -97.5) *NS*	-50.7 (-99.8 to -1.6) *NS*	-1.0 (-2.2 to 0.1) *NS*
	CON	-3593.4 (-6496.8 to -690.1) *NS*	-59.9 (-108.3 to -11.5) *NS*	-1.6 (-2.8 to -0.3) *NS*

NS = No significant time change ($p > 0.05$). *Difference in Δ MVPA between IS and CON ($p = 0.01$).

7.9.3 Main effects

Maulchly's test for sphericity was not computed for within-subject effects, as this assumption is met when the number of levels (in this case time, which had two levels) is two or less (Field, 2009). Levine's test for equality of variance was not significant for any of the PA outcomes ($p > 0.05$).

The main effect of time was significant for TC ($F(1,44) = 4.6$, $p = 0.04$) and CPM ($F(1,44) = 4.6$, $p = 0.04$), but not MVPA ($F(1,44) = 3.5$, $p = 0.07$). Bonferroni-adjusted paired *t*-tests revealed no differences in TC, CPM or MVPA between baseline and end-point in all conditions ($p > 0.016$). However, TC ($p = 0.019$), CPM ($p = 0.019$) and MVPA ($p = 0.016$) were approaching a significant decline in CON participants.

The main effect of condition was not significant for TC ($F(2,44) = 0.5$, $p = 0.60$), CPM ($F(2,44) = 0.5$, $p = 0.60$) or MVPA ($F(2,44) = 0.2$, $p = 0.80$).

The main effect for baseline PA level was significant for TC ($F(1,44) = 44.1$, $p = 0.01$), CPM ($F(1,44) = 44.1$, $p = 0.01$) and MVPA ($F(1,44) = 44.4$, $p = 0.01$). As expected TC (45,673.7 ± 7851.4 vs. 33,863.0 ± 7099.3 $cts.hr^{-1}$, $p = 0.01$), CPM (761.2 ± 130.9 vs. 564.4 ± 118.3 $cts.min^{-1}$, $p = 0.01$) and MVPA (15.3 ± 3.2 vs. 10.5 ± 3.1 min, $p = 0.01$) were greater in *more-active* children at intervention end-point.

7.9.4 Interaction effects

There was an interaction between time and condition for MVPA ($F(2,44) = 5.8$, $p = 0.01$), but not TC ($F(2,44) = 3.3$, $p = 0.05$) or CPM ($F(2,44) = 3.3$, $p = 0.05$). The Δ in MVPA did not differ (Bonferroni-corrected alpha = 0.016) between IS and GS children ($p = 0.04$) or GS and CON children ($p = 0.56$). Change in MVPA differed between IS and CON children ($p = 0.01$). MVPA increased in IS children by 0.8 ± 2.8 min, whilst MVPA decreased by -1.6 ± 2.1 min in CON.

There were no interactions between baseline PA and condition for TC ($F(2,44) = 1.9$, $p = 0.16$), CPM ($F(2,44) = 1.9$, $p = 0.16$) or MVPA ($F(2,44)=2.3$, $p = 0.12$). There were also no interactions between baseline PA, time and condition for TC ($F(2,44) = 0.1$, $p = 0.92$), CPM ($F(2,44) = 0.1$, $p = 0.92$) and MVPA ($F(2,44) = 0.1$, $p = 0.93$). From Figure 7.2, TC increased in *less-active* children in both IS and GS, and decreased in CON children. In *more-active* children TC declined in IS, GS and CON children. From Figure 7.3, MVPA increased in *less-active* children in the IS, whilst decreasing in the GS and CON children. MVPA in *more-active* IS children did not change, whilst it decreased in GS and CON children.

7.9.5 Weight-status moderation

There were no differences in Δ in TC, CPM or MVPA between N-OW and AL-OW children within in any condition ($p > 0.05$).

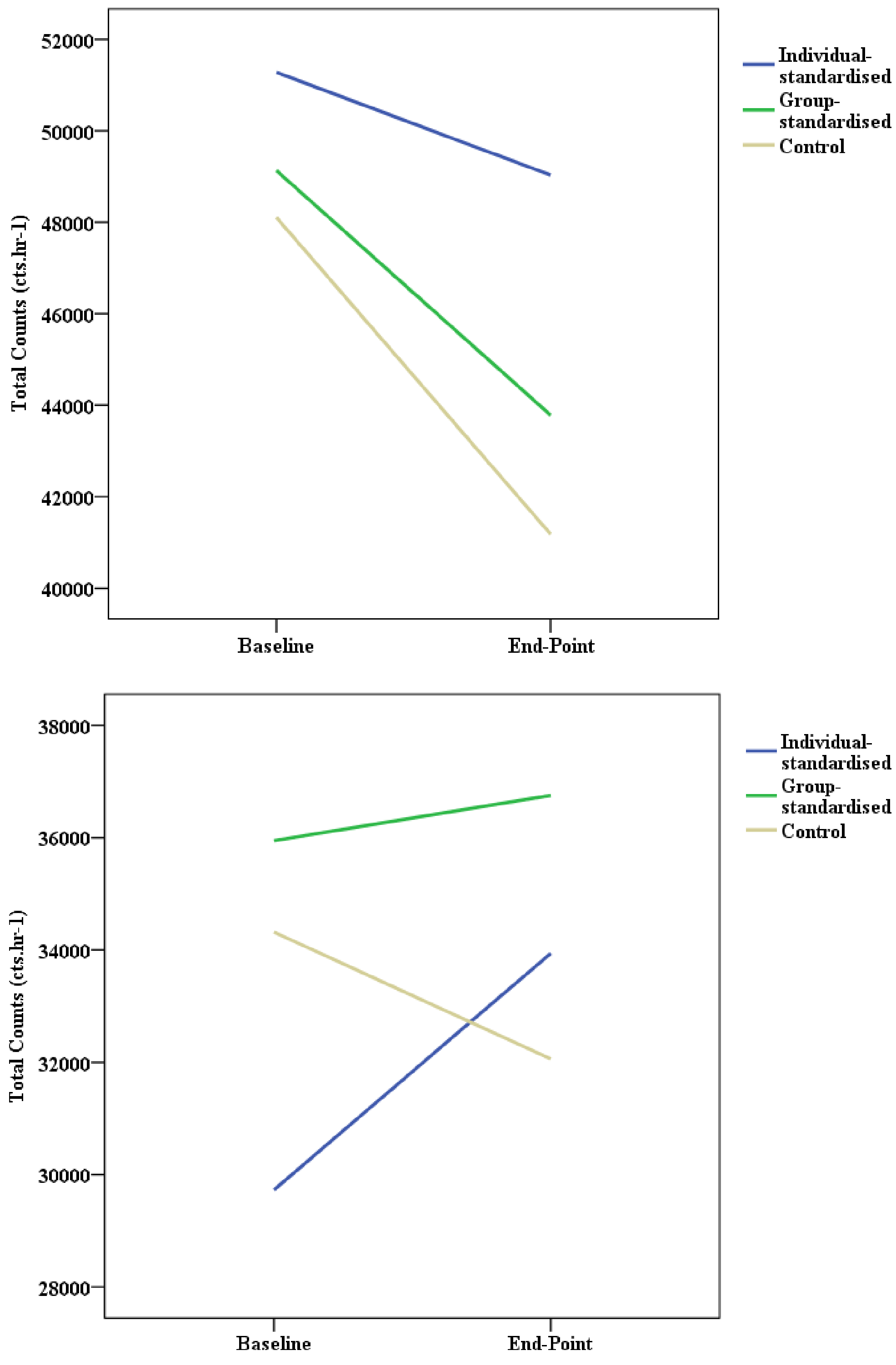


Figure 7.2 Total activity counts (TC) for less-active (bottom panel) and more-active (top panel) children in all conditions. Values are means, error bars are not shown for ease of interpretation.

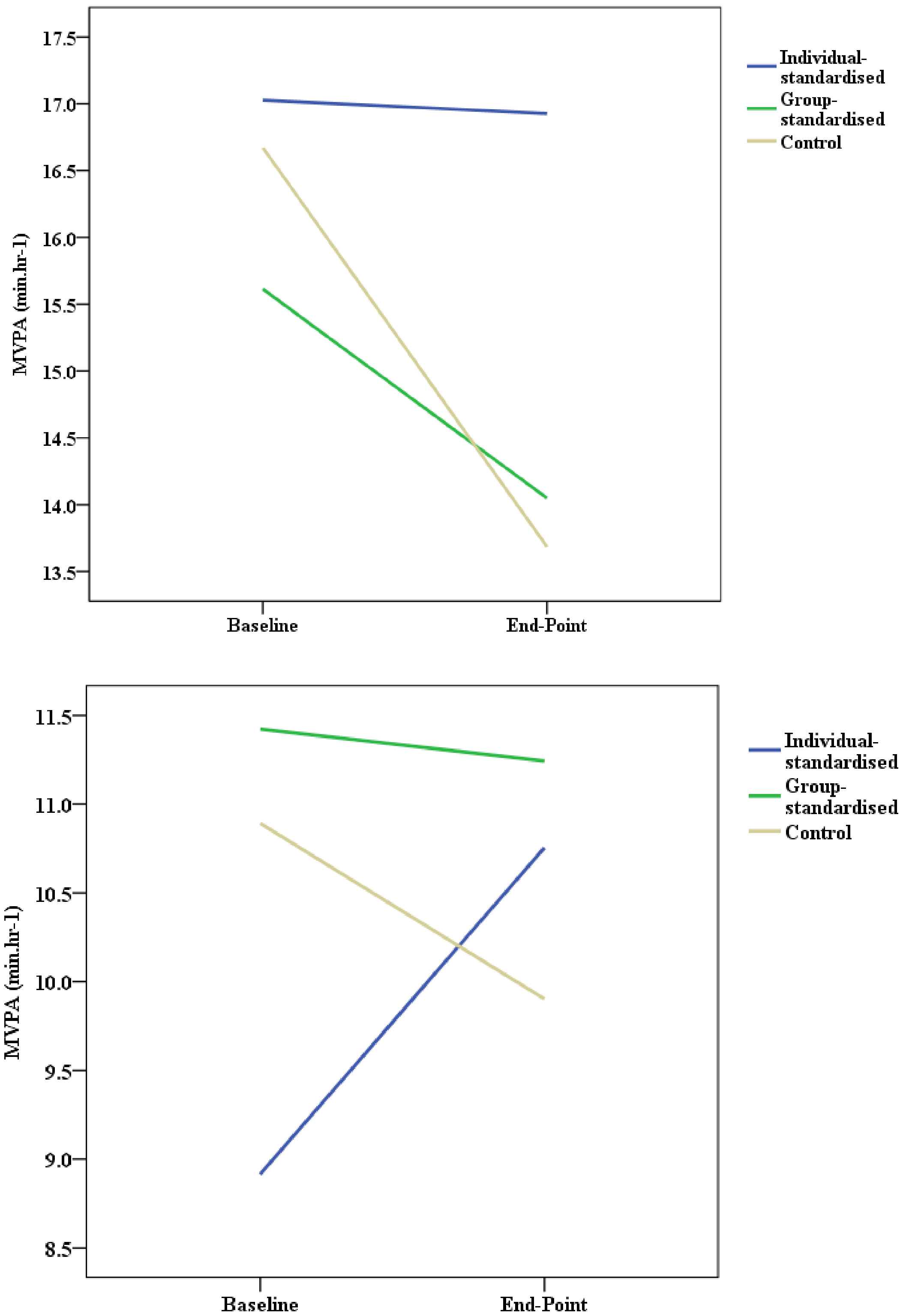


Figure 7.3 MVPA for less-active (bottom panel) and more-active (top panel) children in all conditions. Values are means, error bars are not shown for ease of interpretation.

7.10 Phase Two Discussion

The primary aim of Phase Two of this study was to examine the impact of 3-week pedometer-based interventions upon accelerometer-measured PA, with the specific aim of determining the relative impact of IS and GS step-goal conditions. Results for the whole sample showed that there were no differences between any groups in PA variables at intervention end-point, and no change in PA variables as a result of the intervention. This suggests that a combination of step-count feedback, goal setting and self-monitoring alone may not be sufficient to induce PA behaviour change in children.

The failure of this pedometer-based intervention to increase PA is not consistent with prior pedometer-based studies in children. For example, both 'Fit 'n' Fun Dude' interventions (Hardman et al., 2011; Horne et al., 2009) employed goal setting, self-monitoring and feedback strategies in similar age British children and observed increases in steps of ~2-3000 steps.day^{-1} over baseline. However both of these interventions also included peer modelling, tangible and contingent reward components as well as lengthy maintenance phases. The only directly comparable study, by Kang and Brinthaupt (2009), tested a 6-week step-count feedback, goal-setting (IS and GS conditions) and self-monitoring intervention. Across step-goal conditions there was a 19% increase in daily steps from baseline (5454 ± 1432 steps.day^{-1}) to end-point (6478 ± 2053 steps.day^{-1}). However the intervention duration was two times greater than the current study, and children only wore pedometers within the school day. In addition step goals were set as increments of the previous two weeks' step count as opposed to goals derived from percent over baseline. This disparity in findings implies that the effectiveness of combining step-count feedback, goal setting and self-monitoring alone in children is unclear, and warrants further examination over longer periods, including analyses using full week and weekend day data.

There were no main effects for step-goal condition, or interactions between time and condition (except for MVPA), suggesting that IS and GS goal setting was equally ineffective. The congruence in impact between step-goal conditions is comparable to the findings of Kang and Brinthaupt (2009). They reported no differences in daily step totals between children assigned to IS and GS conditions at intervention end-point. However they did report that daily step counts increased by ~19% from baseline when averaged over the two conditions. In adults, Sidman et al. (2004) observed no group differences at intervention end-point in adult women assigned to either a personalised standardised goal condition (self-set increment over baseline) or a universal (10,000 $steps.day^{-1}$) goal condition. Previous systematic reviews of pedometer-based interventions in adults have reported similar effect sizes (ES) for individual- and universal-/group-standardised step goals. Kang et al. (2009a) found adult studies using a strategy of 10,000 $steps.day^{-1}$ as a goal had the highest ES (0.84), while those that prescribed individualised goals (mixed adult and child data) reported a mean ES of 0.72.

