Understanding and validating accelerometry as a measure of physical activity in children with intellectual disabilities

Arlene Marie McGarty
B.A. (Hons.)

Thesis submitted for the degree of
Doctor of Philosophy

Institute of Health and Wellbeing
College of Medical, Veterinary and Life Sciences

December, 2015

Abstract

Background

Physical activity in children with intellectual disabilities is a neglected area of study, which is most apparent in relation to physical activity measurement research. Although objective measures, specifically accelerometers, are widely used in research involving children with intellectual disabilities, existing research is based on measurement methods and data interpretation techniques generalised from typically developing children. However, due to physiological and biomechanical differences between these populations, questions have been raised in the existing literature on the validity of generalising data interpretation techniques from typically developing children to children with intellectual disabilities. Therefore, there is a need to conduct population-specific measurement research for children with intellectual disabilities and develop valid methods to interpret accelerometer data, which will increase our understanding of physical activity in this population.

Methods

Study 1: A systematic review was initially conducted to increase the knowledge base on how accelerometers were used within existing physical activity research involving children with intellectual disabilities and to identify important areas for future research. A systematic search strategy was used to identify relevant articles which used accelerometry-based monitors to quantify activity levels in ambulatory children with intellectual disabilities. Based on best practice guidelines, a novel form was developed to extract data based on 17 research components of accelerometer use. Accelerometer use in relation to best practice guidelines was calculated using percentage scores on a study-by-study and component-by-component basis.

Study 2: To investigate the effect of data interpretation methods on the estimation of physical activity intensity in children with intellectual disabilities,

a secondary data analysis was conducted. Nine existing sets of child-specific ActiGraph intensity cut points were applied to accelerometer data collected from 10 children with intellectual disabilities during an activity session. Four one-way repeated measures ANOVAs were used to examine differences in estimated time spent in sedentary, moderate, vigorous, and moderate to vigorous intensity activity. Post-hoc pairwise comparisons with Bonferroni adjustments were additionally used to identify where significant differences occurred.

Study 3: The feasibility on a laboratory-based calibration protocol developed for typically developing children was investigated in children with intellectual disabilities. Specifically, the feasibility of activities, measurements, and recruitment was investigated. Five children with intellectual disabilities and five typically developing children participated in 14 treadmill-based and free-living activities. In addition, resting energy expenditure was measured and a treadmill-based graded exercise test was used to assess cardiorespiratory fitness. Breath-by-breath respiratory gas exchange and accelerometry were continually measured during all activities. Feasibility was assessed using observations, activity completion rates, and respiratory data.

Study 4: Thirty-six children with intellectual disabilities participated in a semi-structured school-based physical activity session to calibrate accelerometry for the estimation of physical activity intensity. Participants wore a hip-mounted ActiGraph wGT3X+ accelerometer, with direct observation (SOFIT) used as the criterion measure. Receiver operating characteristic curve analyses were conducted to determine the optimal accelerometer cut points for sedentary, moderate, and vigorous intensity physical activity.

Study 5: To cross-validate the calibrated cut points and compare classification accuracy with existing cut points developed in typically developing children, a sub-sample of 14 children with intellectual disabilities who participated in the school-based sessions, as described in Study 4, were included in this study. To examine the validity, classification agreement was investigated between the criterion measure of SOFIT and each set of cut points using sensitivity, specificity, total agreement, and Cohen’s kappa scores.

Results

Study 1: Ten full text articles were included in this review. The percentage of review criteria met ranged from 12%−47%. Various methods of accelerometer use were reported, with most use decisions not based on population-specific research. A lack of measurement research, specifically the calibration/validation of accelerometers for children with intellectual disabilities, is limiting the ability of researchers to make appropriate and valid accelerometer use decisions.

Study 2: The choice of cut points had significant and clinically meaningful effects on the estimation of physical activity intensity and sedentary behaviour. For the 71-minute session, estimations for time spent in each intensity between cut points ranged from: sedentary = 9.50 (± 4.97) to 31.90 (± 6.77) minutes; moderate = 8.10 (± 4.07) to 40.40 (± 5.74) minutes; vigorous = 0.00 (± .00) to 17.40 (± 6.54) minutes; and moderate to vigorous = 8.80 (± 4.64) to 46.50 (± 6.02) minutes.

Study 3: All typically developing participants and one participant with intellectual disabilities completed the protocol. No participant met the maximal criteria for the graded exercise test or attained a steady state during the resting measurements. Limitations were identified with the usability of respiratory gas exchange equipment and the validity of measurements. The school-based recruitment strategy was not effective, with a participation rate of 6%. Therefore, a laboratory-based calibration protocol was not feasible for children with intellectual disabilities.

Study 4: The optimal vertical axis cut points (cpm) were ≤ 507 (sedentary), 1008−2300 (moderate), and ≥ 2301 (vigorous). Sensitivity scores ranged from 81−88%, specificity 81−85%, and AUC .87−.94. The optimal vector magnitude cut points (cpm) were ≤ 1863 (sedentary), ≥ 2610 (moderate) and ≥ 4215 (vigorous). Sensitivity scores ranged from 80−86%, specificity 77−82%, and AUC .86−.92. Therefore, the vertical axis cut points provide a higher level of accuracy in comparison to the vector magnitude cut points.

Study 5: Substantial to excellent classification agreement was found for the calibrated cut points. The calibrated sedentary cut point (κ =.66) provided comparable classification agreement with existing cut points (κ =.55−.67). However, the existing moderate and vigorous cut points demonstrated low sensitivity (0.33−33.33% and 1.33−53.00%, respectively) and disproportionately high specificity (75.44−.98.12% and 94.61−100.00%, respectively), indicating that cut points developed in typically developing children are too high to accurately classify physical activity intensity in children with intellectual disabilities.

Conclusions

The studies reported in this thesis are the first to calibrate and validate accelerometry for the estimation of physical activity intensity in children with intellectual disabilities. In comparison with typically developing children, children with intellectual disabilities require lower cut points for the classification of moderate and vigorous intensity activity. Therefore, generalising existing cut points to children with intellectual disabilities will underestimate physical activity and introduce systematic measurement error, which could be a contributing factor to the low levels of physical activity reported for children with intellectual disabilities in previous research.

Contents

List of Tables

List of Figures

Publications arising from this thesis

Articles

1. McGarty, A. M., Penpraze, V., & Melville, C. A. (2015). Feasibility of a laboratory-based accelerometer calibration protocol for children with intellectual disabilities. *Pilot and Feasibility Studies*, *1*(1), 18.

2. McGarty, A. M., Penpraze, V., & Melville, C. A. (2014). Accelerometer use during field-based physical activity research in children and adolescents with intellectual disabilities: a systematic review. *Research in Developmental Disabilities*, *35*(5), 973-981.

Conference proceedings

1. McGarty, A. M., Penpraze, V., & Melville, C. A. (2015). *Calibration of the ActiGraph wGT3X+ accelerometer for the estimation of physical activity intensity in children with intellectual disabilities.* Paper presented at International Conference on Ambulatory Monitoring of Physical Activity and Movement, Limerick, Ireland.

2. McGarty, A. M., Penpraze, V., & Melville, C. A. (2015). *Feasibility of an accelerometer calibration protocol in children with intellectual disabilities*. Paper presented at International Society of Behavioral Nutrition and Physical Activity, Edinburgh, UK.

Acknowledgements

Firstly, I would like to give a huge thank you to the children who participated in this research. Not only did they make this thesis possible, but they brought a very welcomed amount of fun and nonsense into the last three years, which I am most grateful for. Thank you also to the parents and teachers who assisted with participation.

I owe a great deal of thanks to Dr. Craig Melville for all his time and efforts throughout this project. Craig has been a constant source of support and I would particularly like to acknowledge and thank him for the kind and patient manner with which he approached supervision. I am also very grateful to Viki Penpraze for all her guidance and support, especially during data collection.

Thank you to Antonia Gates for all her administrative support; Leanne Harris and Louise Bleazard for their assistance during data collection; and the team at Active Schools, South Ayrshire, for their help with recruitment.

Finally, my family and friends deserve a massive thank you for reminding me of the world outside of research. Special thanks go to my parents and Niall for their continued support and belief in me – I could not have done this without them.

Authors Declaration

“I hereby declare that I am the sole author of this thesis, except where the assistance of others has been acknowledged.

It has not been submitted in any form for another degree or professional qualification.”

Arlene Marie McGarty

December, 2015.

Abbreviations

AAIDD	American Association on Intellectual and Developmental Disabilities
ASD	autism spectrum disorders
AUC	area under the curve
BMI	body mass index
bpm	beats per minute
CARS	Children's Activity Rating Scale
CPAF	Children's Physical Activity Form
CO_2	carbon dioxide
cpm	counts per minute
CV	coefficient of variation
EPOC	excess post-exercise oxygen consumption
GPS	global positioning system
HR	heart rate
ICC	intraclass correlation coefficient
ICD	International Classification of Diseases
ID	participants with intellectual disabilities
IOA	inter-observer agreement
IQ	intelligence quotient
MET	metabolic equivalent of task
MEMS	Microelectro-Mechanical-System
NHS	National Health Service
OSRAC-P	Observation System for Recording Physical Activity in Children – Preschool Version
PRISMA	preferred reporting items for systematic reviews and meta-analyses
REE	resting energy expenditure
RER	respiratory exchange ratio
ROC	receiver operating characteristic
RQ	research question
SOFIT	System for Observing Fitness Instruction Time
TD	typically developing participants

UK	United Kingdom
USA	United States of America
$\dot{V}CO_2$	carbon dioxide production
$\dot{V}O_2$	oxygen uptake
$\dot{V}O_{2max}$	maximal oxygen uptake
$\dot{V}O_{2peak}$	peak oxygen uptake
WHO	World Health Organization

Chapter 1 – Background

1.1 Overview of this chapter

This chapter will broadly discuss the literature relating to physical activity in children with intellectual disabilities and discuss the three main areas covered within this thesis: physical activity, intellectual disabilities, and the measurement of physical activity. To ensure clarity, definitions will be presented for each of the major themes. As physical activity research relating to children with intellectual disabilities is lacking, previous research conducted in typically developing children will also be discussed to allow a greater understanding of the importance of physical activity and to highlight the need for more high quality research to be conducted in children with intellectual disabilities. This chapter will conclude with discussion on the important role of measurement in physical activity research and provide a rationale for this thesis focussing specifically on accelerometers.