Earlier studies in children (Kang & Brinthaupt, 2009) and adults (Sidman et al., 2004; Sidman, Corbin, & Rhea, 2003) have suggested goal attainment (i.e. meeting daily step goal) may be dependent upon both step-goal type and baseline PA level. Kang and Brinthaupt (2009) found goal attainment to be higher in low-active children prescribed IS goals (16.4 ± 3.1) than in those with GS step goals (6.2 ± 4.9), but there were no differences between conditions for medium- or high-active children. Sidman et al. (2003) found that very-low active and low-active women assigned a universal goal (10,000 $steps.day^{-1}$) achieved their step goals on a reduced number of occasions in comparison to more-active women. In addition, Sidman et al. (2004) reported low-active women assigned a universal step goal (10,000 $steps.day^{-1}$) had lower goal attainment compared to both medium- and high-active women, and compared to all activity levels in women assigned a self-selected individual-based goal. These data suggest that goal attainment appears to be dependent upon baseline PA level, with limited evidence (Kang & Brinthaupt, 2009) that individual-based goals may be more appropriate for low-active children. Interestingly lower goal attainment did not result in any differences in daily step totals between low-active children in either goal condition at end-point in the study by Kang and Brinthaupt (2009). In the study by

Sidman and colleagues (2004) there were also no differences in daily step totals in women assigned either a universal or a personalised step goal at end-point. In the present study there were an insufficient number of step-goal diaries to permit meaningful analyses due to both children and teachers losing the diaries, damage to the diaries, pedometer malfunctions and inappropriately recorded values.

There were no interactions between baseline PA, time and condition. However, from Figures 7.2 and 7.3 total PA (TC) increased in *less-active* children, markedly in the IS condition (~4000 $cts.hr^{-1}$), and marginally in the GS condition (~500 $cts.hr^{-1}$). *More-active* children's TC declined in all conditions. The decline in *more-active* children may result from a ceiling level whereby initial increases in PA are unsustainable, and accordingly PA levels regress. If children failed to meet step goals in week 1, the goal in week 2 and 3 was increased regardless as it was derived as an increment over baseline. This would have increased the difficulty of goal attainment. Goal attainment has not been found to influence end-point steps in prior studies (Kang & Brinthaupt, 2009; Sidman et al., 2004), but the present child study population may be less resistant to goal failure. Reduced goal attainment and negative step feedback in these *more-active* children may have reduced self-efficacy, a known mediator of children's PA (Lubans, Foster, & Biddle, 2008), negatively impacting on motivation to increase PA. The greater decline in TC in *more-active* children assigned a GS goal may also be because a number of children in the GS group were set lower goals than their baseline values, thus offering no motivational stimulus to increase PA.

Interestingly MVPA increased (~1.5 $mi.hr^{-1}$) in *less-active* children prescribed IS goals, but decreased marginally in those prescribed GS goals. In *more-active* children MVPA appeared stable in the IS children, but declined markedly in GS and CON children. *Post-hoc* tests revealed that Δ in MVPA differed between IS and CON children (p = 0.01). MVPA increased in IS children by 0.8 ± 2.8 min, whilst MVPA decreased by -1.6 ± 2.1 min in CON. It appears that the primary driver of participant response to this pedometer intervention was baseline PA level (consistent with previous studies e.g. Oliver et al., 2006), with IS goal setting to some extent mitigating the unfavourable response in *more-active* children. There were however no main effects for step-goal condition, or time x condition x PA level interactions. Regarding this interaction, *post-hoc* power values revealed that the current sample size did not have sufficient power to detect effects in this sub-group analysis. Despite low power, visual analysis-of-means plots provide tentative evidence that IS goal setting may be more suitable for both *less-active* and *more-active* children, but appears to impact to a greater extent on *less-active* children.

Analysis of Δ in PA outcomes between *non-overweight* and *at least overweight* individuals within each condition revealed no differences between groups. Despite the small number of cases in each group, this analysis provides some evidence that children of differing BMI-determined weight-status groups did not respond differently to the intervention. However, as this analysis was likely underpowered and previous pedometer studies have called for the examination of weight status as a moderating variable (Horne et al., 2009; Tudor-Locke & Lutes, 2009) it is important for future pedometer-based studies to further examine this issue.

The change in MVPA observed between baseline and end-point, across all conditions was not significant and thus may not appear instantaneously meaningful. In the IS condition MVPA increased by 0.8 (95% CI: -0.5 to 2.1) min.hr^{-1} (p = 0.23), in the GS condition it decreased by -1.0 (95% CI: -2.2 to 0.1) min.hr^{-1} (p = 0.07) and in the CON condition it decreased by -1.6 (95% CI: -2.8 to -0.3) min.hr^{-1} (p = 0.02). However, when multiplying these values by the mean wear time from both time points (15.0 ± 0.6 hr), the decrease in MVPA in the GS condition was ~15 min, and ~24 minutes in the CON condition. When considering that a difference of ~15 min of MVPA has been shown to reduce odds of obesity in similar age children (Ness et al., 2007; Riddoch et al., 2009), this is a biologically meaningful decline in health-enhancing PA. Therefore the fact that the IS condition maintained MVPA is arguably as important as a behavioural increase. Indeed, due to limited funding and resources, and the age-related decline in PA, there have been suggestions that preference should be given to PA maintenance as opposed to PA promotion (Jago, 2011).

There are numerous factors that may explain the failure of this combined step-count feedback, self-monitoring and step-goal intervention to significantly improve PA. The intervention duration was limited (3 weeks), and may not have been of sufficient length to modulate PA. Nonetheless, short-duration pedometer self-monitoring interventions have reported significant increases in step counts over baseline (Butcher et al., 2007; Shimon & Petlichkoff, 2009). It is plausible that any initial increases in PA had diminished by intervention end-point. Yet this hypothesis is unfounded as mid-point PA evaluations were not conducted. When considering that the staggered baseline evaluation meant that the study ran from May to July 2011, it is also surprising that PA declined in the CON condition, when an increase in daylight hours would have provided more opportunity for children to partake in outdoor play/activities.

The finding that this intervention did not increase PA serves to strengthen arguments that multiple levels of influence on PA should be targeted to provide maximum likelihood for increasing PA in youth (Van Der Horst, Paw, Twisk, & Van Mechelen, 2007). Systematic reviews of youth interventions have emphasised the importance of the physical environment, altering policies and parental/familial involvement for PA modification (Kriemler et al., 2011; van Sluijs et al., 2007). School-based interventions may be limited in their ability to address wider social and physical environment correlates, but there are factors that could be included within school-based interventions including the introduction of playground markings (Ridgers, Fairclough, & Stratton, 2010), standing desks (Benden, Blake, Wendel, & Huber, 2011) and active class rooms (Lanningham-Foster et al., 2008). To increase the likelihood of a positive effect, future pedometer-based interventions delivered in schools could implement environmental modifications and involve parents/significant others to create a more PA-permissive physical and social environment.

There are several limitations to Phase Two of this study. The final sample (*n* = 50) had borderline sufficient power to detect main intervention effects, and clearly did not have sufficient power to be able to detect change in PA outcomes when sub-group analyses were conducted. Further, the two intervention groups were recruited from within the same school and therefore were open to cross-over contamination effects; however the dissimilar time change in PA at the sub-level (i.e. baseline PA) between groups suggests contamination may have been minimal. A particular limitation was the measurement of only weekday data (due to disruption in the original baseline measurement schedule), therefore possible intervention effects on these days were missed. Future interventions should capture weekend data as there are known differences in PA pattern between weekdays and weekend days (Rowlands, Pilgrim, & Eston, 2009).

The strengths of Phase Two of this study are that the PA outcomes were generated using AW4 accelerometers that had previously been laboratory tested and found to exhibit good intra- and inter-instrument reliability (i.e. CV~5%). These devices were affixed to the same monitor position, and where possible the exact same model was used for each participant at each time-point to minimise artificial data variance. The low number of invalid accelerometer data files due to insufficient wear time provides indirect evidence for acceptability of wrist-worn activity monitors, supporting the assertions made in Study Three of this thesis. This is the first study to employ an accelerometer to measure PA in a field-based pedometer intervention in children. Pedometers are limited by their inability to provide a measure of PA intensity and pattern, which are key outcomes for determining activity-health relationships. Where possible, future pedometer-based interventions should utilise activity monitors capable of capturing these dimensions of PA.

7.11 Chapter Conclusion

The development of two pedometer step goal interventions in Phase One using a reduced format of the intervention mapping protocol ensured evidence-based and transparent interventions (i.e. clarity on active BCTs) were produced. By briefly piloting the intervention materials with the chosen study population and conducting a focus group, practical issues regarding pedometer step self-monitoring, device function, comfort and wear adherence were highlighted prior to implementing and evaluating the interventions in Phase Two. The evaluation of the interventions (developed in Phase One) in Phase Two of this study is unique as it is the first to examine the impact of pedometer-based self-monitoring, feedback and goal setting *per se* against pedometer-wearing controls. It is also only the second study to examine the importance of step-goal type in children. The primary findings were that in the whole sample a combination of pedometer-based self-monitoring, feedback and goal setting did not lead to an increase in PA. Despite no difference in PA or PA change between step-goal conditions, there was a trend towards IS goal setting maintaining PA volume and MVPA in *more-active* children, and increasing PA volume and MVPA in *less-active* children; whereas in children prescribed a GS goal there was a marginal increase in PA volume in *less-active*, a marked decline in *more-active*, and declines in MVPA in both *less-active* and *more-active* children. Whilst not supported statistically these data provide tentative preliminary findings to suggest that IS goal setting is more effective than GS for improving PA using pedometer programming, and is less influenced by baseline PA level.

Future studies that involve pedometer programming as a key component of an intervention should consider the use of an intervention theory to guide development by giving clarity on active BCTs, ensuring intervention content is evidence based, and to pre-test the suitability of intervention materials and equipment prior to implementation and evaluation. In regard to intervention evaluation, future work should use larger samples and longer duration interventions, with a measure of fidelity (i.e. step diary analyses), to clarify if pedometer-based goal-setting, feedback and self-monitoring *per se* is sufficient to change PA. These studies should also seek to determine if gender, PA level and weight status moderate goal attainment and the main intervention effect, and further compare the IS approach against the GS approach, before other goal type comparisons are considered.

8 Chapter Eight - General Discussion

The primary aim of this thesis was to determine the effectiveness of pedometer-based goal-setting, self-monitoring and step-feedback interventions for increasing PA in primary-school-aged children. This research question was informed by a review of pedometer-based PA interventions in children. The literature review demonstrated that there was emerging evidence that pedometer use is associated with increased daily step counts in children. It was clear however that the optimal deployment and independent contribution of pedometer-based strategies within these interventions was unclear, due to the inclusion of additional components such as counselling, peer-rewards and curricular changes etc. in many intervention studies. Alongside evidence from systematic reviews of pedometer literature in children (Lubans et al., 2009b) and adults (Bravata et al., 2007; Kang et al., 2009a), there was a lack of evidence to support the relative impact of setting different types of step goals. Therefore this gap in the pedometer literature was also examined in the intervention study (Chapter Seven) of this thesis.

Before the effectiveness of this step-goal intervention could be examined, there were a number of research issues that warranted investigation. When examining moderators of pedometer-based intervention effects in children, studies have been shown to be more successful in increasing PA in girls or low-active children (Horne et al., 2009; Oliver et al., 2006). Despite calls to examine pedometer-based intervention effects by weight-status groups (Horne et al., 2009; Tudor-Locke & Lutes, 2009) no prior pedometer investigations have determined intervention effects by weight-status groups. As weight-status categorisation using BMI could potentially be affected by diurnal variation in weight and height, it was therefore necessary to determine if the timing of measurements within a school day could impact upon weight-status categorisation (Chapter Four).

Another issue considered in line with the examination of pedometer step-goal conditions was the use of an accelerometer-based PA monitor as an outcome measure, since prior interventions had only used pedometer step counts, providing a metric of the volume of PA only. As the technical reliability of accelerometers set the upper limit of unit validity (Esliger & Tremblay, 2006), and the extent of this had not been quantified in the accelerometer model (AW4) selected for use in this thesis, it was first obligatory to examine the reliability of the AW4 (Chapter Five). Alongside device reliability, further monitoring issues associated with AW4 use were identified. As an issue of practicality it was deemed preferable to place the device on the wrist, as opposed to the waist to potentially increase participant compliance (van Hees et al., 2011). However the agreement in AW4 output between hip and wrist placements, the potential effect of body side dominance and the possibility of deriving hip-measured intensity minutes from wrist data was hitherto unknown (Chapter Six).

Finally, no pedometer-based intervention studies in children have utilised a programme theory/mapping framework to guide intervention development. As intervention mapping provides a transparent audit trail of intervention development, and provides a practical guide for the design of behavioural interventions, the development of the intervention study of this thesis (Chapter Seven) followed a revised format of the intervention mapping protocol, with particular reference to outlining active BCTs and pilot testing intervention materials in the identified study population.

The summary above represents a brief review of what this thesis set out to investigate. The following section will revisit the purpose of each study, and discuss in more detail the main findings and implications for future research.

The first objective of this thesis was to determine the impact of diurnal variation in weight and height upon BMI-determined weight-status categorisation in children. In Chapter Four measurements of weight and height were taken in morning and afternoon measurement sessions, and BMI weight status computed using both population monitoring (85th and 95th centile) and clinical BMI cut-offs (91st and 98th centile). In addition, a sub-sample of children was re-measured following the initial morning measurement session to determine intra-observer measurement reliability.