1.2 Physical activity

Physical activity is an integral aspect of human behaviour and has been throughout evolutionary history. Humans are naturally designed to be active, with a hunter-gatherer biological heritage (Astrand, 1994). This has been consistent with our lifestyles for most of history, with physical activity required for survival, and manual labour commonplace until the mid-20th century. For over 99% of Homo sapiens' existence, our lives have been dominated by the outdoors and physical activity (Astrand, 1986). More recently, however, the technological advances which spread through the developed world have altered the lifestyles that people lead. People have become less active, which has led to an increase of diseases associated with physical inactivity. This is in contrast to the medical advances of recent years, which have eradicated many diseases and contributed to increased life-expectancy. As a result, physical inactivity is now regarded as one of the leading causes of worldwide mortality (World Health Organization; WHO, 2009).

Unlike previous centuries where physical activity was necessary for survival, we are faced with the conundrum of people with hunter-gatherer genes living a twenty-first century lifestyle (Biddle, Mutrie, & Gorely, 2015). Physical activity is no longer a necessary aspect of life, but viewed as a lifestyle choice. The discord between the health benefits of physical activity and the current high prevalence of inactivity has made this an important area for research. In the last half-century or so, there has been a surge in research which has aimed to increase our understanding of physical activity. This research has not only focussed on the physiological aspects and health benefits of activity, but it has also aimed to increase our understanding of why people are inactive/active and how this can be used to develop behaviour change interventions to increase overall health at a population level.

Our understanding of physical activity in day-to-day life has been increasing and the construct of physical activity has evolved into the primary focus of many national and international health recommendations. In recent years, there has been a shift within health-related research from a focus on physical fitness to a greater focus on the promotion of day-to-day physical activity. However, this has led to a lack of clarity surrounding the definition of these separate concepts. Previous research has used the terms physical activity and physical fitness interchangeably and, therefore, produced questionable conclusions on physical activity levels based on parameters of physical fitness (Fernhall, 1993; Frey, Stanish, & Temple, 2008). Therefore, it is important to clarify definitions.

1.2.1 Definition of physical activity

To address the lack of clarity about the key concepts relevant to physical activity research, Caspersen, Powell, and Christenson (1985) proposed definitions of physical activity, exercise, and physical fitness. Physical activity generally describes any form of skeletal muscle movement that results in energy expenditure. Caspersen et al. (1985) also describe physical activity as an overarching term which includes the subcategories of exercise and physical fitness. Exercise refers to physical activity that is planned, structured, and repetitive, which is conducted for the improvement or maintenance of physical fitness. Subsequently, physical fitness is an outcome of exercise and refers to a person having sufficient energy to carry out tasks with vigour, and without unnecessary fatigue. Physical fitness can be categorised as health-related fitness, e.g. cardiorespiratory endurance and muscular strength, and skill-related fitness, e.g. speed and power.

This original definition of physical activity has been expanded upon in more recent years and now encompasses active living, recreational activity, sport, exercise, play, and dance (Scottish National Physical Activity Task Force, 2003). However, an important distinction to be made with regards to individuals with disabilities is that physical activity only refers to movements which are voluntary and not involuntary movements associated with certain disabilities (Cervantes & Porretta, 2010). Fundamentally, physical activity contains three major dimensions: behavioural, movement, and energy expenditure dimensions (Mahar & Rowe, 2002). Furthermore, physical activity can be categorised as containing the following sub-dimensions, which equate to total physical activity: frequency, intensity, duration, type, and context (Corder, Ekelund, Steele, Wareham, & Brage, 2008; Mahar & Rowe, 2002).

Physical activity is most commonly defined in relation to energy expenditure and the associated activity intensity, specifically light, moderate, vigorous, or moderate to vigorous intensity. Until more recently, light intensity activity received little research attention, which is primarily due to this intensity not being sufficient to promote increased levels of fitness and deemed not to be health-enhancing (Troiano & Bucher, 2012). However, from a public health perspective, light intensity activity is more favourable than sedentary behaviours. Therefore, light intensity activity could be utilised as a means to transition inactive populations to health-enhancing intensity activity, or as an alternative for elderly or disabled populations who may be at an increased risk from higher intensity activity, thus making it relevant to children with intellectual disabilities (Gando & Muraoke, 2015). Figure 1.1, which is adapted from Biddle et al. (2015), illustrates the activity intensity and energy expenditure continuum.

In addition to understanding physical activity, researchers are becoming increasingly interested in the evolving concept of sedentary behaviour. Sedentary behaviour describes activities and movements which do not increase energy expenditure above a resting rate (approximately ≤ 1.5 metabolic equivalent of task; METs), such as lying down and sitting (Pate, O'Neill, & Lobelo, 2008). Furthermore, it is a separate construct from physical activity and has distinct health effects for children and youth (Biddle, Gorely, Marshall, Murdey, & Cameron, 2004; Chinapaw, Proper, Brug, van Mechelen, & Singh, 2011). Sedentary behaviour therefore needs to be measured independent of physical activity to further our understanding of this type of behaviour and to develop effective methods for behaviour change (Biddle et al., 2004). However, a limitation within past research is the classification of participants who completed low intensity or low levels of physical activity as "sedentary", even though sedentary behaviour was not specifically measured (Paffenbarger, Hyde, Wing, & Hsieh, 1986; Pate et al., 2008). Individuals who are not physically activity, or do not meet physical activity guidelines, should therefore be described as "inactive", if sedentary behaviour was not measured. Furthermore, the emerging importance of sedentary behaviour as a concept independent of physical activity is highlighted by its inclusion in physical activity guidelines.

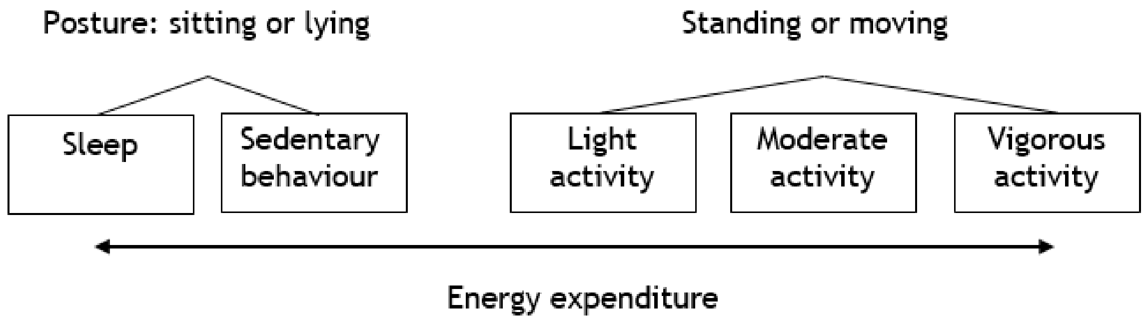


Figure 1.1. Sedentary behaviour and physical activity intensity continuum
Adapted from Biddle et al. (2015)

1.2.2 Physical activity guidelines

One of the primary ways in which physical activity research has real-world impact is by influencing health promotion policy and practice, such as its translation into physical activity guidelines. The promotion of physical activity has increased in the previous two decades, which includes the development of the first consensus physical activity guidelines for children in the United Kingdom (UK) in 1998 (Biddle, Sallis, & Cavill, 1998). These guidelines were originally developed based on expert consensus and review of existing literature to address the need for a public health framework for health-enhancing physical activity specific to children and young people. Until this point, there were conflicting guidelines on the duration and frequency of activity required for positive health outcomes, with activity recommendations for children based on adult literature (Corbin, Pangrazzi, & Welk, 1994; Sallis & Patrick, 1994; U.S. Department of Health and Human Services, 1996).

The original guidelines were developed based on evaluations of results which investigated the effects and relationships between physical activity and various physical and mental health outcomes, such as psychological wellbeing, self-esteem, moral and social development, obesity, and chronic disease risk factors (Cavill, Biddle, & Sallis, 2001). As a result, the following three recommendations were made: 1) all children should participate in one hour of physical activity per day; 2) children who are currently inactive should increase their activity to 30 minutes per day at a moderate intensity; and, 3) activities which enhance or maintain muscular strength, flexibility, and bone health should be conducted twice per week (Biddle et al., 1998). However, it was acknowledged when these guidelines were developed that the strength of the included evidence regarding the relationships between physical activity and various health outcomes was weak and often inconsistent (Riddoch, 1998). That was partially attributed to inappropriate definitions of physical activity and the use of subjective self-report measures, which demonstrate low reliability and criterion validity in children (Cavill et al., 2001; Kohl, Fulton, & Caspersen, 2000). Therefore, there was a need to continue to investigate physical activity in children to increase the evidence-base. Furthermore, this had to be done in conjunction with measurement research aiming to increase the validity and usability of objective, free-living methods of physical activity measurement.

In 2011, the UK physical activity guidelines were updated to reflect the current knowledge on physical activity and positive health outcomes. These guidelines contain three specific recommendations for children aged 5 to 18 years: 1) children should be active at a moderate to vigorous intensity for a minimum of 60 minutes, and up to several hours, each day; 2) children should participate in vigorous intensity activity, and activities that strengthen muscle and bone, at least three times per week; 3) sedentary behaviours, e.g. sitting, should be minimised (Chief Medical Officers, 2011). These guidelines are consistent with physical activity recommendations from other organisations and countries, such as the WHO, United States of America (USA), Canada, and Australia (Tremblay et al., 2011b; WHO, 2010; WHO, 2015). The development of these recommendations was also a result of an expert panel reviewing relevant research (Bull et al., 2010).

Similar to the original guidelines, however, a limitation with the current recommendations is the lack of included evidence that was based on objective physical activity measurements. As subjective measures introduce a higher level of measurement error and bias, the expert panel which developed the guidelines recommend the use of objective and time-stamped measurement methods for future physical activity research (Bull et al., 2010). Furthermore, the expert panel also recommend that a consensus is reached on standardised methods of data cleaning, reduction, and analysis for objective measures. This highlights the important role that physical activity measurement has in the wider dissemination of physical activity research. It also highlights the impact that the limited use of validated objective measures in large-scale epidemiological research, and the lack of consensus on how to deal with objectively measured data, are hindering physical activity research at the highest level. Therefore, increasing the availability of validated objective measures, and producing clear guidelines for using these devices, are important areas for research.