Results showed that height decreased systematically in boys and girls, but weight increased from morning to afternoon in girls only. When expressed using the population monitoring cut-offs this resulted in three of the girls (7.5%) increasing BMI weight category between morning and afternoon measurements with two moving from the *healthy weight* to *overweight* category and one moving from *overweight* to *obese*. When using the clinical cut-offs, change in category from morning to afternoon was observed in one girl, moving from the *healthy weight* to the *overweight* category. Differential categorisation between morning and afternoon was not observed in boys however.

This is the first study to demonstrate the effect of additive interactions between diurnal variation in height and weight upon BMI-determined weight category in children. These findings suggest that at the individual level BMI category may potentially be classified differently in a number of girls when measured in the afternoon compared to the morning. To minimise differential weight-status categorisation due to the timing of measurements*,* the time of day when measurements of height and weight are conducted should be standardised. As notionally if a greater proportion of *healthy weight* or *overweight* individuals were measured in the afternoon within a particular study group, the intervention effect could a) be biased i.e. not a reflection of 'true' intervention response within a given weight category and b) comparison of weight status categories between intervention groups would be unsound as 'like for like' comparison is not being made. Indeed it is arguable that the children who sit on the boundary of *overweight* and *obese* weight categories could respond quite differently to a physical activity intervention compared to individuals in the upper percentiles of the *overweight* and *obese* categories. A simple standardisation of time when measurements are conducted would remove this issue.

There are a few limitations of this particular study however. The sample size was relatively small (*n* = 74) when compared to larger studies examining BMI-determined weight status measurement issues (Townsend, Rutter, & Foster, 2011). Allied with the relatively small sample is the fact that both measurements were not taken at the extremes of the day. If measurements had have been taken immediately following the start of school (i.e. 09:00) and in the later afternoon (i.e. 17:00) it is possible a greater number of children may have been classified differently. This time frame would also more accurately represent the time period used by PA researchers (e.g. 09:00-18:00) in intervention measurement protocols in children. Logistical constraints precluded examination of change across this time period within the current thesis however.

Future research incorporating larger samples and undertaken at greater extremes of the day (i.e. early morning and later afternoon) is required to confirm the additive interaction between diurnal variation in height and weight upon BMI-determined weight category in children. The implications of these findings are also relevant to public health surveillance programmes such as the National Child Measurement Programme (NCMP), who annually track BMI and weight status in Year 6 and reception- age children. In regard to public health surveillance, future research should examine additional sources of measurement variation in NCMP data. This could for example include variation attributable to the many school nurses (inter-observer reliability) who conduct these measurements, with limited anthropometric training, and whom conduct measurements under time constraints, in school conditions not well suited to maintaining accuracy.

The second objective of this thesis was to quantify the intra- and inter-instrument reliability of the AW4 accelerometer. In Study Two batches of AW4 units were attached to an isokinetic dynamometer and accelerated at speeds representative of moderate and vigorous intensity on two separate occasions. Results showed the AW4 to have acceptable intra- and inter-instrument reliability (i.e. CV ~5%) in terms of raw activity counts, with improved reliability observed when the devices were accelerated at vigorous intensity speeds. The reliability statistics (both intra- and inter-) observed equated with previous research in other accelerometer models (Esliger & Tremblay, 2006; Krasnoff et al., 2008; Powell et al., 2003). When placing the raw continuous activity counts into discrete categories of time spent in MPA and VPA, variance within and between devices decreased, with the greatest reliability exhibited when data were expressed as combined MVPA. Importantly the level of percent error between units (i.e. individual unit agreement with group-unit values) was improved when data were expressed as MVPA.

This is the first study to demonstrate the AW4 to have acceptable reliability for use as a PA monitor either in a subject-mounted or mechanical reliability experiment. While there was a 'small' significant difference (~20 cts.10 s^{-1}) between trials for raw activity counts, when accelerated at speeds representative of moderate and vigorous intensity the AW4 can reliably categorise data into the health-enhancing intensity category of MVPA, the main implication being that when separate variables of MPA and VPA are not required, the expression of AW4 data as MVPA optimises data reliability. On this basis the AW4 was deemed reliable for use in Studies Three and Five of this thesis. Due to the degree of inter-instrument error observed, where possible all participants wore the exact same unit at both measurement time points in Study Five.

Particular weaknesses of this study include the limited range of test speeds and repeat trials used. The reliability of the AW4 was only tested under speeds representative of moderate and vigorous intensity, thus providing no information on the consistency of this device at the lower end of the intensity spectrum (i.e. sedentary behaviour, light PA). Ideally the test speeds would also elicit count outputs that were closer to the intensity cut-point thresholds, so that a more stringent evaluation of the misclassification rate of raw count data between intensity categories is made. Finally inclusion of two to three additional repeat trials would provide evidence for the consistency of the AW4 on several occasions. This is of interest as PA interventions often involve baseline, end-point and multiple follow-up evaluations.

Following the completion of Study Two, a recent validation/calibration study using the AW4 also examined the intra- and inter-instrument reliability using a mechanical shaker (Ekblom et al., 2012) at four different intensities (180, 240, 280 and 320 rpm) on three occasions. The CVintra for raw activity counts was 1.72% at 180 rpm, 0.94% at 240 rpm, 1.17% at 280 rpm and 0.72% at 320 rpm. The CVinter for the four intensities ranged between 6.6 and 8.4%. These low CV values confirm the findings of Study Two, and provide additional evidence that the AW4 provides a reliable output. This study by Ekblom et al. (2012) builds upon the limitations of Study two, as the performance of the AW4 was examined within a wider range of test speeds. The lower CVintra values reported by Ekblom et al. (2012) likely result from the use of a mechanical shaker, the differing experimental design, and possibility that these were unused, and therefore newly calibrated AW4 units.

In Study Two variance in AW4 output was negatively related to the speed and frequency of the test condition. Future studies should seek to determine if variance in AW4 output differs across a greater range of physiologically meaningful test conditions (i.e. sedentary to vigorous) using an engineering grade mechanical shaking device as per Esliger and Tremblay (2006). However, whilst mechanical laboratory experiments elucidate the variance solely attributable to the accelerometer unit itself, studies where unit attachment, placement and experimental protocols are highly controlled are also required to determine the reliability of the AW4 during a range of free-living activities- as ultimately these devices will be deployed for use in the field. Finally, accelerometer end-users should calibrate all units (i.e. test unit reliability) before deployment in the field and ensure units are within acceptable limits of instrument error (i.e. <5-10%). Unit calibration should also be conducted following field use to determine the extent of calibration drift and provide information on whether unit reliability increases with extended use/age. Acceptable parameters of instrument variance will depend on the research question and study population (i.e. sedentary or highly active etc.).

The third objective of this thesis was to determine whether AW4-measured PA counts and MVPA differed between the hip and wrist, worn on the dominant and non-dominant sides in children and to investigate the predictive ability of wrist-measured

MVPA and demographic variables against hip-measured MVPA. In Study Three children wore AW4 accelerometers on the dominant wrist, non-dominant wrist, dominant hip and non-dominant hip during a large segment of the school day. Results showed that raw activity counts and minutes of MVPA were greater at both wrists compared to hip placement sites. There were however no differences in any PA outcomes between either body sides at the wrist or hip respectively. When applying hip-derived intensity cut-points, the regression model to predict hip-measured MVPA from wrist-derived data was accurate at the group level only. Due to the magnitude of the standard error and width of the 95% limits of agreement, the developed equation is deemed inaccurate for use at the individual level.

This is the first study in either children or adults to explicitly compare accelerometer output between hip- and wrist-placement sites and is the only to compare output between body sides. The greater accelerometer output at the wrist is supported by previous observations of a stronger acceleration signal at upper limb sites compared to trunk placements during mixed static/dynamic lifestyle activities in adults (Esliger et al., 2011; Kumahara et al., 2004). The greater wrist output is likely a reflection of differential movement patterns and activation periods during free living and orientation of the AW4 piezoelectric sensor within a different directional plane. Clearly, these data show that comparisons between wrist- and hip-derived accelerometer data would not be valid. The finding that side dominance does not impact upon AW4 output is novel as prior contralateral placement site studies have simply compared left versus right placements (Cook & Lambert, 2009; Fairweather et al., 1999; McClain et al., 2007).

According to the current results, and converse to the manufacturer’s instructions (Cambridge Neurotechnology, 2008), side dominance is not a monitoring issue when the AW4 is placed at the hip or wrist in primary-school-aged children. Thus the AW4 (in this population) may be placed on either hip or wrist without risk of biasing data. Confirmation of the absence of a side-dominance variance component is important as this will allow research participants in similar populations to select their preference for which side the accelerometer is worn for the duration of monitoring. Therefore to boost compliance, participants in Study Five wore the AW4 on their preferred wrist. As accelerometer data quality is related to participant compliance to monitoring protocols (Trost, McIver, & Pate, 2005b) it is paramount that accelerometer end-users seek methods to maximise participant compliance. In young children in particular, preference for monitor placement may increase compliance to monitoring protocols, especially over extended monitoring periods used for habitual PA assessment.

The derived regression equation to predict hip-derived MVPA from wrist data displayed a large SEE, and examination of Bland-Altman plots showed considerable variation in prediction accuracy at the individual level. As noted earlier, whilst group-level prediction was accurate, this variation could reduce the ability to detect between group differences in predicted non-dominant hip-derived MVPA due to the introduction of additional variance to the data set.

Without such variation this equation would have provided an efficient field-based alternative to a full-value calibration study in the laboratory, and allow MVPA intensity data to be derived from wrist-placed units. As wrist placement may improve acceptability (and potentially compliance) (van Hees et al., 2011), and the hip-derived cut-points of Puyau et al. (2004) produced questionably high MVPA values in this population, it was decided that the AW4 would be placed on the wrist, and data be expressed as raw PA output (i.e. total and mean counts per time unit) in Study Four. However, as noted above (page 148), following completion of Study Three a full-value calibration of the AW4 at the wrist placement site has been conducted in children aged 9-10 years (Ekblom et al., 2012). Ekblom et al. (2012) calibrated the AW4 against PAEE, with children performing a mixture of eight locomotor and lifestyle activities. Validity coefficients for the non-dominant wrist placement were good (r = 0.7) and cut-points to define sedentary, light, moderate and vigorous intensity were derived with acceptable levels of sensitivity and specificity (74-98%). Therefore these cut-points were used to reduce AW4 data in Study Five.

The limitations of Study Three include the small sample size (*n* = 24) and the relatively short monitoring period. A larger sample may have reduced the variation in the individual-level prediction of the derived regression equation, and therefore led to an equation that was usable in the ensuing studies of this thesis. It is possible that collecting data over a longer monitoring period in a larger more heterogeneous sample including a greater range in participant characteristics (i.e. leg length, body composition) and activity patterns may provide different results. A free-living study over 7 days including full weekday and weekend days would provide a more conclusive picture of differences in AW4 output within and between placement sites.

Future studies should also cross-validate the cut-points of Ekblom et al. (2012) as these will be accurate only to the extent that the calibration source activities and participants are representative of the population they will be used in. For optimal accuracy sample-specific cut-points should be derived, however this is time consuming and may not alter activity-health relations (Stone et al., 2009). Therefore, there is still an on-going requirement for future AW calibration studies to be conducted in larger samples, using standardised protocols to ensure more representative intensity thresholds. Further studies can then investigate the diagnostic accuracy of the most commonly used cut-points for the AW (as per Trost, Loprinzi, Moore, and Pfeiffer (2011)) which would provide evidence to support the use of one set of cut-points over another.

In addition, as both wrist and hip placements have been validated for use in the AW4, the question must be which placement should be used in future? Placement should be dictated by the magnitude of association between counts and PAEE. Whilst not as strongly related to PAEE as a trunk placement (Puyau et al., 2002; Puyau et al., 2004), wrist placed AW have been validated (Ekblom et al., 2012). The wrist placement would appear preferable as it is inherently less obtrusive than the waist, shows good acceptability in adults and may enhance participant compliance (van Hees et al., 2011). Further work is required to compare compliance as a function of placement site, and also to examine the acceptability of wearable motion sensors in children considering not only monitor placement, but monitor size and design, duration of wear and wearing protocol as well as methods to promote protocol compliance (e.g. automatic wear time reminders).

The fourth objective of this thesis was to develop two pedometer step-goal interventions to increase habitual MVPA in children aged 10-11 years. In Phase One of Study Four an intervention mapping (IM) framework (Bartholomew et al., 1998) was used to guide the development of the pedometer interventions, evaluated in Phase Two of Study Four. Following the IM protocol, the behavioural strategies of self-monitoring, performance feedback and goal setting to be implemented in Phase Two were clearly defined, and the methods of goal setting described. Following this, the intervention setting, duration and sequence of the intervention were determined and described. Once the intervention materials (i.e. step diary, pedometer instruction sheet etc.) were designed, they were piloted in a sub-sample of 10-11 year old children. Following, a focus-group interview was conducted to determine the suitability of the programme materials and use of the pedometer in this population. Results revealed a number of practical concerns regarding pedometer use (i.e. comfort of pedometer wear, forgetting to wear pedometer and uncertainty over how to use the pedometer) and the suitability of the daily step chart as a self-monitoring tool.

Feedback from the focus-group interview was important, as it lead to alterations in the programme materials (i.e. removal of step chart, design of pedometer instruction sheet with schematic diagram, implementation of daily reminder sticker chart) and the verbal instruction given at the start of the intervention. Whilst appearing tenuous, these process-related issues if unattended to may have impacted upon participant fidelity to the intervention. There is a requirement for future work to examine practical issues regarding pedometer use in this population. These include issues of pedometer design to increase comfort, methods to improve wearing compliance, and more engaging methods to promote self-monitoring of pedometer steps.

Importantly the functional behavioural techniques (pedometer strategies) evaluated in Phase Two were clearly defined, so that judgement could be made upon the active intervention ingredients and the intervention outcome. The poor reporting of intervention components has been discussed at great length by behavioural scientists (Michie et al., 2009b), and is seen as a barrier to explaining the success or failure of a given intervention. To date, no pedometer-based interventions have been developed using an IM approach, and the interventions themselves have been vaguely described, leaving the optimal method of pedometer programming unclear. Future pedometer-based interventions should use an IM or programme theory framework to guide development so that intervention methods are clearly described and consideration is given to the suitability of programme materials.

The fifth and primary objective of this thesis was to determine the impact of a 3-week pedometer-based intervention to increase PA in children aged 10-11 years. A secondary aim was to determine the impact of individual- and group-standardised step-goal conditions. In Phase Two of Study Four three classes of primary school children were randomised to one of three conditions: (a) individual-standardised goal group (IS), (b) group-standardised goal group (GS), or (c) open-pedometer control group (CON). Children in both intervention groups received a pedometer and a step diary with their respective daily step goal inside. Children in the control condition received an open pedometer and no further materials/support. At the intervention end-point there were no differences between groups, nor were there any significant time differences for any accelerometer-measured PA outcomes (i.e. TC, CPM and MVPA). There was an interaction between condition and time for MVPA with Δ in MVPA differing between IS and CON children. MVPA increased in IS children by 0.8 ± 2.8 min, whilst MVPA decreased by -1.6 ± 2.1 min in CON. There were no further interactions for condition and time, or baseline PA level, condition and time. Examination of means plots stratified by baseline PA level revealed that there was a trend towards IS goal setting maintaining MVPA, and minimising reductions in TC in *more-active* children, and increasing both in *less-active* children. While in children set a GS goal there was a small increase in total PA in *less-active*, a noticeable decline in *more-active*, and declines in MVPA in both *less-active* and *more-active* children.

This is the first study to examine the impact of a goal-setting, step-count feedback, and self-monitoring intervention against a control group in children. It is also only the second study to explicitly investigate the relative impact of different step-goal types in children. Data from Study Four suggests a combination of these pedometer strategies alone may not be sufficient to induce change in PA behaviour in primary school children. This finding opposes Kang and Brinthaupt (2009) who reported, in the only other study to examine similar pedometer strategies in children, a 19% increase in mean daily steps from baseline following a 6-week school-based intervention. In addition prior pedometer-based interventions in children have reported increases in mean daily steps of +300-3000 over baseline (Berry et al., 2007; Butcher et al., 2007; Goldfield et al., 2000; Goldfield et al., 2006; Hardman et al., 2011; Horne et al., 2009; Kang & Brinthaupt, 2009; Oliver et al., 2006; Rooney et al., 2005), however these have employed additional behavioural strategies alongside. The short duration of the intervention, potential lack of compliance to the intervention protocol and small sample size may account for the lack of significant findings in the current intervention study.

In regard to the differential impact of step-goal condition, data from Study Four and from analogous studies in children (Kang & Brinthaupt, 2009) and adults (Sidman et al., 2004) provide no hard evidence to support the prescription of individual-based step goals above group-based or universal goals. Whilst not supported statistically, in the current study it appeared that individual-standardised goal setting may be more successful, regardless of baseline PA level. This concurs with the assertions of Tudor-Locke (2002) who suggests that the most appropriate approach is to personalise step goals using baseline values, alongside considering health goals and long-term sustainability. Indeed Sidman et al. (2003) suggest that the “effort adherence trade-off” may also determine the appropriateness of goal type, that is as the effort required to meet a step goal increases, adherence to the goal may decrease. There is however no consistent empirical data to support one type of goal over another in children.

Despite the equivocal evidence regarding step-goal type, in the current study it was apparent that the intervention only impacted positively on *less-active* individuals, potentially as a result of a ceiling effect in *more-active* individuals. Thus baseline PA level may dictate response to pedometer-based interventions in children, which is supported by the findings of Oliver et al. (2006) and Horne et al. (2009). In addition, the examination of BMI-determined weight status as an intervention modifier revealed no differences in intervention impact between *non-overweight* and *at least overweight* children. This is a positive finding, in as much that it appears that *at least overweight* children do not respond any less favourably than *non-overweight* children, suggesting that pedometer-based programming may be suitable for all weight-status groups. Further research in larger samples is required to examine this assertion across the full weight-status spectrum (i.e. *underweight, healthy weight, overweight obese*) however.

There are a number of limitations regarding this intervention study which are worthy of mention. Firstly whilst the sample had nearly (n = 50 vs. n = 54) sufficient a-priori calculated power (~0.80) to detect main effects, the interaction/effect-modifier analyses were underpowered according to SPSS-derived *post-hoc* power estimates. Second, a measure of goal attainment was not available due to the number of incomplete/missing step diaries. It is therefore unknown (a) if goal attainment was related to the marginal differences between goal conditions observed, (b) if sufficient adherence to the intervention protocol was maintained, and (c) the extent to which this may have impacted on either goal condition. Finally the intervention was of a short duration (3 weeks), which may not have been of sufficient length to induce PA behaviour change.

To address these limitations and build upon the findings of Study Four, there are a number of recommendations that can be made. There is a necessity for a larger sample and longer duration study to be carried out to really determine the impact of pedometer goal-setting, step-count feedback and self-monitoring against controls. With a longer duration also comes the need to determine the long-term acceptability of pedometer use. As yet the longest duration pedometer-based intervention in children was ~16 weeks (Hardman et al., 2011). It is unknown if regular pedometer wear can be sustained in children beyond this threshold. For example, does the novelty factor wear off over prolonged periods? Prospective qualitative investigations should examine if issues such as pedometer function (i.e. steps, calories, distance walked, self-monitoring software etc.) and design (i.e. size, colour, and comfort to user) are related to adherence to prolonged pedometer use.

A larger sample investigation will permit sufficiently powered sub-group analyses to determine if baseline PA level, weight status or even gender moderate the impact of pedometer-based interventions in children. It is important to determine which sub-group responds best to these types of interventions if maximum health impact is to be made of limited resources. Indeed Jago, Fox, Page, Brockman, and Thompson (2010) have suggested that many interventions designed to increase children's PA have failed due to their aim of maximising reach by delivering an intervention to an entire school, or school year group.

As individual-standardised goal setting appeared to be slightly more successful (although not supported statistically), the effectiveness of this goal approach against group-standardised goals should be examined further. It is important to address this issue first as personalised goal setting may require more advanced self-regulatory skills and intensive training from a researcher, which may not be appropriate for primary school aged children. Further, the use of a universal step goal, whilst low on researcher burden, is not tailored to the study population's PA level. The individual-standardised approach however, anchors to the participant's baseline PA without as much researcher burden. If the individual-standardised approach is not found to be any more or less effective than group-based goal setting in future work, group-based goals should then be compared against universal goals (ideally criterion-referenced step thresholds related to BMI-determined risk of overweight/obesity for example) and personalised goal setting. As suggested by Tudor-Locke and Lutes (2009), it may be that working towards any step goal that represents an increase over baseline values is, for children at least, the most important. It is significant to determine if this is the case before deploying pedometers in more comprehensive behaviour change interventions. This future work should also include an assessment of goal attainment, and even measures of perception of effort related to step goals, as suggested by Sidman et al. (2003), to permit explanation of why one goal type may be more successful than another.

In sum, by understanding how pedometers are best deployed, and in which sub-population they have the greatest impact, public health practitioners and researchers will be able to maximise the value of pedometers within PA interventions in children. It is hoped that continued research on goal type, intervention moderators, pedometer design and practical issues related to prolonged pedometer use will be able to provide an evidence-based template on how to optimise the use of these simple step-counting devices in children. To build upon the findings of this thesis, and address the limitations of the work in this thesis the following issues should be addressed in future pedometer intervention work in children:

- Larger samples should be used to ensure adequate statistical power and permit sub-group analyses.
- Interventions durations should be lengthened to the maximum feasible to increase the likelihood of PA change.
- Further goal setting, self-monitoring and step-feedback pedometer-based interventions should be run to determine if this approach *per se* is sufficient to change PA.
- Measures of intervention fidelity (such as analysis of returned step goal diaries) should be implemented.
- Individual-standardised approaches should be compared against group-standardised before other goal type comparisons are considered.

- Main intervention effects should be examined by PA level and weight status to determine which sub-group responds best or indeed if any groups respond detrimentally to the intervention.
- Studies should measure goal attainment and perception of effort required to meet goals, to be able to explain why a given goal type is more or less successful.
- Formative qualitative data should be generated with the target study population to pre-test intervention materials and address practical intervention issues.
- Accelerometers should be used as an outcome measure to determine if change in health-enhancing MVPA, or indeed in any other intensities of PA or SB has occurred.
- Researchers using BMI-determined weight status as a moderating variable should standardise the time of day when measurements are conducted.

9 Chapter Nine - Conclusion

The primary aim of this thesis was to determine the effectiveness of goal setting, self-monitoring and step-feedback pedometer-based interventions for increasing physical activity in Year 6 children. To address this research question a number of interlinked measurement-related objectives were set out. The objective of Study One was to determine the impact of diurnal variation in weight and height upon BMI-determined weight status. Results showed a small proportion of *healthy weight* girls were susceptible to differential and unfavourable weight status classification when measured in the afternoon compared to morning. To ensure reliable categorisation for all participants in Studies Three and Four of this thesis, measurement timing was standardised to the morning - at the start of the school day. The objective of Study Two was to ascertain the reliability of the Actiwatch 4 in a controlled mechanical experiment. Data from this experiment suggested that the Actiwatch 4 displays satisfactory intra- and inter-instrument reliability, with data reliability improved when separate categories of moderate and vigorous intensity are combined into a moderate and vigorous (MVPA) intensity category. Consequently intensity data in Study Three and Study Four were presented as MVPA, and participants in Study Four (intervention study) wore the same unit, where possible, for repeat accelerometer assessment.

The objective of Study Three was to examine the degree of dissonance in Actiwatch 4 output between the hip- and wrist-placement sites, and if appropriate to create a regression equation to predict hip-derived MVPA from wrist data. Results showed wrist-derived data were greater than hip derived. The regression equation was not deemed acceptable for future use due to inaccurate individual level prediction and the large standard error of the estimate observed. At this point a decision was made not to use this prediction equation, nor to use previously published cut-points for the hip placement and to express future data as raw activity count variables. However following the later publication of new wrist intensity cut-points for the Actiwatch 4 (Ekblom et al., 2012), MVPA data were reduced using these cut-points in the intervention study of this thesis (Study Four). The objective of Phase One of Study Four was to develop the goal setting, self-monitoring and step-feedback pedometer-based interventions using a reduced format of the intervention mapping protocol. This clearly outlined the active behaviour change ingredients used in Phase Two, and a pilot study highlighted important practical pedometer issues (i.e. pedometer attachment/ wear comfort, wearing compliance and suitability of instruction sheets) that required addressing before the interventions were implemented and evaluated in Phase Two of Study Four.

The findings of Studies One to Three, and Phase One of Study Four, all contributed to the methods employed in Phase Two of Study Four, which aimed to determine the effectiveness of goal setting, self-monitoring and step-feedback pedometer-based interventions. A secondary purpose was to determine the relative effectiveness of individualised-standardised versus group-standardised pedometer step goals. Findings from this intervention study revealed that a 3-week goal-setting, self-monitoring and step-feedback intervention is not sufficient to increase physical activity in 10-11-year old children. In addition individual-standardised goal setting may be a more promising intervention strategy as this approach appeared to mitigate any decline in MVPA in *more-active* children, and increased MVPA in *less-active* children. Therefore future researchers should examine the effectiveness of the individual-standardised goal type against group-standardised further in a longer-duration intervention and using a larger sample of children, which will also allow sub-group analyses to be conducted. Of most importance however is future clarification on the effectiveness of goal setting, self-monitoring and step-feedback pedometer-based interventions *per se* for changing PA in children.

10 Chapter Ten - References

Abraham, C., & Michie, S. (2008). A taxonomy of behavior change techniques used in interventions. *Health Psychology, 27*(3), 379-387.

Adamo, K. B., Prince, S. A., Tricco, A. C., Connor-Gorber, S., & Tremblay, M. (2009). A comparison of indirect versus direct measures for assessing physical activity in the pediatric population: a systematic review. *International Journal of Pediatric Obesity, 4*, 2-27.

Adams, M. A., Dolan, P., & Hutton, W. C. (1987). Diurnal variations in the stresses on the lumbar spine. *Spine, 12*(2), 130-137.

Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., . . . Leon, A. S. (2000). Compendium of physical activities: an update of activity codes and MET intensities. *Medicine and Science in Sports and Exercise, 32*(9 Suppl), S498-504.

Andersen, L. B., Harro, M., Sardinha, L. B., Froberg, K., Ekelund, U., Brage, S., & Anderssen, S. A. (2006). Physical activity and clustered cardiovascular risk in children: a cross-sectional study (The European Youth Heart Study). *Lancet, 368*(9532), 299-304.

Baquet, G., Stratton, G., Van Praagh, E., & Berthoin, S. (2007). Improving physical activity assessment in prepubertal children with high-frequency accelerometry monitoring: a methodological issue. *Preventive Medicine, 44*(2), 143-147.

Barfield, J. P., Rowe, D. A., & Michael, T. J. (2004). Interinstrument consistency of the Yamax Digiwalker pedometer in elementary school-aged children. *Measurement in Physical Education and Exercise Science, 8*(2), 109-116.

Barlow, S. E., & Dietz, W. H. (1998). Obesity evaluation and treatment: expert committee recommendations. The maternal and child health bureau, health resources and services administration and the department of health and human services. *Pediatrics, 102*(3), E29.

Bartholomew, L. K., Parcel, G. S., & Kok, G. (1998). Intervention mapping: a process for developing theory- and evidence-based health education programs. *Health Education and Behaviour, 25*(5), 545-563.

Bassett, D. R. (2000). Validity and reliability issues in objective monitoring of physical activity. *Research Quarterly for Exercise and Sport, 71*(2), 30-36.