Another limitation of physical activity guidelines is that they are based on research involving healthy populations without disabilities. Therefore, the relevance of these guidelines to other population groups may be limited. Although this was acknowledged in the development of the UK guidelines, the relevance and validity of these guidelines for individuals with disabilities has not been empirically investigated. Therefore, it is important that physical activity behaviours are better understood in populations with disabilities, rather than continuing the trend of generalising findings and recommendations from typically developing populations to populations with disabilities.

1.2.3 Physical activity in typically developing children

Physical inactivity is the fourth leading risk factor for worldwide mortality, accounting for 6% of deaths globally (WHO, 2009). Furthermore, physical inactivity has been consistently shown to have a causal relationship with all-cause mortality (WHO, 2009). The health effects associated with inactivity are so great that it costs the UK National Health Service (NHS) an estimated £1-1.8 billion per year overall, and an estimated £91 million per year for the NHS in Scotland (Chief Medical Officer, 2009).

The prevalence of physical inactivity is surprising, considering evidence on the health benefits of physical activity has long been established. One of the earliest and most notable studies was by Morris, Kagan, Pattison, and Gardiner (1966), which compared the incidence of ischaemic heart disease between sedentary bus drivers and physically active bus conductors and found a significantly higher prevalence in bus drivers (8.5 per 100) compared to conductors (4.7 per 100) over a five year period. Since this study, health outcomes associated with physical activity has been extensively investigated throughout the life course.

Physical activity has subsequently been identified as a determinant of many positive health outcomes, including cardiovascular health (20-30% reduced risk of coronary heart disease and stroke), cancer prevalence (30% and 20% reduced risk for colon and breast cancer, respectively), metabolic health (30%-40% reduced risk of metabolic syndrome and type 2 diabetes), and mental health (20-30% reduced risk of depression, dementia, and anxiety; Chief Medical Officers, 2011). Due to the many health benefits associated with physical activity, it has been described as a “wonder drug” and “miracle cure” (Chief Medical Officer, 2009, pg.1).

A considerable amount of research has been conducted to investigate various aspects of physical activity in typically developing children. Therefore, a full review of this research is outwith the scope of the thesis. A brief overview of this research, however, will highlight the breadth of the existing knowledge-base for typically developing children, which is comparatively lacking for children with intellectual disabilities. The following sections will discuss the current stage of typically developing research, specific to the health benefits of physical activity.

1.2.3.1 Physical activity benefits for typically developing children

Research investigating the benefits of physical activity has been widely conducted in typically developing children. Due to the large volume of existing research, numerous systematic reviews, meta-analyses, and systematic reviews of systematic reviews and meta-analyses have been conducted to consolidate this research. Therefore, this section will discuss relevant systematic reviews and meta-analyses with a focus on the health outcomes and benefits of physical activity in typically developing children.

1.2.3.1.1 Physical health benefits

Observational studies show that physical activity is correlated with many physical health benefits and reduced risk factors in children (Janssen & LeBlanc, 2010). Furthermore, physical activity in childhood has a preventive effect on many factors relating to ill-health in adulthood, such as bone health (Hallal, Victora, Azevedo, & Wells, 2006). One of the most extensively studied areas in observational research is the relationship between physical activity and obesity, with children's physical activity levels showing an inverse relationship with relative weight gain (Must & Tybor, 2005). Furthermore, the strength and consistency of this relationship increases with the intensity and duration of activity, indicating additional health benefits are associated with increased activity (Janssen & LeBlanc, 2010).

However, the observed relationships between physical activity and health benefits vary between outcomes. The evidence relating to the relationship between hypertension, cholesterol, and metabolic syndrome in observational studies of children is generally weak and somewhat limited (Janssen and LeBlanc, 2010). An interesting finding in the review by Janssen and LeBlanc (2010) is the effect that the method used to measure physical activity has on results. The authors highlight that subjective self-/proxy-reports produce weak to moderate relationships with health outcomes, whereas objective measures produce consistently strong, positive relationships. Therefore, the wide use of subjective measures in this area of research is potentially underestimating the strength of relationships between physical activity and positive health outcomes, and limiting our understanding of the amount of activity required to achieve these benefits.

The evidence relating to the benefits of physical activity has additionally been investigated in experimental studies. Similar to observational research, the effect of physical activity on obesity is one of the most widely investigated outcomes, although there are conflicting findings. Reviews by Steinbeck (2001) and Waters et al. (2011) report that increasing physical activity through interventions is effective in preventing obesity in children. Furthermore, increasing physical activity using home- and clinical-based interventions is not only effective in preventing obesity, but it is also effective in reducing the body mass index (BMI) of obese and overweight children, with a summary effect size of −0.36 (95% CI −0.64, −0.08; Ruotsalainen, Kyngäs, Tammelin, & Kääriäinen, 2015).

However, the effects of physical activity on BMI reported in experimental research is affected by environment. In contrast to the effective non-school-based interventions included in the previously discussed reviews, a meta-analysis by Harris, Kuramoto, Schulzer, and Retallack (2009) reports that school-based physical activity interventions are not effective in significantly reducing BMI (weighted mean difference –0.05 kg/m^2; 95% CI –0.19, 0.10). This is concurrent with a Cochrane review which also found no reductions in BMI based on school-based physical activity interventions, although more positive effects were found for experimental groups in comparison with control groups, suggesting a weight maintenance effect (Dobbins, Husson, DeCorby, & LaRocca, 2013). A reason for the effect of environment is that interventions which aim to increase school-based physical activity are generally not effective at increasing activity levels, or do not increase activity levels enough to promote health benefits. Therefore, it is important to be aware of this effect when interpreting findings.

There is also a growing body of evidence suggesting that reducing sedentary behaviours has positive health outcomes in typically developing children. Tremblay et al. (2011a) identified that reducing sedentary time has significant effects on reducing BMI ($\bar{d}$ = −0.81; 95% CI −1.44, −0.17, p = .01), whereas over two hours of sedentary time per day is associated with obesity (Tremblay et al., 2011a). Similarly, a systematic review of randomised controlled trials by Leung, Agaronov, Grytsenko, and Yeh (2011) showed that reducing sedentary time resulted in lower levels of obesity and body composition measurements. On the other hand, Chinapaw et al. (2011) and Marshall, Biddle, Gorely, Cameron, and Murdey (2004) found no clinically meaningful effects of sedentary time on BMI, with Marshall et al. (2004) noting a mean sample-weighted effect size of r = 0.066 (95% CI 0.056, 0.078). However, both these studies focussed on screen-time, which Marshall et al. (2004) suggested did not capture all sedentary time, thus underestimating the effect size.

Increases in physical activity in experimental studies also improve health outcomes relating to blood lipid levels, blood pressure, metabolic syndrome, and bone health, with Janssen and LeBlanc (2010) reporting improved summary effects for triglycerides (−3.03; 95% CI −3.22, −2.84), systolic blood pressure (aerobic = −1.39, 95% CI −2.53, −0.24; non-aerobic = −0.61, 95% CI −2.27, 1.05), diastolic blood pressure (aerobic = −0.39, 95% CI −1.72, 0.93; non-aerobic = −0.51, 95% CI −2.18, 1.06), and fasting insulin (aerobic = −0.60, 95% CI −1.71, 0.50; resistance training = −0.31, 95% CI − 0.82, 0.19). Interestingly, this review also highlights that it is not only the intensity of activity, but also the type of activity, which has an effect on health outcomes.

1.2.3.1.2 Mental health benefits

In addition to physical health benefits, mental health benefits have also been found for physical activity in observational research. The systematic review by Janssen and LeBlanc (2010) reports positive associations for various parameters of mental health, including anxiety, depression, global and physical self-concept, with positive but weak relationships for social and academic self-concept. However, there was insufficient evidence from observational studies to conclude the intensity or duration of activity required for children to gain these benefits, which could be partially attributed to the use of self-report measures. On the other hand, the experimental studies within this review provide some initial evidence that increasing activity can have a positive effect on mental health outcomes. Furthermore, activity at a higher intensity is more effective in significantly reducing depression and stress scores, in comparison with lower intensity activity.

These findings are supported by a meta-analysis conducted by Ahn and Fedewa (2011), which included 73 studies. The authors report that increasing physical activity has medium to large effect sizes on various mental health outcomes, including depression ($\bar{d}$ = − 0.41, SE = 0.13), anxiety ($\bar{d}$ = − 0.35, SE = 0.18), psychological distress ($\bar{d}$ = − 0.61, SE = 0.30), and self-esteem ($\bar{d}$ = 0.29, SE = 0.01). A review of reviews by Biddle and Asare (2011) also reports that physical activity is associated with improved mental health, with sedentary behaviour associated with poorer mental health. However, this study reports that effect sizes are generally weak to moderate; −0.15 to −0.66 for depression, −0.15 to −0.48 for anxiety, and 0.12 to 0.89 for self-esteem.

1.2.3.1.3 Cognitive functioning benefits

An area which has received a greater focus in more recent years is the cognitive functioning benefits of being physically active. Janssen and Le Blanc (2010) report positive associations between physical activity and academic performance, specifically standardised test scores and memory. Concurrent findings are reported in a systematic review by Howie and Pate (2012), which reviewed 125 articles and found many positive effects relating to physical activity and constructs of academic achievement; however, many studies within this review are limited by weak study designs and subjective measures. Therefore, the authors also recommend a future research focus on understanding the intensity and duration of activity required for health outcomes. Again, these finds are concurrent with meta-analyses results, as Sibley and Etnier (2003) report an effect size of 0.32 (SD = 0.27) for increased cognitive function. Furthermore, Fedewa and Ahn (2011) also found that physical activity has a positive effect on children’s cognitive function and academic achievement ($\bar{d}$ = 0.35, SE = 0.04, 95% CI = 0.27, 0.43). Another interesting finding of this study is that children with a higher level of fitness have higher cognitive function ($\bar{d}$ = 0.32, SE = 0.03, 95% CI = 0.26, 0.37), which suggests higher intensity and duration of activity promotes greater health outcomes. It is important to note, however, that there is great debate surrounding the possible direct and indirect mechanisms of these effects.

1.2.3.1.4 Summary of health benefits

There is a large volume of data demonstrating various health benefits of physical activity in typically developing children, such as reducing and preventing obesity, lowering blood pressure, fasting insulin levels, depression and anxiety, and increasing global and physical self-worth, and academic performance.

However, the generalisability of research relating to physical activity and positive health outcomes is somewhat limited as the majority of this research has included samples of high risk children, i.e. children who already have the outcome of interest, such as obesity or metabolic syndromes, thus limiting the generalisability to children with a healthy weight and without chronic health conditions (Strong et al., 2005).

A consistent limitation reported within the discussed reviews is the wide use of subjective measures, such as questionnaires, which further limits our understanding of how the duration and intensity of activity effects health outcomes. This is concurrent with the issues previously discussed in relation to the development of physical activity guidelines, which relied on weak evidence and was limited by subjective measurement methods. Therefore, to increase our understanding of how physical activity affects health, objective measures need to be more widely used to investigate the dose-response relationship.

1.2.3.1.5 Dose-response relationship

The association between increased physical activity and increased health benefits is known as the dose-response relationship, i.e. how the intensity and duration of activity affect positive health outcomes. The previous sections in this chapter on the benefits of physical activity in typically developing children highlights that there is evidence showing increasing the duration and intensity of physical activity is associated with increased physical and mental health benefits, and improved cognitive functioning. Furthermore, there is emerging evidence showing that sedentary time has a negative relationship with obesity, with reductions in sedentary time promoting positive health outcomes. However, a current limitation with this experimental research in children relates to what sedentary behaviour is replaced with, i.e. what intensity of activity, which has implications from a public health perspective. With an increased research focus on light intensity activity, it is important to understand the health benefits of this ‘dose’ of activity. In addition, it is also important to investigate whether replacing sedentary behaviours with light intensity activity has associated health benefits, and whether this is an effective method of increasing physical activity levels at the higher end of the intensity continuum, as illustrated in Figure 1.1 (Gando & Muraoke, 2015).

The relationship between physical activity and reduced risk of disease in typically developing populations has been used for the development of physical activity guidelines and for health promotion (Chief Medical Officers, 2011). However, the development of these guidelines is limited by our lack of understanding of the dose-response relationship. Furthermore, there is insufficient evidence to make recommendations on the volume of activity required to reduce the risk of specific diseases (Bull et al., 2010; Chief Medical Officers, 2011).

The primary reason given for this lack of evidence is the use of subjective methods to measure physical activity. A limitation of subjective measures is that these methods are affected by recall bias and have limited validity for the reporting of intensity and duration of activity (Matthews, 2002). To better understand the dose-response relationship and more accurately inform health promotion, researchers need to be able to accurately and objectively measure the duration and intensity of physical activity to increase our knowledge of the interactions between these physical activity dimensions and health outcomes. Furthermore, when objective measures are used, the methods employed to analyse and interpret data need to be better understood, with standardised methods developed (Bull et al., 2010).

As previously discussed, the development of physical activity guidelines and the understanding of the dose-response relationship is based on data from typically developing children. However, it is important that this relationship is investigated in individuals with disabilities so that the effect of duration and intensity of activity on health outcomes can be better understood. Furthermore, as children with intellectual disabilities are reported to be a sedentary population with complex health needs, the importance of reducing sedentary behaviour and increasing light intensity activity for the improvement of functional fitness may be more important for this population. As a result, targeted health promotion guidelines can be developed and evidence-based interventions designed.

1.2.4 Need to study physical activity in children with intellectual disabilities

Children with intellectual disabilities are a neglected population in physical activity research (Frey et al., 2008). The breath, depth, and overall quality of intellectual disabilities research is limited in comparison with physical activity research conducted in typically developing children. There is a fundamental lack of knowledge within physical activity research involving children with intellectual disabilities, ranging from the basics of valid measurement methods to the design and implementation of effective interventions (Frey et al., 2008; Hinckson & Curtis, 2013). As a result, research is based on measurement methods with questionable validity, uncertain and contradictory conclusions on physical activity levels, and ineffective interventions to increase activity levels. Furthermore, there is a trend in intellectual disabilities research of generalising findings and study designs from research involving people without intellectual disabilities.

Considering that people with intellectual disabilities have a higher prevalence of both physical and mental ill-health in comparison with people without intellectual disabilities, this population group would potentially benefit greatly from increased physical activity (Maiano, 2010). Furthermore, conducting research in children is important as physical activity in childhood is associated with improved health and increased activity levels in adulthood (Cavill et al., 2001; Telama, 2009; Telama et al., 2005). Therefore, introducing active lifestyles in childhood could increase physical activity levels throughout the life course.

For these reasons, it is important to build an evidence-base that will enable physical activity to be accurately measured and facilitate a better understanding of activity behaviours and health benefits in children with intellectual disabilities. As a result, this will aid in the development and implementation of effective interventions and population-specific health promotion guidelines.

1.3 Intellectual disabilities

1.3.1 Definition of intellectual disabilities

In the current version of the International Classification of Diseases (ICD-10), the WHO uses the term “mental retardation” to describe:

“*a condition of arrested or incomplete development of the mind, which is especially characterized by impairment of skills manifested during the developmental period, skills which contribute to the overall level of intelligence, i.e. cognitive, language, motor, and social abilities*” (WHO, 1993, pg. 70).

A similar and widely cited definition is that by the American Association on Intellectual and Developmental Disabilities (AAIDD):

“*significant limitations both in intellectual functioning and in adaptive behavior as expressed in conceptual, social, and practical adaptive skills. This disability originates before age 18*” (Schalock et al., 2010, pg.1)

Although both these definitions have been widely accepted and used within research in previous years, more recently the use of the term “mental retardation” has been questioned (Schalock et al., 2010). Various alternative terms have been used, including intellectual disabilities, mental handicap, mental deficiency, learning disabilities, and developmental disabilities (WHO, 2007). There is currently an international debate surrounding the definition and assessment of mental retardation and its classification within the forthcoming version of the ICD (ICD-11). Although the specifics of the proposed changes are outwith the scope of this thesis, its magnitude highlights the difficulties in establishing an accurate and universally accepted definition and classification criteria (Bertelli et al., 2014; Carulla et al., 2011). As part of these proposed changes, the use of the term “intellectual disabilities” is suggested to describe the functional/disability condition, which was formally mental retardation.

In recent years, there has been a significant increase in the use of the term “intellectual disabilities” over “mental retardation” (Russell, Mammen, & Russell, 2005). This is evident from the number of international organisations and journals which have adopted the term. In contrast to the definition described previously, the WHO now uses the term intellectual disabilities rather than mental retardation in its publications. Furthermore, the AAIDD was renamed in 2007, from the American Association on Mental Retardation, in keeping with the evolving terminology. This change in terminology is primarily important to individuals with disabilities as mental retardation has been described as “offensive to persons with disabilities” (Schalock, Luckasson, & Shogren., 2007, pg. 118). In addition, it also allows a greater level of consistency in research regarding terminology.

The lack of clarity surrounding terminology in research has led to discrepancies surrounding the specific condition each term describes. For example, in the UK the term “learning disabilities” can be used interchangeably with intellectual disabilities. However, in the USA, learning disabilities is not synonymous with intellectual disabilities, as it specifically describes conditions which impact on learning but not intelligence, such as dyslexia. Another commonly used term is “developmental disabilities”, which relates to chronic conditions that cause physical and/or mental impairments, such as autism, Down syndrome, and cerebral palsy. Although this term has previously been used synonymously with intellectual disabilities, an important distinction to be made is that although intellectual disabilities can be categorised as a developmental disability, not all individuals with developmental disabilities will meet the classification criteria for intellectual disabilities.

Currently, the classification of intellectual disabilities is based on three fundamental criteria: 1) impaired intellectual functioning, which is generally measured as an intelligence quotient (IQ) score of < 70, or two standard deviations below the mean; 2) limitations in adaptive behaviour, specifically conceptual, social, and practical skills; 3) the age of onset during the developmental period, i.e. prior to the age of 18 years (McDermott, Durkin, Schupf, Stein, 2007; Schalock et al., 2010; WHO, 2007). Intellectual disabilities are usually classified as mild, moderate, severe, or profound. Specifically, level of intellectual disabilities is generally classified using the IQ criteria described in Table 1.1. In addition, intellectual disabilities can be categorised as “other” if it is not possible to complete the necessary assessment, e.g. due to a severe physical disability, or can be classified as “unspecified” if there is evidence of intellectual disabilities but not enough information to make an accurate classification.

Table 1.1. Level of intellectual disabilities and corresponding IQ range

Level of intellectual disabilities	IQ
Mild	50 – 69
Moderate	35 – 49
Severe	20 – 34
Profound	< 20

Therefore, in keeping with current research and practice, the term “intellectual disabilities” will be used throughout this thesis and will refer to conditions which meet the ICD-10 and AAIDD definitions and classification criteria, as described above. This term will be used when discussing all previous research relating to individuals with intellectual disabilities, regardless of differing terminology used within individual studies. Furthermore, children who do not meet the criteria for intellectual disabilities and have no reported developmental disabilities will be described as “typically developing”.

1.3.2 Causes of intellectual disabilities

There are a number of potential causes of intellectual disabilities. Causes can generally be categorised into genetic abnormalities, biological factors, and environmental factors, which can occur prenatal, perinatal, or postnatal (Guralnick, 2005). Genetic abnormalities are prenatal causes, with examples including Down syndrome and Fragile X syndrome. Perinatal causes occur around the time of birth and include infections (biological factor), low birth weight, and asphyxia during birth (environmental factors). Postnatal causes can occur up to the age of 18 years and include epilepsy (biomedical factor), head injury, and child neglect (environmental factors; Carnaby, 2007; Croen, Grether, & Selvin, 2001).

When considering potential causes for intellectual disabilities, it is important to understand the interactions between biological and environmental factors, and the cumulative effect of exposure to risk factors (Burchinal, Roberts, Hooper, & Zeisel, 2000; Guralnick, 2005). Furthermore, many of the risk factors associated with intellectual disabilities are also factors associated with lower socio-economic status, such as malnutrition, limited access to healthcare, and level of maternal education (Croen et al., 2001; Leonard et al., 2002). Therefore, country/region of residence, socio-economic status, and ethnicity can impact on the prevalence rates of intellectual disabilities. It is important to note, however, that an estimated 30% to 50% of cases of intellectual disabilities are attributed to an unknown cause (Curry et al., 1997).