Basterfield, L., Adamson, A. J., Parkinson, K. N., Maute, U., Li, P. X., Reilly, J. J., & Gateshead Millennium Study Core, T. (2008). Surveillance of physical activity in the UK is flawed: validation of the Health Survey for England Physical Activity Questionnaire. *Archives of Disease in Childhood, 93*(12), 1054-1058.

Bauman, A. E., Sallis, J. F., Dzewaltowski, D. A., & Owen, N. (2002). Toward a better understanding of the influences on physical activity: the role of determinants, correlates, causal variables, mediators, moderators, and confounders. *American Journal of Preventive Medicine, 23*(2 Suppl), 5-14.

Beets, M. W., Patton, M. M., & Edwards, S. (2005). The accuracy of pedometer steps and time during walking in children. *Medicine and Science in Sports and Exercise, 37*(3), 513-520.

Benden, M. E., Blake, J. J., Wendel, M. L., & Huber, J. C., Jr. (2011). The impact of stand-biased desks in classrooms on calorie expenditure in children. *American Journal of Public Health, 101*(8), 1433-1436.

Berry, D., Savoye, M., Melkus, G., & Grey, M. (2007). An intervention for multiethnic obese parents and overweight children. *Applied Nursing Research, 20*(2), 63-71.

Biddle, S. J., Gorely, T., & Stensel, D. J. (2004). Health-enhancing physical activity and sedentary behaviour in children and adolescents. *Journal of Sports Sciences, 22*(8), 679-701.

Biddle, S. J. H., Brehm, W., Verheijden, M., & Hopman-Rock, M. (2011). Population physical activity behaviour change: a review for the european college of sport science. *European Journal of Sport Science, 12*(4), 367-383.

Blair, S. N., Kohl, H. W., 3rd, Paffenbarger, R. S., Jr., Clark, D. G., Cooper, K. H., & Gibbons, L. W. (1989). Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA : The Journal of the American Medical Association, 262*(17), 2395-2401.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 8;1*(8476), 307-310.

Boreham, C., & Riddoch, C. (2001). The physical activity, fitness and health of children. *Journal of Sports Sciences, 19*(12), 915-929.

Bornstein, D. B., Beets, M. W., Byun, W., Welk, G., Bottai, M., Dowda, M., & Pate, R. (2011). Equating accelerometer estimates of moderate-to-vigorous physical activity: in search of the Rosetta Stone. *Journal of Science and Medicine in Sport, 14*(5), 404-410.

Brage, S., Brage, N., Wedderkopp, N., & Froberg, K. (2003). Reliability and validity of the computer science and applications accelerometer in a mechanical setting. *Measurement in Physical Education and Exercise Science, 7*(2), 101-119.

Bravata, D. M., Smith-Spangler, C., Sundaram, V., Gienger, A. L., Lin, N., Lewis, R., . . . Sirard, J. R. (2007). Using pedometers to increase physical activity and improve health. *JAMA: The Journal of the American Medical Association, 298*(19), 2296-2304.

Brawley, L. R., & Latimer, A. E. (2007). Physical activity guides for Canadians: messaging strategies, realistic expectations for change, and evaluation. *Canadian Journal of Public Health, 98* (Suppl 2), S170-184.

Brug, J., Oenema, A., & Ferreira, I. (2005). Theory, evidence and intervention mapping to improve behavior nutrition and physical activity interventions. *International Journal of Behavioral Nutrition and Physical Activity, 2*(1), 2. doi: 10.1186/1479-5868-2-2

Butcher, Z., Fairclough, S., Stratton, G., & Richardson, D. (2007). The effect of feedback and information on children's pedometer step counts at school. *Pediatric Exercise Science, 19*(1), 29-38.

Calle, E. E., Thun, M. J., Petrelli, J. M., Rodriguez, C., & Heath, C. W., Jr. (1999). Body-mass index and mortality in a prospective cohort of U.S. adults. *New England Journal of Medicine, 341*(15), 1097-1105.

Cambridge Neurotechnology. (2007). Information bulletin no 4B: Actigraphy activity analysis. Cambridge.

Cambridge Neurotechnology. (2008). The Actiwatch guide to getting started. Cambridge.

CDC. (2011). Physical activity for everyone.   Retrieved 08/03/2012, 2012, from http://www.cdc.gov/physicalactivity/everyone/guidelines/children.html

Chen, K. Y., & Bassett, D. R., Jr. (2005). The technology of accelerometry-based activity monitors: current and future. *Medicine and Science in Sports and Exercise, 37*(11 Suppl), S490-500.

Cheuvront, S. N., Carter, R., 3rd, Montain, S. J., & Sawka, M. N. (2004). Daily body mass variability and stability in active men undergoing exercise-heat stress. *International Journal of Sport Nutrition and Exercise Metabolism, 14*(5), 532-540.

Cole, T. J., Bellizzi, M. C., Flegal, K. M., & Dietz, W. H. (2000). Establishing a standard definition for child overweight and obesity worldwide: international survey. *British Medical Journal, 320*(7244), 1240-1243.

Cole, T. J., Freeman, J. V., & Preece, M. A. (1995). Body mass index reference curves for the UK, 1990. *Archives of Disease in Childhood, 73*(1), 25-29.

Colley, R. C., Garriguet, D., Janssen, I., Craig, C. L., Clarke, J., & Tremblay, M. S. (2011). Physical activity of Canadian children and youth: accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. *Health Reports, 22*(1), 15-23.

Cook, I., & Lambert, E. V. (2009). Monitor placement, sources of variance and reliability of free-living physical activity: a pilot investigation. *South African Journal of Sports Medicine, 21*(1), 13-18.

Cooper, A. R., Page, A., Fox, K. R., & Misson, J. (2000). Physical activity patterns in normal, overweight and obese individuals using minute-by-minute accelerometry. *European Journal of Clinical Nutrition, 54*, 887-894.

Corder, K., Ekelund, U., Steele, R. M., Wareham, N. J., & Brage, S. (2008). Assessment of physical activity in youth. *Journal of Applied Physiology, 105*(3), 977-987.

Corder, K., van Sluijs, E. M., McMinn, A. M., Ekelund, U., Cassidy, A., & Griffin, S. J. (2010). Perception versus reality awareness of physical activity levels of British children. *Americal Journal of Preventive Medicine, 38*(1), 1-8.

Corder, K., van Sluijs, E. M., Wright, A., Whincup, P., Wareham, N. J., & Ekelund, U. (2009). Is it possible to assess free-living physical activity and energy expenditure in young people by self-report? *American Journal of Clinical Nutrition, 89*(3), 862-870.

Craig, C. L., Tudor-Locke, C., Cragg, S., & Cameron, C. (2010). Process and treatment of pedometer data collection for youth: the Canadian Physical Activity Levels among Youth study. *Medicine and Science in Sports and Exercise, 42*(3), 430-435.

Croteau, K. A. (2004). A preliminary study on the impact of a pedometer-based intervention on daily steps. *American Journal of Health Promotion, 18*(3), 217-220.

Crouter, S. E., Churilla, J. R., & Bassett, D. R., Jr. (2006a). Estimating energy expenditure using accelerometers. *European Journal of Applied Physiology, 98*(6), 601-612.

Crouter, S. E., Clowers, K. G., & Bassett, D. R., Jr. (2006b). A novel method for using accelerometer data to predict energy expenditure. *Journal of Applied Physiology, 100*(4), 1324-1331.

Crouter, S. E., Schneider, P. L., Karabulut, M., & Bassett, D. R., Jr. (2003). Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. *Medicine and Science in Sports and Exercise, 35*(8), 1455-1460.

CSEP. (2011). Canadian physical activity guidelines for children aged 5-11 years. Retrieved 08/03/2012, 2012, from http://www.csep.ca/CMFiles/Guidelines/CSEP-InfoSheets-child-ENG.pdf

De Puky, P. (1935). The physiological oscillation of the length of the body. *Acta Orthopaedica Scandinavica, 6*, 338-347.

Dinsdale, H., Ridler, C., & Ells, L. (2011). *A simple guide to classifying body mass index in children*. Oxford: National Obesity Observatory.

Dittmar, M., Raschka, C., & Koch, H. J. (2002). Effect of diurnal variation on body composition under consideration of selected chronobiological marker systems. *Anthropological Review, 65*, 17-26.

Dobbins, M., De Corby, K., Robeson, P., Husson, H., & Tirilis, D. (2009). School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6-18. *Cochrane Database of Systematic Reviews*(1), CD007651. doi: 10.1002/14651858.cd007651

DOH. (2008). *Health survey for England 2008, Volume 1, Physical activity and fitness*. The Health and Social Care Information Centre, NHS.

DOH. (2011). Physical activity guidelines for children and young people aged 5-18 years. Retrieved 08/03/2012, 2012, from http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_128144.pdf

Dollman, J., Norton, K., & Norton, L. (2005). Evidence for secular trends in children's physical activity behaviour. *British Journal of Sports Medicine, 39*(12), 892-897.

Duncan, J. S., Schofield, G., & Duncan, E. K. (2006). Pedometer-determined physical activity and body composition in New Zealand children. *Medicine and Science in Sports and Exercise, 38*(8), 1402-1409.

Duncan, J. S., Schofield, G., Duncan, E. K., & Hinckson, E. A. (2007). Effects of age, walking speed, and body composition on pedometer accuracy in children. *Research Quarterly for Exercise and Sport, 78*(5), 420-428.

DuRant, R. H., Baranowski, T., Puhl, J., Rhodes, T., Davis, H., Greaves, K. A., & Thompson, W. O. (1993). Evaluation of the Children's Activity Rating Scale (CARS) in young children. *Medicine and Science in Sports and Exercise, 25*(12), 1415-1421.

Edwardson, C. L., & Gorely, T. (2010). Epoch length and it's effect on physical activity intensity. *Medicine and Science in Sports and Exercise, 42*(5), 928-934.

Ekblom, O., Nyberg, G., Bak, E. E., Ekelund, U., & Marcus, C. (2012). Validity and comparability of a wrist-worn accelerometer in children. *Journal of Physical Activity and Health, 9*(3), 389-393.

Ekelund, U., Aman, J., & Westerterp, K. (2003). Is the ArteACC index a valid indicator of free-living physical activity in adolescents? *Obesity Research, 11*(6), 793-801.

Ekelund, U., Luan, J. a., Sherar, L. B., Esliger, D. W., Griew, P., & Cooper, A. (2012). Moderate to vigorous physical activity and sedentary time and cardiometabolic risk factors in children and adolescents. *JAMA: The Journal of the American Medical Association, 307*(7), 704-712.

Ekelund, U., Tomkinson, G., & Armstrong, N. (2011). What proportion of youth are physically active? measurement issues, levels and recent time trends. *British Journal of Sports Medicine, 45*(11), 859-865.

Esliger, D. W., Copeland, J. L., Barnes, J. D., & Tremblay, M. S. (2005). Standardising and optimizing the use of accelerometer data for free-living physical activity monitoring. *Journal of Physical Activity and Health, 3*, 366-383.

Esliger, D. W., Rowlands, A. V., Hurst, T. L., Catt, M., Murray, P., & Eston, R. G. (2011). Validation of the GENEA Accelerometer. *Medicine and Science in Sports and Exercise, 43*(6), 1085-1093.

Esliger, D. W., & Tremblay, M. S. (2006). Technical reliability assessment of three accelerometer models in a mechanical setup. *Medicine and Science in Sports and Exercise, 38*(12), 2173-2181.

Esliger, D. W., & Tremblay, M. S. (2007). Physical activity and inactivity profiling: the next generation. *Canadian Journal of Public Health, 98* (Suppl 2), S195-207.

Esliger, D. W., Tremblay, M. S., Copeland, J. L., Barnes, J. D., Huntington, G. E., & Bassett, D. R., Jr. (2010). Physical activity profile of Old Order Amish, Mennonite, and contemporary children. *Medicine and Science in Sports and Exercise, 42*(2), 296-303.

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*(1), 362-371.

Fairweather, S. C., Reilly, J. J., Grant, S., Whittaker, A., & Paton, J. Y. (1999). Using the computer science and applications (CSA) activity monitor in preschool children. *Pediatric Exercise Science, 11*, 413-420.

Field, A. (2009). *Discovering statistics using SPSS* (Third ed.). London: SAGE Publications.

Finn, K. J., & Specker, B. (2000). Comparison of actiwatch activity monitor and children's activity rating scale in children. *Medicine and Science in Sports and Exercise, 32*(10), 1794-1797.

Freedman, D. S., Wang, J., Maynard, L. M., Thornton, J. C., Mei, Z., Pierson, R. N., . . . Horlick, M. (2005). Relation of BMI to fat and fat-free mass among children and adolescents. *International Journal of Obesity, 29*(1), 1-8.

Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of accelerometer output for children. *Medicine and Science in Sports and Exercise, 37*(11 Suppl), S523-530.

Gardner, P. J., & Campagna, P. D. (2011). Pedometers as Measurement Tools and Motivational Devices: New Insights for Researchers and Practitioners. *Health Promotion Practice, 12*(1), 55-62.

Godfrey, A., Conway, R., Meagher, D., & OLaighin, G. (2008). Direct measurement of human movement by accelerometry. *Medical Engineering and Physics, 30*(10), 1364-1386.

Goldfield, G. S., Kalakanis, L. E., Ernst, M. M., & Epstein, L. H. (2000). Open-loop feedback to increase physical activity in obese children. *International Journal of Obesity and Related Metabolic Disorders, 24*(7), 888-892.

Goldfield, G. S., Mallory, R., Parker, T., Cunningham, T., Legg, C., Lumb, A., . . . Adamo, K. B. (2006). Effects of open-loop feedback on physical activity and television viewing in overweight and obese children: a randomized, controlled trial. *Pediatrics, 118*(1), e157-166.

Goran, M. I. (1994). Application of the doubly labeled water technique for studying total energy expenditure in young children: a review. *Pediatric Exercise Science, 6*, 11-30.