1.3.3 Prevalence of intellectual disabilities

It has been estimated that the global prevalence of intellectual disabilities ranges between 1% and 3%, although rates as low as 0.16% and has high as 16% have been reported (Harris, 2006; Maulik, Mascarenhas, Mathers, Dua, & Saxena, 2011). The prevalence of mild intellectual disabilities is highest and affects an estimated 85% of people with intellectual disabilities; rates of moderate, severe, and profound are estimated to be approximately 10%, 4%, and 2%, respectively (King, Toth, Hodapp, & Dykens, 2009). Furthermore, the prevalence of intellectual disabilities is generally higher in boys compared to girls (Croen et al., 2001)

As previously discussed, numerous factors have been identified which are thought to affect the prevalence of intellectual disabilities, such as those associated with lower socio-economic status. Therefore, prevalence rates of intellectual disabilities vary greatly between countries, with notable differences found between developed and developing countries (Maulik et al., 2011). For example, Stein, Belmont, and Durkin (1987) report prevalence rates of 15.60%, 6.43%, and 4.03% in Bangladesh, Brazil, and India, respectively. In contrast, prevalence rates of 0.35% were previously reported in both Norway and Canada (Bradley, Thompson, & Bryson, 2002; Stromme, 1998).

However, ascertaining the prevalence of intellectual disabilities poses many challenges; therefore, prevalence rates should be interpreted with caution. Research in this area is limited by a number of factors, including a lack of reliable data collection procedures, research predominantly being conducted in high-income countries, and differing definitions and classifications of intellectual disabilities (Maulik et al., 2011; WHO, 2007).

1.3.4 Physical activity in children with intellectual disabilities

1.3.4.1 Benefits of physical activity for children with intellectual disabilities

Children with intellectual disabilities are at a higher risk from secondary health problems compared to their typically developing peers, such as obesity and its associated risk factors (Maiano, 2010). Therefore, if the positive health effects from physical activity seen in typically developing children are similar in children with intellectual disabilities, then promoting and increasing physical activity could be highly beneficial for this population group. However, in comparison with the breadth and depth of research conducted in typically developing children, research relating to the health benefits of physical activity in children with intellectual disabilities is very limited. Previous research has predominantly focussed on effects of exercise and physical activity interventions. This section will provide an overview of review studies which investigated the health benefits of increased physical activity.

The type of interventions conducted in children with intellectual disabilities are somewhat different to those conducted in typically developing children. A systematic review by Johnson (2009) found that studies generally focus on the effects of exercise and therapeutic activity programmes on health outcomes in children with developmental disabilities, rather than the effects of daily/habitual physical activity. Although, increasing activity in these interventions was effective in improving respiratory function, motor function, muscle strength, and fitness. However, Johnson (2009) concludes that the overall quality of this research evidence is low due to weak study designs and small sample sizes. Furthermore, there is insufficient evidence to make recommendations on the required duration and intensity of activity to promote health benefits.

A more recent meta-analysis of randomised controlled trials by Harris, Hankey, Murray, and Melville (2015) investigates the effect of physical activity on body composition in adolescents and young adults with intellectual disabilities. Similar to the review by Johnson (2009), all studies within this review included exercise training interventions, such as cycle ergometry and plyometrics. However, the meta-analysis results show no significant improvements in weight (−0.17 kg; 95% CI −1.04, 0.72 kg) or measures of body composition, including BMI (−0.07 kg/m^2; 95% CI −0.64, 0.51 kg/m^2) and waist circumference (−1.14 cm; 95% CI −4.03, 1.75 cm), for the experimental group. The authors describe that the dose of physical activity of the interventions was not sufficient to promote positive health outcomes, with small sample sizes also limiting conclusions.

Sibley & Etnier (2003) conducted a meta-analysis investigating the relationship between physical activity and cognition in children. One of the merits of this study was that it did not exclude research conducted in children with intellectual disabilities. This review highlights that physical activity in children with intellectual disabilities has a similarly positive and significant ($p < .05$) relationship with cognition as that seen in typically developing children. Furthermore, the effect size for children with intellectual disabilities (ES = 0.43) is higher than in typically developing children (ES = 0.25). However, only two studies involving children with intellectual disabilities were included in this review, both of which were published in the 1960's, and focussed on the effects of physical education. A more recent review conducted by Howie and Pate (2012) investigates the effects of physical activity and academic achievement in children. Similar to the review by Sibley and Etiner (2003), this review did not exclude studies which included a sample of children with intellectual disabilities. However, of the 125 studies included in the review, only one focussed on children with intellectual disabilities, highlighting the dearth of research in comparison with typically developing children. Furthermore, the included study, which was by Bluechardt and Shepard (1995), investigated self-perceptions of academic competence, rather than academic performance specifically.

A limitation with the structured interventions reported within these reviews is that sustainability is limited post-intervention. Furthermore, these intervention designs do not increase our understanding of daily physical activity and how physical activity conducted outwith the intervention programme effects health outcomes. More recently there has been a focus on increasing daily physical activity through behaviour-change interventions, which have a greater longevity and prolonged benefits compared to organised exercise interventions (Biddle et al., 2015). However, the implementation of daily physical activity interventions is limited in children with intellectual disabilities. Hinckson, Dickinson, Water, Sands, and Penman (2013) conducted a complex 10-week physical activity and nutrition intervention in 17 children with autism and intellectual disabilities, which included physical activity and education elements, and aimed to increase daily physical activity. However, there were no beneficial outcomes for any body composition outcomes, with BMI and waist circumference increasing at post-intervention and follow-up. This study was also limited by the use of subjective measures of physical activity.

In summary, the quality of research relating to the health benefits of physical activity in children with intellectual disabilities is generally weak and is limited by small sample sizes and subjective measures (Maiano, Normand, Aime, & Bergarie, 2014). The lack of breadth and depth in this research area is preventing definitive conclusions being made. Many of the trends which are present in research relating to typically developing children, such as the increased health benefits associated with a higher duration and intensity of activity, have not yet been established in children with intellectual disabilities. Furthermore, with existing research predominately focussing on exercise interventions, there is little evidence relating to reducing sedentary time and the health outcomes associated with various intensities of activity. Therefore, further research is required to increase our understanding of the relationship between physical activity and positive health outcomes in children with intellectual disabilitie.

1.3.4.2 Physical activity levels of children with intellectual disabilities

As the extent of health benefits is somewhat determined by levels of physical activity, it is important to understand the amount of physical activity that children actually do. Therefore, the aim of this section is to provide an overview of research relating to the physical activity levels of children with intellectual disabilities.

Consolidating previous research in this area poses many difficulties as there are multiple parameters of physical activity that can be measured, e.g. type and frequency, which result in non-comparable outcomes. For example, studies have measured daily step-count using pedometry (Suzuki et al., 1991), while others have used doubly-labeled water to measure total energy expenditure (van Mil et al., 2000). Although both these methods have merits, they are not comparable. Furthermore, these outcomes do not allow inferences to be made regarding the intensity or the patterns of activity, i.e. bouts, which are important when comparing activity levels to the recommended guidelines. Therefore, this section will focus on studies which report outcomes relating to levels of activity, i.e. minutes per day and activity intensity. This will also maximise the comparisons which can be made between individual studies.

McDonald, Esposito, and Ulrich (2011) used the Actical accelerometer to objectively measure physical activity levels in children with autism spectrum disorders (ASD; n = 72), some of whom additionally had intellectual disabilities, over a seven-day period. This study reports that children aged 12 to 18 years were on average active at a moderate to vigorous intensity for 90.02 (± 97.89) minutes per day, with children aged 9 to 11 years completing significantly ($p < .05$) higher levels of activity (131.57 ± 84.23 minutes per day). These high levels of physical activity are concurrent with Tyler, MacDonald, and Menear (2014), who measured the physical activity levels of children with ASD over a seven-day period using the ActiGraph GT3X+ accelerometer, and report that children spent 154.90 (± 50.10) minutes in moderate intensity activity and 165.90 (± 58.70) minutes in moderate to vigorous intensity activity per day.

In children with Down syndrome, Whitt-Glover, O'Neill, and Stettler (2006) report that children (n = 23) achieve an average of 153.10 (± 56.40) minutes per day of moderate to vigorous physical activity, when measured over seven-days using the Actitrac accelerometer. Similarly, Shields, Dodd, and Abblitt (2009) report that the study sample (n = 19) participated in an average of 104.50 (± 35.30) minutes of moderate to vigorous intensity activity per day, measured over seven-days using the RT3 accelerometer. Furthermore, younger children (7 to 12 years) are significantly ($p < .05$) more active (+36.40 minutes, 95% CI = 7.50, 65.30 minutes) than older children (aged 13 to 17 years). This is concurrent with the previous findings by McDonald et al. (2011) who also report that activity levels decrease with age. These high levels of activity are further supported by Pitetti, Beets, and Combs (2009) who measured physical activity during school recess and physical education using heart rate, and report that children ($n = 15$) were active at a moderate to vigorous intensity for an average of 83.50 minutes per day, suggesting that school-based activity alone is sufficient to achieve the physical activity guidelines.

These results suggest that children with intellectual disabilities are greatly exceeding the recommended levels of physical activity, although, some of the high standard deviations suggest that these mean levels of activity may not be representative of all children in the study samples. However, these findings are not consistent across all previous research, with multiple studies reporting levels of physical activity which are below the recommended guidelines.

Kozub (2003) reports the activity levels of children with intellectual disabilities ($n = 7$) ranged from 14 to 55 minutes of moderate to vigorous intensity activity per day, with levels of activity also reducing with age. Esposito, MacDonald, Hornyak, and Ulrich (2012) report that children ($n = 104$) with Down syndrome are insufficiently active to achieve the physical activity guidelines when activity was quantified using 7-day Actical accelerometer measurements. Furthermore, differences were identified between age groups, as children aged 8 to 9 years were active at a moderate intensity for 43.88 (± 15.95) minutes per day, which was significantly ($p < .01$) higher than children aged 14 to 15 years (23.79 ± 16.38 minutes). A similar trend was reported for vigorous intensity activity, as children aged 8 to 9 years were active at this intensity for 1.50 (± 1.89) minutes, whereas children aged 14 to 15 years were only active for 0.91 (± 1.48) minutes, although this difference was not significant.