Goran, M. I., Peters, E. J., Herndon, D. N., & Wolfe, R. R. (1990). Total energy expenditure in burned children using the doubly labeled water technique. *American Journal of Physiology, 259*(4 Pt 1), E576-585.

Goto, R., & Mascie-Taylor, C. G. (2007). Precision of measurement as a component of human variation. *Journal of Physiological Anthropology, 26*(2), 253-256.

Graves, L. E., Ridgers, N. D., Atkinson, G., & Stratton, G. (2010). The effect of active video gaming on children's physical activity, behavior preferences and body composition. *Pediatric Exercise Science, 22*(4), 535-546.

Guinhouya, B. C., Hubert, H., & Zitouni, D. (2011). Need for unbiased computation of the moderate-intensity physical activity of youth in epidemiologic studies. *American Journal of Preventive Medicine, 41*(1), e1-2.

Guinhouya, B. C., Lemdani, M., Vilhelm, C., Durocher, A., & Hubert, H. (2009). Actigraph-defined moderate-to-vigorous physical activity cut-off points among children: statistical and biobehavioural relevance. *Acta Paediatrica, 98*(4), 708-714.

Ham, S. A., Martin, S., & Kohl, H. W., 3rd. (2008). Changes in the percentage of students who walk or bike to school-United States, 1969 and 2001. *Journal of Physical Activity and Health, 5*(2), 205-215.

Hands, B., Parker, H., & Larkin, D. (2006). Physical activity measurement methods for young children: a comparative study. *Measurement in Physical Education and Exercise Science, 10*(3), 203-214.

Hardman, C. A., Horne, P. J., & Fergus Lowe, C. (2011). Effects of rewards, peer-modelling and pedometer targets on children's physical activity: a school-based intervention study. *Psychology and Health, 26*(1), 3-21.

Harrell, J. S., McMurray, R. G., Baggett, C. D., Pennell, M. L., Pearce, P. F., & Bangdiwala, S. I. (2005). Energy costs of physical activities in children and adolescents. *Medicine and Science in Sports and Exercise, 37*(2), 329-336.

Harris, K. C., Kuramoto, L. K., Schulzer, M., & Retallack, J. E. (2009). Effect of school-based physical activity interventions on body mass index in children: a meta-analysis. *Canadian Medical Association Journal, 180*(7), 719-726.

Heath, G. W., Parra, D. C., Sarmiento, O. L., Andersen, L. B., Owen, N., Goenka, S., . . . Brownson, R. C. (2012). Evidence-based intervention in physical activity: lessons from around the world. *The Lancet, 380*(9838), 272-281.

Horne, P. J., Hardman, C. A., Lowe, C. F., & Rowlands, A. V. (2009). Increasing children's physical activity: a peer modelling, rewards and pedometer-based intervention. *European Journal of Clinical Nutrition, 63*(2), 191-198.

Howley, E. T. (2001). Type of activity: resistance, aerobic and leisure versus occupational physical activity. *Medicine and Science in Sports and Exercise, 33*(6), S364-S369.

Jago, R. (2011). Commentary: Age-related decline in physical activity during adolescence - an opportunity to reflect on intervention design and key research gaps. *International Journal of Epidemiology, 40*(3), 699-700.

Jago, R., Fox, K. R., Page, A. S., Brockman, R., & Thompson, J. L. (2010). Physical activity and sedentary behaviour typologies of 10-11 year olds. *International Journal of Behavioral Nutrition and Physical Activity, 7*, 59. doi: 10.1186/1479-5868-7-59

Jago, R., Watson, K., Baranowski, T., Zakeri, I., Yoo, S., Baranowski, J., & Conry, K. (2006). Pedometer reliability, validity and daily activity targets among 10- to 15-year-old boys. *Journal of Sports Sciences, 24*(3), 241-251.

Jago, R., Zakeri, I., Baranowski, T., & Watson, K. (2007). Decision boundaries and receiver operating characteristic curves: new methods for determining accelerometer cutpoints. *Journal of Sports Sciences, 25*(8), 937-944.

Janssen, I. (2007). Physical activity guidelines for children and youth. *Canadian Journal of Public Health, 98* (Suppl 2), S109-121.

Janssen, I., & Leblanc, A. G. (2010). Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *International Journal of Behavioral Nutrition and Physical Activity, 7*, 40. doi: 10.1186/1479-5868-7-40

Jimmy, G., Dossegger, A., Seiler, R., & Mader, U. (2012). Metabolic thresholds and validated accelerometer cutoff points for the actigraph gt1m in young children based on measurements of locomotion and play activities. *Measurement in Physical Education & Exercise Science, 16*, 23-40.

Kang, M., & Brinthaupt, T. M. (2009). Effects of group and individual-based step goals on children's physical activity levels in school. *Pediatric Exercise Science, 21*(2), 148-158.

Kang, M., Marshall, S. J., Barreira, T. V., & Lee, J. O. (2009a). Effect of pedometer-based physical activity interventions: a meta-analysis. *Research Quarterly for Exercise and Sport, 80*(3), 648-655.

Kang, M., Rowe, D. A., Barreira, T. V., Robinson, T. S., & Mahar, M. T. (2009b). Individual information-centered approach for handling physical activity missing data. *Research Quarterly for Exercise and Sport, 80*(2), 131-137.

Kesaniemi, Y. K., Danforth, E., Jr., Jensen, M. D., Kopelman, P. G., Lefebvre, P., & Reeder, B. A. (2001). Dose-response issues concerning physical activity and health: an evidence-based symposium. *Medicine and Science in Sports and Exercise, 33*(6 Suppl), S351-358.

Kilanowski, C. K., Consalvi, A. R., & Epstein, L. H. (1999). Validation of an electronic pedometer for measurement of physical activity in children. *Pediatric Exercise Science, 11*, 63-68.

Kimm, S. Y., Glynn, N. W., Kriska, A. M., Barton, B. A., Kronsberg, S. S., Daniels, S. R., . . . Liu, K. (2002). Decline in physical activity in black girls and white girls during adolescence. *New England Journal of Medicine, 347*(10), 709-715.

Knuth, A. G., & Hallal, P. C. (2009). Temporal trends in physical activity: a systematic review. *Journal of Physical Activity and Health, 6*(5), 548-559.

Krasnoff, J. B., Kohn, M. A., Choy, F. K., Doyle, J., Johansen, K., & Painter, P. L. (2008). Interunit and intraunit reliability of the RT3 triaxial accelerometer. *Journal of Physical Activity and Health, 5*(4), 527-538.

Kriemler, S., Meyer, U., Martin, E., van Sluijs, E. M., Andersen, L. B., & Martin, B. W. (2011). Effect of school-based interventions on physical activity and fitness in children and adolescents: a review of reviews and systematic update. *British Journal of Sports Medicine, 45*(11), 923-930.

Kumahara, H., Tanaka, H., & Schutz, Y. (2004). Daily physical activity assessment: what is the importance of upper limb movements vs whole body movements? *International Journal of Obesity and Related Metabolic Disorders, 28*(9), 1105-1110.

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*(6 Suppl), S370-378.

Lampl, M. (1992). Further observations on diurnal variation in standing height. *Annals of Human Biology, 19*(1), 87-90.

Lanningham-Foster, L., Foster, R. C., McCrady, S. K., Manohar, C. U., Jensen, T. B., Mitre, N. G., . . . Levine, J. A. (2008). Changing the school environment to increase physical activity in children. *Obesity 16*(8), 1849-1853.

LaPorte, R. E., Montoye, H. J., & Caspersen, C. J. (1985). Assessment of physical activity in epidemiologic research: problems and prospects. *Public Health Reports, 100*(2), 131-146.

LeBlanc, A. G., & Janssen, I. (2010). Difference between self-reported and accelerometer measured moderate-to-vigorous physical activity in youth. *Pediatric Exercise Science, 22*(4), 523-534.

Livingstone, M. B., Coward, W. A., Prentice, A. M., Davies, P. S., Strain, J. J., McKenna, P. G., . . . Kerr, M. J. (1992). Daily energy expenditure in free-living children: comparison of heart-rate monitoring with the doubly labeled water (2H2(18)O) method. *American Journal of Clinical Nutrition, 56*(2), 343-352.

Loprinzi, P. D., & Cardinal, B. J. (2011). Measuring children's physical activity and sedentary behaviors. *Journal of Exercise Science and Fitness, 9*(1), 15-23.

Lubans, D., & Morgan, P. (2008). Evaluation of an extra-curricular school sport programme promoting lifestyle and lifetime activity for adolescents. *Journal of Sports Sciences, 26*(5), 519-529.

Lubans, D. R., Foster, C., & Biddle, S. J. (2008). A review of mediators of behavior in interventions to promote physical activity among children and adolescents. *Preventive Medicine, 47*(5), 463-470.

Lubans, D. R., Morgan, P. J., & Callister, R. (2012). Potential moderators and mediators of intervention effects in an obesity prevention program for adolescent boys from disadvantaged schools. *Journal of Science and Medicine in Sport, 15*(6), 519-525.

Lubans, D. R., Morgan, P. J., Callister, R., & Collins, C. E. (2009a). Effects of integrating pedometers, parental materials, and e-mail support within an extracurricular school sport intervention. *Journal of Adolescent Health, 44*(2), 176-183.

Lubans, D. R., Morgan, P. J., & Tudor-Locke, C. (2009b). A systematic review of studies using pedometers to promote physical activity among youth. *Preventive Medicine, 48*(4), 307-315.

Lutes, L. D., & Steinbaugh, E. K. (2010). Theoretical models for pedometer use in physical activity interventions. *Physical Therapy Reviews, 15*(3), 143-153.

Mattocks, C., Ness, A., Leary, S., Tilling, K., Blair, S. N., Shield, J., . . . Riddoch, C. (2008a). Use of accelerometers in a large field-based study of children: protocols, design issues, and effects on precision. *Journal of Physical Activity and Health, 5* (Suppl 1), S98-111.

Mattocks, C., Tilling, K., Ness, A., & Riddoch, C. (2008b). Improvements in the measurement of physical activity in childhood obesity research; lessons from large studies of accelerometers. *Clinical Medicine Insights: Pediatrics, 2*, 1-10.

McClain, J. J., Abraham, T. L., Brusseau, T. A., Jr., & Tudor-Locke, C. (2008). Epoch length and accelerometer outputs in children: comparison to direct observation. *Medicine and Science in Sports and Exercise, 40*(12), 2080-2087.

McClain, J. J., Sisson, S. B., & Tudor-Locke, C. (2007). Actigraph accelerometer interinstrument reliability during free-living in adults. *Medicine and Science in Sports and Exercise, 39*(9), 1509-1514.

McClain, J. J., & Tudor-Locke, C. (2009). Objective monitoring of physical activity in children: considerations for instrument selection. *Journal of Science and Medicine in Sport, 12*(5), 526-533.

McKenzie, T. L. (2002). The use of direct observation to assess physical activity. In G. Welk (Ed.), *Physical activity assessments for health-related research* (pp. 179-195). Champaign, IL: Human Kinetics.

McNamara, E., Hudson, Z., & Taylor, S. J. (2010). Measuring activity levels of young people: the validity of pedometers. *British Medical Bulletin, 95*, 121-137.

Metcalf, B. S., Curnow, J. S., Evans, C., Voss, L. D., & Wilkin, T. J. (2002). Technical reliability of the CSA activity monitor: The EarlyBird Study. *Medicine and Science in Sports and Exercise, 34*(9), 1533-1537.

Metcalf, B. S., Hosking, J., Jeffery, A. N., Voss, L. D., Henley, W., & Wilkin, T. J. (2011). Fatness leads to inactivity, but inactivity does not lead to fatness: a longitudinal study in children (EarlyBird 45). *Archives of Disease in Childhood, 96*(10), 942-947.

Michie, S., Abraham, C., Whittington, C., McAteer, J., & Gupta, S. (2009a). Effective techniques in healthy eating and physical activity interventions: a meta-regression. *Health Psychology, 28*(6), 690-701.

Michie, S., Fixsen, D., Grimshaw, J. M., & Eccles, M. P. (2009b). Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implementation Science, 4*, 40. doi: 10.1186/1748-5908-4-40

Mirwald, R. L., Baxter-Jones, A. D., Bailey, D. A., & Beunen, G. P. (2002). An assessment of maturity from anthropometric measurements. *Medicine and Science in Sports and Exercise, 34*(4), 689-694.

Mitchell, J. A., Pate, R. R., & Blair, S. N. (2012). Screen-based sedentary behavior and cardiorespiratory fitness from age 11 to 13. *Medicine and Science in Sports and Exercise, 44*(7), 1302-1309. doi: 10.1249/MSS.0b013e318247cd73.

Morris, J. N., Heady, J. A., Raffle, P. A., Roberts, C. G., & Parks, J. W. (1953). Coronary heart-disease and physical activity of work. *Lancet, 265*(6795), 1053-1057.

Nader, P. R., Bradley, R. H., Houts, R. M., McRitchie, S. L., & O'Brien, M. (2008). Moderate-to-vigorous physical activity from ages 9 to 15 years. *JAMA : The Journal of the American Medical Association, 300*(3), 295-305.

Nakae, S., Oshima, Y., & Ishii, K. (2008). Accuracy of spring-levered and piezo-electric pedometers in primary school japanese children. *Journal of Physiological Anthropology, 27*(5), 233-239.

Ness, A. R., Leary, S. D., Mattocks, C., Blair, S. N., Reilly, J. J., Wells, J., . . . Riddoch, C. (2007). Objectively measured physical activity and fat mass in a large cohort of children. *PLoS Medicine, 4*(3), e97. doi: 10.1371/journal.pmed.0040097

Nilsson, A., Ekelund, U., Yngve, A., & Sjostrom, M. (2002). Assessing physical activity among children with accelerometers using different time sampling intervals and placements. *Pedicatric Exercise Science, 14*, 87-96.

NHS Information Centre. (2009). *National Child Measurement Programme: England, 2008/09 School Year.* London, UK: The Health and Social Care Information Centre.

NHS Information Centre. (2010). *National Child Measurement Programme: England, 2009/10 school year*. London, UK: The Health and Social Care Information Centre.

NHS Information Centre. (2011). *National child measurement programme: England, 2010/11 school year*. London, UK: The Health and Social Care Information Centre.

O'Donovan, G., Blazevich, A. J., Boreham, C., Cooper, A. R., Crank, H., Ekelund, U., . . . Stamatakis, E. (2010). The ABC of Physical Activity for Health: a consensus statement from the British Association of Sport and Exercise Sciences. *Journal of Sports Sciences, 28*(6), 573-591.

Oliver, M., Schofield, G., & McEvoy, E. (2006). An integrated curriculum approach to increasing habitual physical activity in children: a feasibility study. *Journal of School Health, 76*(2), 74-79.

Omron. (2003). Omron instruction manual: omron pedometer model hj-109. Retrieved Jan 20 2010, 2010, from http://www.omron-healthcare.com.sg/pdf/HJ-109.pdf

Owen, C. G., Nightingale, C. M., Rudnicka, A. R., Cook, D. G., Ekelund, U., & Whincup, P. H. (2009). Ethnic and gender differences in physical activity levels among 9-10-year-old children of white European, South Asian and African-Caribbean origin: the Child Heart Health Study in England (CHASE Study). *International Journal of Epidemiology, 38*(4), 1082-1093.

Ozdoba, R., Corbin, C., & Le Masurier, G. (2004). Does reactivity exist in children when measuring activity levels with unsealed pedometers? *Pediatric Exercise Science, 16*, 158-166.

Paffenbarger, R. S., Hyde, R., Wing, A. L., & Hsieh, C.-c. (1986). Physical Activity, All-Cause Mortality, and Longevity of College Alumni. *New England Journal of Medicine, 314*(10), 605-613.

Paffenbarger, R. S., Laughlin, M. E., Gima, A. S., & Black, R. A. (1970). Work Activity of Longshoremen as Related to Death from Coronary Heart Disease and Stroke. *New England Journal of Medicine, 282*(20), 1109-1114.

Page, A., Cooper, A. R., Stamatakis, E., Foster, L. J., Crowne, E. C., Sabin, M., & Shield, J. P. H. (2005). Physical activity patterns in nonobese and obese children assessed using minute-by-minute accelerometry. *International Journal of Obesity and Related Metabolic Disorders, 29*(9), 1070-1076.

Pan, H., & Cole, T. (2009). LMSGrowth, a Microsoft Excel add-in to access growth references based on the LMS method Retrieved from http://www.healthforallchildren.com/index.php/shop/product/Software/Gr5yCsMCONpF39hF/0

Park, C. O. (1997). Diurnal variation in lumbar MRI. Correlation between signal intensity, disc height, and disc bulge. *Yonsei Medical Journal, 38*(1), 8-18.

Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., . . . et al. (1995). Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *JAMA : The Journal of the American Medical Association, 273*(5), 402-407.

Penpraze, V., Reilly, J., Maclean, C. M., Montgomery, C., Kelly, L. A., Paton, J. Y., . . . Grant, S. (2006). Monitoring of physical activity in young children: how much is enough? *Pediatric Exercise Science, 18*, 483-491.

Powell, S. M., Jones, D. I., & Rowlands, A. V. (2003). Technical variability of the RT3 accelerometer. *Medicine and Science in Sports and Exercise, 35*(10), 1773-1778.

Powell, S. M., & Rowlands, A. V. (2004). Intermonitor variability of the RT3 accelerometer during typical physical activities. *Medicine and Science in Sports and Exercise, 36*(2), 324-330.

Prins, R. G., van Empelen, P., Beenackers, M. A., Brug, J., & Oenema, A. (2010). Systematic development of the YouRAction program, a computer-tailored physical activity promotion intervention for dutch adolescents, targeting personal motivations and environmental opportunities. *BMC Public Health, 10*, 474. doi: 10.1186/1471-2458-10-474

Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obesity Research, 10*(3), 150-157.

Puyau, M. R., Adolph, A. L., Vohra, F. A., Zakeri, I., & Butte, N. F. (2004). Prediction of activity energy expenditure using accelerometers in children. *Medicine and Science in Sports and Exercise, 36*(9), 1625-1631.

Reilly, J. J., Dorosty, A. R., & Emmett, P. M. (2000). Identification of the obese child: adequacy of the body mass index for clinical practice and epidemiology. *International Journal of Obesity and Related Metabolic Disorders, 24*(12), 1623-1627.

Reilly, J. J., Methven, E., McDowell, Z. C., Hacking, B., Alexander, D., Stewart, L., & Kelnar, C. J. (2003). Health consequences of obesity. *Archives of Disease in Childhood, 88*(9), 748-752.

Reilly, J. J., Penpraze, V., Hislop, J., Davies, G., Grant, S., & Paton, J. Y. (2008). Objective measurement of physical activity and sedentary behaviour: review with new data. *Archives of Disease in Childhood, 93*(7), 614-619.

Riddoch, C. J., Leary, S. D., Ness, A. R., Blair, S. N., Deere, K., Mattocks, C., . . . Tilling, K. (2009). Prospective associations between objective measures of physical activity and fat mass in 12-14 year old children: the Avon Longitudinal Study of Parents and Children (ALSPAC). *British Medical Journal, 339*, b4544.

Riddoch, C. J., Mattocks, C., Deere, K., Saunders, J., Kirkby, J., Tilling, K., . . . Ness, A. R. (2007). Objective measurement of levels and patterns of physical activity. *Archives of Disease in Childhood, 92*(11), 963-969.

Ridgers, N. D., & Fairclough, S. (2011). Assessing free-living physical activity using accelerometry: Practical issues for researchers and practitioners. *European Journal of Sport Science, 11*(3), 205-213.

Ridgers, N. D., Fairclough, S. J., & Stratton, G. (2010). Twelve-month effects of a playground intervention on children's morning and lunchtime recess physical activity levels. *Journal of Physical Activity and Health, 7*(2), 167-175.

Ridley, K., Ainsworth, B. E., & Olds, T. S. (2008). Development of a compendium of energy expenditures for youth. *International Journal of Behavioral Nutrition and Physical Activity, 5*, 45. doi: 10.1186/1479-5868-5-45

Rodearmel, S. J., Wyatt, H. R., Stroebele, N., Smith, S. M., Ogden, L. G., & Hill, J. O. (2007). Small changes in dietary sugar and physical activity as an approach to preventing excessive weight gain: the America on the Move family study. *Pediatrics, 120*(4), e869-879.

Rooney, B. L., Gritt, L. R., Havens, S. J., Mathiason, M. A., & Clough, E. A. (2005). Growing healthy families: family use of pedometers to increase physical activity and slow the rate of obesity. *Wisconsin Medical Journal, 104*(5), 54-60.

Rowland, T. W. (2002). Declining cardiorespiratory fitness in youth: fact or supposition? *Pediatric Exercise Science, 14*, 1-8.

Rowland, T. W. (2007). Promoting physical activity for children's health: rationale and strategies. *Sports Medicine, 37*(11), 929-936.

Rowlands, A. V. (2007). Accelerometer assessment of physical activity in children: an update. *Pediatric Exercise Science, 19*(3), 252-266.

Rowlands, A. V., & Eston, R. G. (2007). The measurement and interpretation of children's physical activity. *Journal of Sports Science and Medicine, 6*, 270-276.

Rowlands, A. V., Pilgrim, E. L., & Eston, R. G. (2009). Seasonal changes in children's physical activity: an examination of group changes, intra-individual variability and consistency in activity pattern across season. *Annals of Human Biology, 36*(4), 363-378.

Rowlands, A. V., Powell, S. M., Humphries, R., & Eston, R. G. (2006). The effect of acceleormeter epoch on physical activity output measures. *Journal of Exercise Science and Fitness, 4*(1), 52-58.

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*(2), 1-14.

Sardinha, L. B., Andersen, L. B., Anderssen, S. A., Quitério, A. L., Ornelas, R., Froberg, K., . . . Ekelund, U. (2008). Objectively measured time spent sedentary is associated with insulin resistance independent of overall and central body fat in 9- to 10-year-old portuguese children. *Diabetes Care, 31*(3), 569-575.

SBRN. (2012). Letter to the editor: standardized use of the terms “sedentary” and “sedentary behaviours”. *Applied Physiology, Nutrition and Metabolism, 37*(3), 540-542.

Schneider, P. L., Crouter, S. E., Lukajic, O., & Bassett, D. R., Jr. (2003). Accuracy and reliability of 10 pedometers for measuring steps over a 400-m walk. *Medicine and Science in Sports and Exercise, 35*(10), 1779-1784.

Schoeller, D. A. (1983). Energy expenditure from doubly labeled water: some fundamental considerations in humans. *American Journal of Clinical Nutrition, 38*(6), 999-1005.

Schoeller, D. A., & van Santen, E. (1982). Measurement of energy expenditure in humans by doubly labeled water method. *Journal of Applied Physiology, 53*(4), 955-959.

Schofield, L., Mummery, W. K., & Schofield, G. (2005). Effects of a controlled pedometer-intervention trial for low-active adolescent girls. *Medicine and Science in Sports and Exercise, 37*(8), 1414-1420.

Schutz, Y., Weinsier, R. L., & Hunter, G. R. (2001). Assessment of free-living physical activity in humans: an overview of currently available and proposed new measures. *Obesity Research, 9*(6), 368-379.

Shephard, R. J. (2003). Limits to the measurement of habitual physical activity by questionnaires. *British Journal of Sports Medicine, 37*(3), 197-206.

Shimon, J. M., & Petlichkoff, L. M. (2009). Impact of pedometer use and self-regulation strategies on junior high school physical education students daily step counts. *Journal of Physical Activity and Health, 6*(2), 178-184.

Sidman, C. L., Corbin, C. B., & Le Masurier, G. (2004). Promoting physical activity among sedentary women using pedometers. *Research Quarterly for Exercise and Sport, 75*(2), 122-129.

Sidman, C. L., Corbin, C. B., & Rhea, M. (2003). An examination of the 10,000 step goal in sedentary women with different baseline step goals. *Women in Sport & Physical Activity Journal 12*(2), 111-126.

Siklar, Z., Sanli, E., Dallar, Y., & Tanyer, G. (2005). Diurnal variation of height in children. *Pediatrics International, 47*(6), 645-648.

Silva, P., Mota, J., Esliger, D., & Welk, G. (2010). Technical reliability assesment of the Actigraph GT1M accelerometer. *Measurement in Physical Education and Exercise Science, 14*, 79-91.

Sirard, J. R., & Pate, R. R. (2001). Physical activity assessment in children and adolescents. *Sports Medicine, 31*(6), 439-454.

Sirard, J. R., Trost, S. G., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2005). Calibration and evaluation of an objective measure of physical activity in preschool children. *Journal of Physical Activity and Health, 3*, 345-357.

Smith, J. D., & Schroeder, C. A. (2010). Pedometer accuracy in elementary school children while walking, skipping, galloping and sliding. *Measurement in Physical Education and Exercise Science, 14*, 92-103.

Southard, D. R., & Southard, B. H. (2006). Promoting physical activity in children with MetaKenkoh. *Clinical and Investigative Medicine, 29*(5), 293-297.

Steele, R. M., van Sluijs, E. M., Sharp, S. J., Landsbaugh, J. R., Ekelund, U., & Griffin, S. J. (2010). An investigation of patterns of children's sedentary and vigorous physical activity throughout the week. *International Journal of Behavioral Nutrition and Physical Activity, 7*, 88. doi: 10.1186/1479-5868-7-88

Stone, M. R., Rowlands, A. V., & Eston, R. G. (2009). Relationships between accelerometer-assessed physical activity and health in children: impact of the activity-intensity classification method. *Journal of Sports Science and Medicine, 8*, 136-143.

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*(9 Suppl), S465-470.

Stratton, G. (1996). Children's heart rates during physical education lessons: a review. *Pediatric Exercise Science, 8*, 215-233.

Stratton, G., Canoy, D., Boddy, L. M., Taylor, S. R., Hackett, A. F., & Buchan, I. E. (2007). Cardiorespiratory fitness and body mass index of 9-11-year-old English children: a serial cross-sectional study from 1998 to 2004. *International Journal of Obesity, 31*(7), 1172-1178.

Strong, J. P., Malcom, G. T., McMahan, C. A., Tracy, R. E., Newman, W. P., Herderick, E. E., . . . Group, f. t. P. D. o. A. i. Y. R. (1999). Prevalence and extent of atherosclerosis in adolescents and young adults. *JAMA: The Journal of the American Medical Association, 281*(8), 727-735.

Strong, W. B., Malina, R. M., Blimkie, C. J., Daniels, S. R., Dishman, R. K., Gutin, B., . . . Trudeau, F. (2005). Evidence based physical activity for school-age youth. *The Journal of Pediatrics, 146*(6), 732-737.

Sturm, R. (2005). Childhood obesity - what we can learn from existing data on societal trends, part 1. *Preventing Chronic Disease, 2*(1), 1-9.

Sweeting, H. N. (2007). Measurement and definitions of obesity in childhood and adolescence: a field guide for the uninitiated. *Nutrition Journal, 6*, 32.

Tillmann, V., & Clayton, P. E. (2001). Diurnal variation in height and the reliability of height measurements using stretched and unstretched techniques in the evaluation of short-term growth. *Annals of Human Biology, 28*(2), 195-206.

Townsend, N., Rutter, H., & Foster, C. (2011). Variations in data collection can influence outcome measures of BMI measuring programmes. *International Journal of Pediatric Obesity, 6*(5-6), 491-498.

Tremblay, M. S., Colley, R. C., Saunders, T. J., Healy, G. N., & Owen, N. (2010). Physiological and health implications of a sedentary lifestyle. *Applied Physiology, Nutrition and Metabolism, 35*(6), 725-740.