A more recent study by Boddy, Downs, Knowles, and Fairclough (2015) measured physical activity levels of 70 children with intellectual disabilities over 7 days using the ActiGraph GT1M accelerometer. However, unlike previous studies, there was no significant difference in time spent in moderate to vigorous physical activity between children ≤ 11.9 years ($M = 50.10$ minutes, $SE = 5.60$ minutes) and children aged ≥ 12.0 years ($M = 47.40$ minutes, $SE = 6.40$ minutes). Although, children aged ≤ 11.90 years spent significantly less time sedentary than older children ($M = 414.10$ minutes, $SE = 17.20$ minutes and $M = 436.30$ minutes, $SE = 19.50$ minutes, respectively). Einarsson et al. (2015) also report that in a sample of 91 children with intellectual disabilities, none were achieving the recommended 60 minutes of moderate to vigorous activity per day. Furthermore, an interesting aspect of this study was the effect of environment, with children significantly ($p < .001$) less active at the weekend in comparison to weekdays, suggesting that the school environment could be important for physical activity.

The importance of school-based activity has previously been acknowledged, with studies focussing specifically on activity levels in this environment. A study by Horvat and Franklin (2001) investigated physical activity in the school environment using various methods (heart rate, Tritrac accelerometer activity counts, and Scheme for Observing Activity Levels direct observation tool) and found that children were most active in non-inclusive recess and least active during classroom time. MacDonald, Esposito, and Ulrich (2011) also measured the time of day when activity took place, and found that children were more active during school (35.10 to 48.23 minutes), in comparison with after school (10.28 to 17.32 minutes) and in the evening (25.99 to 40.48 minutes). Furthermore, Foley and McCubbin (2009) assessed physical activity levels during school-time using direct observation measurements and report that children with intellectual disabilities ($n = 80$) spend 145.70 to 134.10 minutes per week in moderate to vigorous physical activity. These findings are in contrast to the high school-based activity levels reported by Pitetti et al. (2009), although this could be a result of the various methods used to quantify activity.

In comparison with their typically developing peers, children with intellectual disabilities are generally less active (Einarsson et al., 2015; Frey et al., 2008; Stanish & Mozzochi, 2000; Tyler et al., 2014), although Lorenzi, Horvat, and Pellegrini (2000) report that children with intellectual disabilities are in fact more active than typically developing children. However, typically developing children engage in more vigorous intensity activity, whereas children with intellectual disabilities perform most of their activity at a moderate intensity (Stanish & Mozzochi, 2000). In regards to gender, boys with intellectual disabilities have consistently been noted as being more active than their female peers, with boys additionally recording higher intensity activity than girls (Frey et al., 2008; Lorenzi et al., 2000; Phillips & Holland, 2011).

The contradictory findings in these previous studies is preventing clear conclusions being made regarding the physical activity levels of children with intellectual disabilities. A limitation of studies in this field of research is the wide use of observational, cross-sectional study designs, and small sample sizes, which are common limitations in physical activity research involving children with intellectual disabilities (Frey et al., 2008). Although this section only focusses on research which included physical activity outcomes relating to intensity and duration, an advantage of these studies is the wide use of objective measures of physical activity, specifically accelerometers.

1.4 Measurement of physical activity in children with intellectual disabilities

To further our understanding of associations and effects between physical activity and health-related variables, investigate dose-response relationships, measure the effectiveness of interventions, and quantify compliance with physical activity guidelines, is it important that physical activity is accurately measured (Bull et al., 2010; Mahar & Rowe, 2002; Salmon & Okely, 2009; Warren et al., 2010). However, measuring physical activity in children with intellectual disabilities poses additional difficulties due to the variability within this group. As the term “intellectual disabilities” encompasses many syndromes, with various causes, this is a very heterogeneous group. Furthermore, there are specific disabilities which affect different disorders, e.g. abnormal gait patterns and atypical heart rates are associated with cerebral palsy and Down syndrome, respectively, which will be possible causes of error when measuring physical activity and sedentary behaviours. However, as previous research has predominately focussed on the population of children with intellectual disabilities as a whole, as opposed to specific syndromes, it is important to understand physical activity measurement in this wider population; although, this additionally highlights the need to recruit representative samples for measurement research and the need for researchers to be aware of possible disability-related effects when measuring activity.

When deciding on the best method of measurement to be used in a study, one of the most important considerations for researchers should be the reliability and validity of methods. Therefore, the following sections will define reliability and validity, and discuss the advantages and disadvantages of criterion, subjective, and objective methods of measuring physical activity in children with intellectual disabilities.

1.4.1 Reliability

Reliability, which generally refers to the consistency of a measure or agreement between raters, has many uses in research, such as: investigating the consistency, or stability, of a test administered on separate days; the agreement between tests which purport to measure the same construct; the test-retest reliability (internal consistency) of a test; and the objectivity of raters. However, this traditional view has received criticism for not accounting for participant variation (Linacre, 2000). This is specifically important when assessing reliability in individuals with disabilities, as this population have a high level of variability, i.e. fluctuations in daily behaviour, which can impact on obtaining reliable measures. Furthermore, there is a great amount of between-participant variability, i.e. inter-individual variation, in this population, which can further limit reliability and the generalisation of measurement methods and results (Linacre, 2000; Rikli, 1997).

Reliability is generally expressed as a correlation coefficient or percentage agreement. For the analysis of two different variables, an interclass correlation coefficient is calculated using Pearson r, whereas the coefficient for the analysis of the same variable, such as two raters measuring the same construct, is calculated using an analysis of variance approach or intraclass correlation coefficient (ICC; Thomas, Nelson, & Silverman, 2005). Furthermore, reliability can be independently established for a measure and is not dependent on validity, i.e. a device can be reliable without being valid. This is not the case for validity, which is dependent on reliability, thus making reliability an important aspect of measurement research (Thomas et al., 2005).

1.4.2 Validity

Validity is one of the most important and fundamental principles of measurement (Thomas et al., 2005). Ensuring the accuracy of physical activity measurements is crucial to furthering our understanding of associations between physical activity and health benefits, investigating dose-response relationships, measuring the effectiveness of interventions, and compliance with physical activity guidelines (Mahar & Rowe, 2002; Salmon & Okely, 2009).

The definition and theory behind validity has evolved in the last few decades. Traditionally, validity was viewed as the accuracy of an instrument to measure what it was supposed to measure (Linacre, 2000; Thomas et al., 2005). However, this definition ignores a fundamental concept of validity, which is the appropriateness of inferences made from measurements (Mahar & Rowe, 2002). Cronbrach (1971) argued that measurement devices cannot be validated; instead, the inferences based on these measurements should be the focus of validation. Linacre (2000) more recently suggested:

"*Validity is no longer established, once for all time, for the whole test, by criteria only indirectly related to the content of the test, such as the chronological age of the subjects. Instead, validity is reevaluated every time the test is administered, for each item in the test, according to the substantive theory which the test items are intended to implement.*" pg. 130

Validity should therefore be viewed as a multifaceted ongoing process in which evidence is accumulated using a range of research designs and methods, and which is established for each population group, context, and purpose for which a measure is used (Yun & Ulrich, 2002). In practical terms, validity coefficients cannot be generalised between populations, and the validity of measures and inferences needs to be re-established in different populations. This has important implications for research involving children with intellectual disabilities as, based on this definition of validity, it is not appropriate to assume that if a method is valid in typically developing children, it will also be valid in children with intellectual disabilities.

1.4.3 Sensitivity to change

Another consideration for researchers when choosing a device is sensitivity to change. Sensitivity to change refers to the ability of a measurement tool to detect meaningful changes over time (Cohen, 1977). This responsiveness is independent of validity and reliability and is specifically relevant to research when an increase in activity over time is the outcome of interest. Considering that small increases in physical activity levels or intensity can promote positive health benefits, it is important that the measurement device chosen is sensitive enough to detect this change. This is of primary importance in research involving children, including those with intellectual disabilities, as a limitation of previous interventions is the inability to sufficiently increase activity levels. Therefore, although no increase in activity could be a result of an ineffective intervention design, it could also be due to the measurement device not being sensitive enough to detect small changes in activity, such as 5-10 min/hr (Montoye, Pfeiffer, Suton, & Trost, 2014). Furthermore, sensitivity to change varies between measurement methods (Caballero et al., 2003). As a result, discrepancies within and between studies relating to changes in activity could further be affected by the method chosen, thus making sensitivity to change an important consideration for researchers.

1.4.4 Measurement methods

In a review of physical activity measurement conducted in 1985, LaPorte, Montoye, and Caspersen (1985) reported that there were over 30 methods which could be used to measure physical activity. However, 30 years on, there is still no universally accepted “gold standard” measure of physical activity. There are various methods which can be used to measure different dimensions of activity and, although there are advantages to all these methods, each method has at least one “Achilles’ heel” which prevents its use as a global measure of physical activity (Mahar & Rowe, 2002).

The use of different methods and the measurement of different dimensions of physical activity limits the comparison of results between studies and the consolidation of research. Since no method can accurately measure all dimensions of physical activity, the universal use of the term “physical activity” is somewhat misleading. Therefore, it is important for researchers to define physical activity (theoretical domain) and specify the construct they wish to measure, such as intensity and frequency, and chose an appropriate method of measurement based on the theoretical domain and study outcomes (operational domain; Mahar & Rowe, 2002). For example, if the theoretical definition of physical activity is focussed on movement, and the frequency and intensity of this movement, a measure should be chosen which allows these constructs to be measured.