Tremblay, M. S., Esliger, D. W., Tremblay, A., & Colley, R. (2007). Incidental movement, lifestyle-embedded activity and sleep: new frontiers in physical activity assessment. *Applied Physiology, Nutrition and Metabolism, 32* (Suppl 2E), S208-217.

Tremblay, M. S., LeBlanc, A. G., Kho, M. E., Saunders, T. J., Larouche, R., Colley, R. C., . . . Gorber, S. C. (2011). Systematic review of sedentary behaviour and health indicators in school-aged children and youth. *International Journal of Behavioral Nutrition and Physical Activity, 8*, 98. doi: 10.1186/1479-5868-8-98

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise, 40*(1), 181-188.

Trost, S., McIver, K., & Pate, R. R. (2005a). Conducting accelerometer-based activity assessments in field-based research. *Medicine and Science in Sports and Exercise, 37*(11 Suppl1), S531-S543.

Trost, S. G. (2007). State of the art reviews: measurement of physical activity in children and adolescents. *American Journal of Lifestyle Medicine, 1*(4), 299-314.

Trost, S. G., Loprinzi, P. D., Moore, R., & Pfeiffer, K. A. (2011). Comparison of accelerometer cut points for predicting activity intensity in youth. *Medicine and Science in Sports and Exercise, 43*(7), 1360-1368.

Trost, S. G., McIver, K. L., & Pate, R. R. (2005b). Conducting accelerometer-based activity assessments in field-based research. *Medicine and Science in Sports and Exercise, 37*(11 Suppl), S531-543.

Trudeau, F., Laurencelle, L., & Shephard, R. J. (2004). Tracking of physical activity from childhood to adulthood. *Medicine and Science in Sports and Exercise, 36*(11), 1937-1943.

Tsiros, M. D., Sinn, N., Brennan, L., Coates, A. M., Walkley, J. W., Petkov, J., . . . Buckley, J. D. (2008). Cognitive behavioral therapy improves diet and body composition in overweight and obese adolescents. *American Journal of Clinical Nutrition, 87*(5), 1134-1140.

Tudor-Locke, C. (2002). Taking steps toward increased physical activity: using pedometers to measure and motivate. *Research Digest, 3*(17), 1-8.

Tudor-Locke, C., Bell, R. C., Myers, A. M., Harris, S. B., Ecclestone, N. A., Lauzon, N., & Rodger, N. W. (2004). Controlled outcome evaluation of the First Step Program: a daily physical activity intervention for individuals with type II diabetes. *International Journal of Obesity and Related Metabolic Disorders, 28*(1), 113-119.

Tudor-Locke, C., Craig, C. L., Beets, M. W., Belton, S., Cardon, G. M., Duncan, S., . . . Blair, S. N. (2011). How many steps/day are enough? for children and adolescents. *International Journal of Behavioral Nutrition and Physical Activity, 8*, 78. doi: 10.1186/1479-5868-8-78

Tudor-Locke, C., & Lutes, L. (2009). Why do pedometers work?: a reflection upon the factors related to successfully increasing physical activity. *Sports Medicine, 39*(12), 981-993.

Tudor-Locke, C., Myers, A. M., & Rodger, N. W. (2001). Development of a theory-based daily activity intervention for individuals with type 2 diabetes. *The Diabetes Educator, 27*, 85-93.

Tudor-Locke, C., Sisson, S. B., Lee, S. M., Craig, C. L., Plotnikoff, R. C., & Bauman, A. (2006). Evaluation of quality of commercial pedometers. *Canadian Journal of Public Health, 97 Suppl 1*, S10-S15.

Ulijaszek, S. J., & Kerr, D. A. (1999). Anthropometric measurement error and the assessment of nutritional status. *British Journal of Nutrition, 82*(3), 165-177.

van Cauwenberghe, E., Labarque, V., Trost, S. G., de Bourdeaudhuij, I., & Cardon, G. (2011). Calibration and comparison of accelerometer cut points in preschool children. *International Journal of Pediatric Obesity, 6*(2-2), e582-589.

Van Der Horst, K., Paw, M. J., Twisk, J. W., & Van Mechelen, W. (2007). A brief review on correlates of physical activity and sedentariness in youth. *Medicine and Science in Sports and Exercise, 39*(8), 1241-1250.

van Hees, V. T., Renstrom, F., Wright, A., Gradmark, A., Catt, M., Chen, K. Y., . . . Franks, P. W. (2011). Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. *PLoS One, 6*(7), e22922. doi: 10.1371/journal.pone.0022922

van Sluijs, E., Skidmore, P., Mwanza, K., Jones, A., Callaghan, A., Ekelund, U., . . . Griffin, S. (2008). Physical activity and dietary behaviour in a population-based sample of British 10-year old children: the SPEEDY study (Sport, Physical activity and Eating behaviour: Environmental Determinants in Young people). *BMC Public Health, 8*(1), 388. doi: doi:10.1186/1471-2458-8-388

van Sluijs, E. M., McMinn, A. M., & Griffin, S. J. (2007). Effectiveness of interventions to promote physical activity in children and adolescents: systematic review of controlled trials. *British Medical Journal, 335*(7622), 703. doi: 10.1136/bmj.39320.843947.BE

Vanhelst, J., Theunynck, D., Gottrand, F., & Beghin, L. (2010). Reliability of the RT3 accelerometer for measurement of physical activity in adolescents. *Journal of Sports Sciences, 28*(4), 375-379.

Vincent, S. D., & Pangrazi, R. P. (2002). Does reactivity exist in children when measuring activity levels with pedometers? *Pediatric Exercise Science, 14*, 56-63.

Voss, L. D., & Bailey, B. J. (1997). Diurnal variation in stature: is stretching the answer? *Archives of Disease in Childhood, 77*(4), 319-322.

Voss, L. D., Bailey, B. J., Cumming, K., Wilkin, T. J., & Betts, P. R. (1990). The reliability of height measurement (the Wessex Growth Study). *Arichives of Disease in Childhood, 65*(12), 1340-1344.

Ward, D. S., Evenson, K. R., Vaughn, A., Rodgers, A. B., & Troiano, R. P. (2005). Accelerometer use in physical activity: best practices and research recommendations. *Medicine and Science in Sports and Exercise, 37*(11 Suppl), S582-588.

Welk, G. (2005). Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Medicine and Science in Sports and Exercise, 37*(11 Suppl), S501-511.

Welk, G., Corbin, C. B., & Dale, D. (2000). Measurement issues in the assessment of physical activity in children. *Research Quarterly for Exercise and Sport, 71*(2), 59-73.

Westerterp, K. R. (1999). Physical activity assessment with accelerometers. *International Journal of Obesity and Related Metabolic Disorders, 23 Suppl 3*, S45-49.

Whitehouse, R. H., Tanner, J. M., & Healy, M. J. (1974). Diurnal variation in stature and sitting height in 12-14-year-old boys. *Annals of Human Biology, 1*(1), 103-106.

WHO. (2010). Global reccomendations on physical activity for health. Retrieved 08/03/2012, 2012, from http://www.who.int/dietphysicalactivity/factsheet_young_people/en/index.html

Yngve, A., Nilsson, A., Sjostrom, M., & Ekelund, U. (2003). Effect of monitor placement and of activity setting on the MTI accelerometer output. *Medicine and Science in Sports and Exercise, 35*(2), 320-326.

Zizzi, S., Vitullo, E., Rye, J., O'Hara-Tompkins, N., Abildso, C., Fisher, B., & Bartlett, M. (2006). Impact of a Three-week Pedometer Intervention on High School Students' Daily Step Counts and Perceptions of Physical Activity. *American Journal of Health Education, 37*(1), 35-40.

11 Chapter Eleven – Appendices (see separate document)

Body Mass Index and Accelerometer Measurement Issues for use in the Evaluation of Pedometer-based Physical Activity Interventions in Children

<u>Appendices</u>

Ashley Charles Routen

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

July 2013

University of Worcester

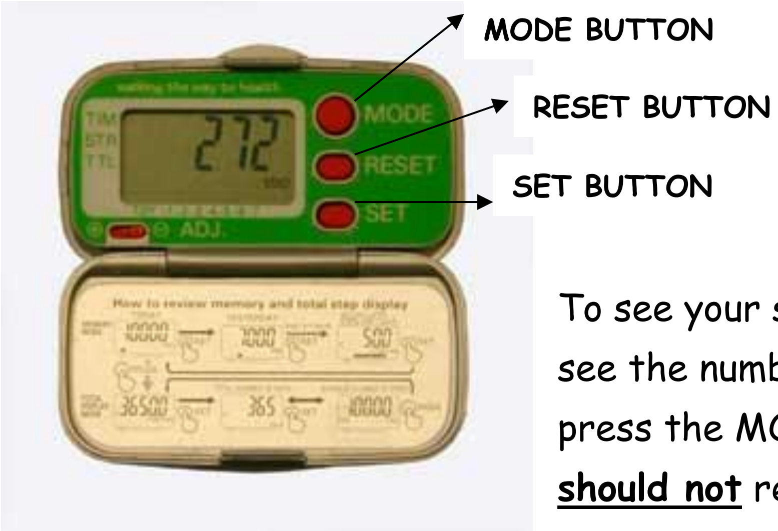
Institute of Sport and Exercise Science

Table of Contents

1 Appendix One - Training courses attended

- ✓ **PGCert in Research Methods** (merit) awarded September 2009, University of Worcester. This included completion of modules:
  - **RTP401:** Processes and Skills, Management and Methods
  - **RTP402:** Dissemination, Engagement and Impact

2 Appendix Two- Pedometer instruction sheets



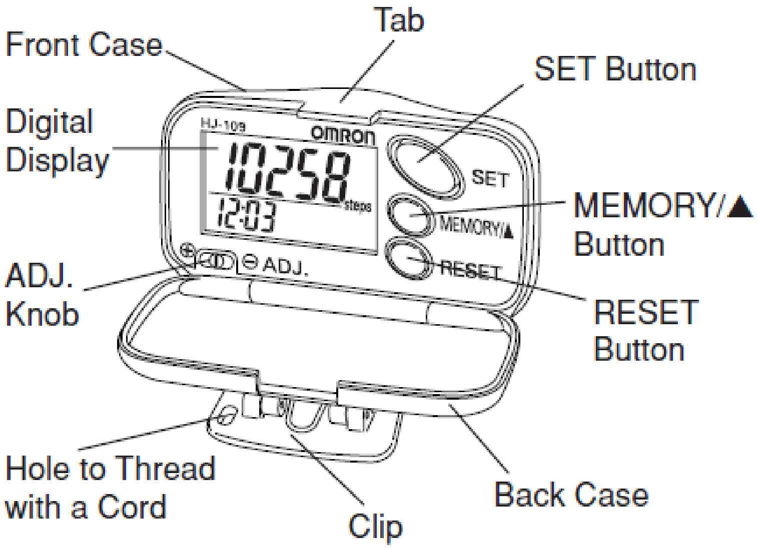
PEDOMETER INSTRUCTION SHEET

To see your step count at the end of the day press the **MODE button** until you see the number of steps, as shown in the screen on the left e.g. 272 steps. If you press the MODE button again it will read the distance walked e.g. 0.27 mile, you **should not** record this in your diary!

Each morning press and hold the **RESET button** to clear the previous days step counts. This will return the counter to 0.

If you press the **MODE button** until a black square appears in the bottom left of the screen, you can see how many steps you have done in previous days (memory function). Press the **SET button** and move the black square along the bottom of the screen to see one day before, two days etc. The days are displayed as TDY (today) 1,2,3,4,5,6,7 in green writing below the screen.

PEDOMETER INSTRUCTION SHEET



To see your step count at the end of the day press the **SET button**. You should see the number of steps, as shown in the screen on the left e.g. 10,258 steps. If you press the SET button again it will read 'aerobic steps' e.g. 100 aerobic steps, you **should not** record this in your diary!

Each morning press and hold the **RESET button** to clear the previous days step counts. This will return the counter to 0.

If you press the MEMORY button once you can see how many steps you did yesterday e.g. 10,000 steps one day before. If you press the MEMORY button again, it will show you two days before.

3 Appendix Three- Pedometer step diary

Your <u>daily step goal</u> this week is:……………………..steps per day

Please record the number of steps you take at the end of each day before you go to bed.

Please fill in the day (e.g. Monday), the date (e.g. 1/1/2011), and the number of steps you have taken according to your pedometer (e.g. 8,000). We would also like you put a tick in the goal achieved column if you managed to do the number of steps set in your daily step goal above.

Below is an example of how to fill your diary in!

Day	**Date**	**Steps**	**Goal Achieved?**
Monday	1/1/2011	7000	
Tuesday	2/1/2011	9110	✔

Office use only Group: 1 ☐ 2☐ 3☐

Complete this section below

⇩

Day	**Date**	**Steps**	**Goal Achieved?**

4 Appendix Four - Additional research dissemination arising from PhD studentship

Routen, A.C. (2011, November). *Accelerometer-based physical activity monitoring in children*. Invited speaker at the 4th Annual International Conference of Physical Education: Sport & Health, University of Pitesti, Romania.

Routen, A.C. (2011). Should our children be sitting comfortably in school? *British Medical Journal*, *343*(7813).

Routen, A.C., Edwards, M.G., Upton, D., & Peters, D.M. (2009, September). *A preliminary investigation of child, parent and programme leader reflections on participation and delivery of a family-based weight intervention programme*. Poster presentation at the BASES Annual Conference, University of Leeds, UK.

Routen, A.C., Edwards, M.G., Upton, D., & Peters, D.M. (2009, April). *A preliminary investigation of child, parent and programme leader reflections on participation and delivery of a family-based weight intervention programme*. Poster presentation at the 'From Cell to Whole Body: The Benefits of Exercise' Conference, University of Birmingham, UK.

5 Appendix Five – Published manuscripts arising from thesis