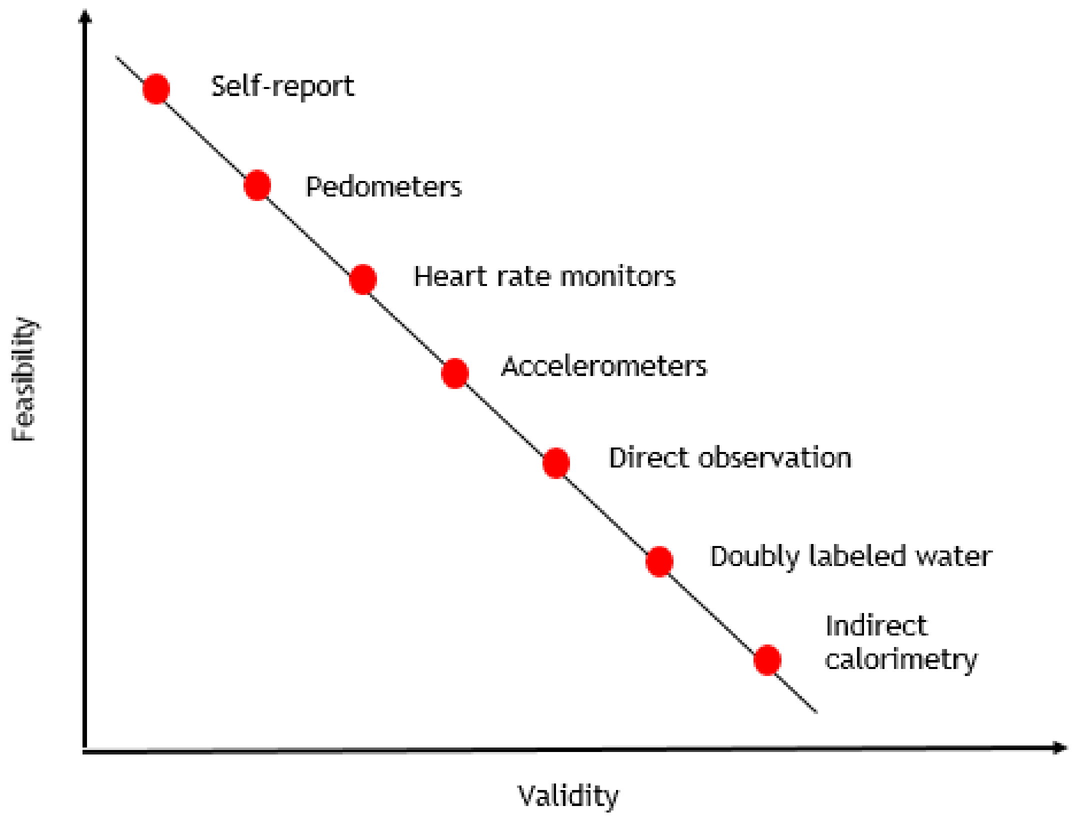


Figure 1.2. Illustration of the trade-off between feasibility and validity for physical activity measurement

As different measurement tools measure different dimensions of physical activity, e.g. energy expenditure or activity type, the method used should, theoretically, be the one which provides the most valid measure of the study outcome of interest. In practice, however, another important consideration when deciding upon a method of measurement is feasibility. The feasibility of a measurement device can relate to many factors, such as cost, participant burden, and complexity of data analysis. The choice of a measurement device is therefore a trade-off between validity and feasibility. Esliger and Tremblay (2007) discuss the interactions between validity and feasibility, a summary of which is presented in Figure 1.1. This illustrates that as the validity of a device increases its feasibility decreases.

The following sections will provide a brief overview of the most commonly used methods to measure dimensions of physical activity and will discuss advantages and disadvantages of each method, both in general and specific to children with intellectual disabilities.

1.4.4.1 Criterion measures

Criterion measures are the most valid methods to measure physical activity, but are also the least feasible. Due to the lack of feasibility with criterion measures, one of the primary uses of these methods is to validate other, more feasible, measures of physical activity; this type of validity is known as criterion validity. The three methods which are generally regarded as criterion measures are doubly labeled water, indirect calorimetry, and direct observation.

1.4.4.1.1 Doubly labeled water

Doubly labeled water measurements require participants to consume a dose of water containing a known concentration of non-radioactive forms of the stable isotopes of hydrogen and oxygen ($^{2}H_{2}\,^{18}O$). In the subsequent days and weeks (usually in the range of 3 to 21 days), labeled hydrogen leaves the body in the form of water, such as sweat, with labeled oxygen expelled as both water and carbon dioxide (CO_2). Through the analysis of salvia or urine, differences in these elimination rates allow total CO_2 production to be directly measured and the estimation of oxygen consumption and total energy expenditure (Katch, McArdle, & Katch, 2011).

Doubly labeled water is generally regarded as the most valid measure of energy expenditure (Kohl et al., 2000). Validity evidence for doubly labeled water is well established in adults, with this method accurate to within 3 to 4% of calorimeter measurements (Schoeller & Webb, 1984). Although criterion validity has been tested in children, this evidence is limited in comparison with adults due to the feasibility issues associated with other criterion measures, such as the practical limitations of conducing multiple days of whole-room calorimetry measurements (Goran, 1994; Sirard & Pate, 2001). In addition to its validity, there are several advantages to using doubly labeled water (Katch et al., 2011; Warms, 2005). Firstly, it is non-invasive and provides long-term free-living measurements. Secondly, it requires minimal participant and researcher burden.

Thirdly, it does not require participants to know that energy expenditure is the primary outcome measure and therefore can reduce reactivity.

However, doubly labeled water has many limitations which affect feasibility and prevent it being extensively used as a measure of physical activity. Although it is a valid measure of total energy expenditure, which is a domain of physical activity, it does not provide any information regarding the subdomains of physical activity, such as frequency, intensity, or type. Therefore, it is not possible to discern from this measure how much energy was expended as a direct result of physical activity. Furthermore, the required stable isotopes are expensive, with the subsequent analysis requiring sophisticated measurement equipment and expertise (Katch et al., 2011). The high costs and complex analysis associated with this method limit its feasibility and its use in large-scale research studies.

Previous studies involving children with intellectual disabilities which used doubly labeled water have utilised case study designs. These studies focus on children with Prader-Willi syndrome, which is associated with life-threatening obesity, where the study outcome of interest is total energy expenditure (Massersmith, Slifer, Gomez-Cabello, Pullbrook-Vetter, & Bellipanni, 2008; Singh et al., 2008). The use of doubly labeled water is an appropriate method of measurement for these studies. However, as both studies include only one participant, this also highlights that this method is generally only feasible for studies with small sample sizes.

Although this method has not been extensively used in children with intellectual disabilities, it is a feasible method for use in this population when the outcome of interested is total energy expenditure and a small sample size is used. However, considering the limitations discussed, doubly labeled water is not a feasible method for the measurement of physical activity in larger-scale studies which aim to measure other dimensions of physical activity (Kohl et al., 2000; Loprinzi & Cardinal, 2011).

1.4.4.1.2 Indirect calorimetry

Open circuit indirect calorimetry measures oxygen uptake ($\dot{V}O_2$) and carbon dioxide production ($\dot{V}CO_2$), from which energy expenditure can be calculated. The changes in oxygen and carbon dioxide percentages in expired air compared with inspired ambient air are used to indirectly measure energy metabolism (Katch et al., 2011). This provides a valid measure of $\dot{V}O_2$ and the energy expenditure requirements of specific types of activity (Warms, 2005). There are various techniques which can be used for indirect calorimetry, specifically the use of a stationary metabolic cart, a portable metabolic cart, or whole-room calorimetry. Although these methods provide the most valid method of measuring the intensity and duration of physical activity, the high cost of the equipment and need for trained personnel to measure and analyse this data negatively affect its feasibility.

Due to the equipment required for indirect calorimetry measurements, all these techniques are limited to a laboratory or controlled environment. This limits the feasibility of measuring unstructured/free-living physical activity using these techniques, although feasibility and validity vary between techniques. The use of a stationary metabolic cart requires a respiratory mask to be directly attached to the metabolic cart, which limits the freedom of movement. The portable metabolic cart varies to the stationary technique as the gas analysers are worn in a backpack by the participant, which allows activity to be almost unrestricted; however, this technique requires monitoring by the research team and therefore requires a controlled environment. A whole-room calorimeter does not require any equipment to be directly attached to the participant; instead the participant completes activities in a confined room (calorimeter) where the air and temperature are controlled and measured constantly. However, there are many feasibility issues associated with this method as it requires participants to remain in the confined calorimeter for hours at a time (Oortwijn, Plasqui, Reilly, & Okely, 2009).

Due to the high validity of these techniques to measure energy expenditure and $\dot{V}O_2$, they are commonly used as criterion measures to validate other devices, such as accelerometry (Bassett, Rowlands, & Trost, 2012; Kim, Beets, & Welk, 2012). On the other hand, due to limited feasibility, these techniques are not used for free-living measurements without researcher supervision. The use of these techniques has been limited for measuring physical activity in children with intellectual disabilities. Previous research in children with intellectual disabilities have used stationary metabolic carts to measure cardiorespiratory fitness during treadmill-based exercise tests (Fernhall, Millar, Pitetti, Hensen, & Vukovich, 2000; Fernhall, Pitetti, Stubbs, & Stadler, 1996). These studies report no issues with the measurement technique used, suggesting that the use of indirect calorimetry is feasible for children with intellectual disabilities.

1.4.4.1.3 Direct observation

Direct observation is the only measurement method which is focussed on physical activity behaviours. Direct observation measurements are conducted by trained observers who code physical activity behaviours, such as duration and type. Measurements are generally recorded using pencil-and-paper or computerized methods, with the use of video recording increasing the reliability of measurements (Loprinzi & Cardinal, 2011). There are various direct observation tools which can be used to measure activity, all of which have shown criterion validity evidence (Sirard & Pate, 2001).

An important advantage of direct observation is that it can capture many subdomains of physical activity, and is the only criterion measure which objectively measures type of activity. This has important implications for research as the type of activity conducted is related to body fat and habitual physical activity levels in children (Rowlands, Ingledew, & Eston, 2000). Furthermore, it also allows a vast amount of contextual data to be recorded, such as child interactions and teacher feedback. As physical activity is affected by environmental and contextual factors, direct observation provides objective measures on when, where, and with whom activity is conducted (McKenzie, 2002). Another advantage of this method is that it puts no measurement burden on the participant as it is non-invasive, which will therefore reduce reactivity and increase the validity of measurements.

Not without limitations, however, this method requires a high researcher burden, in terms of training and data collection/analysis (Warms, 2005). To ensure that the measurements recorded are truly objective, researchers need to be trained on how to accurately code activity, which requires a training period and training resources. Furthermore, as participants need to be in view of the research team or video cameras, use of this method is only feasible in small samples in a confined environment, such as schools. It also does not allow the direct measurement of activity intensity, which is important for understanding dose-response relationships and for use as a criterion measure; however, some measurement tools allow this dimension to be estimated using validated prediction equations (McKenzie, Sallis, & Nader, 1991).

Direct observation has previously been used successfully in children with intellectual disabilities to measure school-based physical activity and as a criterion measure to validate accelerometry (Capio, Sit, & Abernethy, 2010; Faison-Hodge & Porretta, 2004; Sit, McKenzie, Lian, & McManus, 2008). This previous use suggests that direct observation provides a feasible and valid method of measuring physical activity in children with intellectual disabilities.

1.4.4.2 Subjective measures

Subjective methods encompasses the quantitative and qualitative techniques used to measure self- or proxy-reported physical activity. There are various methods which can be used to subjectively measure physical activity, such as self-report questionnaires, interview-administered questionnaires, and physical activity diaries. However, there are many difficulties associated with using subjective measures in children with intellectual disabilities, which are consistent across all measurement methods. Therefore, the strengths and limitations of subjective measures will be collectively discussed specific to self- and proxy-reports.

1.4.4.2.1 Self-reports

Self-report measures are the most commonly used method to measure physical activity and are particularly popular in epidemiological research involving large samples (Bjornson, 2005). Self-report measures can be administered in various ways, including questionnaires, dairies, and interviews. Depending on the method or questionnaire used, self-report methods can measure all or some dimensions of physical activity, e.g. type, frequency, intensity, and duration.

The wide use of self-report measures is due to the high feasibility of this method. Self-report measures require minimal participant burden, are low cost, and are relatively easy to administer to a large number of participants. Furthermore, self-reports not only have the potential to measure all dimensions of physical activity, but can be focussed to include specific study outcomes of interest, such as intensity of activity, or type, e.g. work, household, or transport (Sallis & Saelens, 2000). This is an advantage for researchers as a large volume of specific data can be measured with limited burden on the participants and researchers. However, as highlighted in Figure 1.2, this high feasibility is associated with lower validity.

A review by Kohl et al. (2000) notes that self-report measures have low to moderate validity for the measurement of physical activity in children, with validity coefficients ranging from .03 to .88. However, validation against criterion measures is lacking in children (Sirard & Pate, 2001). A review by Sallis and Saelens (2000) reports that all included self-report measures were validated against objective measures, mostly accelerometry and heart rate, with none validated against a criterion measure. Therefore, this raises questions on the validity evidence established for self-report measures due to a lack of established criterion validity. A reason for this lower reported validity is that self-report measures are dependent on the participant's ability to provide valid information on their physical activity behaviours, which is reliant on cognition and memory/recall abilities (Matthews, 2002).Therefore, the validity of self-report measures is lower in younger children due to their lower cognitive and language development (Sallis, 1991). Furthermore, Baranowski et al. (1984) report that children under 10 years cannot accurately recall activity and are often not capable of understanding the concept of physical activity.

The recall and cognitive demands associated with self-reports restrict the use of this method in children with intellectual disabilities. The only identified study which used self-report measures in children with intellectual disabilities did so in conjunction with objective accelerometry measures (Einarsson et al., 2015). Participants were asked to complete the questionnaire, with assistance from a parent, to give contextual information to the accelerometry data, such as time spent in physical education and mode of transport to school. Therefore, considering the need for parental assistance when completing self-report measures, proxy-reports may be more suitable for children with intellectual disabilities.

1.4.4.2.2 Proxy-reports

As the use of self-report measures is limited in children with intellectual disabilities due to the recall and cognitive demands, an alternative method is to ask an adult close to the participant, such as a parent or teacher, to act as a proxy and report on the child’s physical activity behaviours.

Proxy-reports can overcome issues with recall bias in children with a lower developmental age. However, there are still difficulties associated with adults accurately recalling a child’s activity, as it can be difficult for a proxy to be able to constantly monitor all activity, such as both school- and home-based activity (Corder et al., 2008). Proxy-reports share many of the advantages associated with self-report measures, such as being low cost, easy to administer, and are little burden to the proxy or child. Furthermore, proxy-reported physical activity has been shown to have a moderate and significant relationship with accelerometry ($r = .41-.66$). However, the choice of proxy can have important implications on the validity of measurements, with parental-reported activity being more strongly related to heart rate ($r = .72-.82$) compared to teacher-reported activity ($r = .07-.59$; Sallis, 1991; Sirard & Pate, 2001). Similar to self-reports, however, the validity of proxy-reports varies between studies and is generally validated against objective measures rather than criterion measures. Another limitation with proxy-repots is social desirability, which results in activity levels being over-estimated.

As subjective measures are not suitable for children with a lower developmental age, the use of objective measures is recommended where possible (Trost, 2007b). However, for children with intellectual disabilities, similar to Einarsson et al. (2015), the use of proxy measures in conjunction with objective methods could provide added information on dimensions of activity, such as type, which many objective measures are not able to capture.

1.4.4.3 Objective measures

Objective methods generally measure physiological or biomechanical parameters of activity, which are subsequently used to estimate dimensions of physical activity, such as energy expenditure or activity intensity (Corder et al., 2008). The most commonly used objective measures are heart rate monitors and motion sensors. Motion sensors is the overarching term used to describe devices which measure body motion or movements, specifically pedometers and accelerometers. Considering movement is a fundamental component of physical activity, motion sensors are, theoretically, very pragmatic methods to measure activity.

1.4.4.3.1 Heart rate monitors

The use of heart rate monitors allows the collection of objective data relating to the frequency, duration, and intensity of physical activity. With the exception of indirect calorimetry and doubly labeled water, heart rate is the only measurement method which directly measures the body’s physiological response to activity, through an electrocardiogram transmitter worn around the chest which detects heart rate (Janz, 2002).

The measurement of heart rate has many advantages, primarily its feasibility. It enables the measurement of a physiological variable without the high participant burden associated with some criterion measures. Additionally, heart rate monitoring devices are inexpensive and relatively unobtrusive. Heart rate can provide reliable measures of physical activity and is particularly effective when used in conjunction with other methods (Kohl et al., 2000).

However, there are various factors which limit the validity of heart rate measures. Firstly, the relationship between heart rate and activity energy expenditure is only linear during moderate to vigorous intensity activity (Trost, 2007b). Therefore, heart rate is not deemed a valid measure for low intensity activity, which raises validity issues for the use of this method in free-living measurements involving inactive populations, such as children with intellectual disabilities. Other factors which affect the relationship between heart rate and energy expenditure include stress, age, cardiorespiratory fitness, and room temperature (Bjornson, 2005). As a result, measured changes in heart rate may not be a direct result of physical activity, thus introducing measurement error into the results. Finally, changes in heart rate are not instantaneous and lag behind actual changes in activity and the data recorded using other measures, which could limit the ability of heart rate monitoring to accurately capture the sporadic nature of children's activity. There are, however, various techniques which can be used to limit these effects (Corder et al., 2008; Trost, 2007b).

Heart rate monitoring has previously been used in children with intellectual disabilities for the measurement of physical activity intensity and during cardiorespiratory fitness testing (Baynard, Pitetti, Guerra, Unnithan, & Fernhall, 2008; Capio et al., 2010; Faison-Hodge & Porretta, 2004; Fernhall et al., 2001). Minor feasibility issues have been reported for the use of heart rate monitors in children with intellectual disabilities, as the wrist-worn device receiver has been noted as a distraction to children (Faison-Hodge & Porretta, 2004). Furthermore, some syndromes associated with intellectual disabilities, such as Down syndrome, cause atypical peak and resting heart rates, which needs consideration if heart rate monitoring is to be used in this population (Baynard et al., 2008).

Despite these limitations, heart rate could still be a feasible method of measuring physical which is at, or above, a moderate intensity (Riddoch & Boreham, 1995). Furthermore, the use of heart rate should be considered in conjunction with other measures of physical activity.

1.4.4.3.2 Pedometers

Pedometers are relatively simple devices which primarily measure step count. As walking can be undertaken by most people without any substantial risks or fitness requirements, it is one of the most commonly conducted physical activity behaviours. Therefore, the ability to objectively measure walking is appealing to physical activity researchers.

Pedometers are relatively inexpensive and unobtrusive for participants to wear, thus making them feasible for objective measurements, over multiple days, in large samples (Warms, 2005). Pedometers provide a measure of total walking activity over the measurement period - an output which is simple for researchers to interpret. Therefore, pedometers provide a feasible and simple method of measuring free-living physical activity over multiple days without the high cost and complex analysis associated with more sophisticated objective measures (Rowe, 2011).

Although the simplicity of pedometers is in some respects advantageous, it is also a limiting factor of this method. As the focus of most pedometers is on quantifying step count, no data is collected relating to other behaviours which are contributing to overall physical activity. Therefore, the use of pedometers is not an appropriate measure for measuring children's attainment of physical activity guidelines, nor does this method give any indication on the dose-response relationship. Furthermore, inter-instrument variability is high amongst pedometers due to the differing internal mechanisms between devices, which limits the generalisability of results (Corder et al., 2008).

Pedometers generally operate using spring-lever or piezoelectric mechanisms. Spring-lever devices contain a horizontal arm which moves up and down as a result of pelvic movement and vertical acceleration, specifically walking. This motion opens and closes an electric circuit which subsequently records a step. A limitation of this internal mechanism is that it is only effective when positioned vertically, which poses difficulties in obese populations (Crouter, Schneider, & Bassett, 2005). Newer piezoelectric devices contain a weighted horizontal cantilevered beam which applies pressure to a piezoelectric sensor during movement, which registers a step. These steps are then summed for the duration of the measurement period to provide a total score. A few more recent models can provide additional information on number of steps accumulated during each day of the measurement period, total distance walked, or calories expended (Bjornson, 2005). However, the algorithms used to calculate calories expended are not appropriate for children.

Pedometry has been previously used in children with intellectual disabilities to measure daily step count (Eiholzer et al., 2003; Suzuki et al., 1991). Furthermore, thresholds based on heart rate, and accounting for age and height, have been developed to translate step count into a measure of moderate to vigorous intensity, with an average of 122 steps/min representing moderate to vigorous intensity (Beets & Pitetti, 2011). Criterion validity evidence has also been investigated for children, for both step count and activity time, with mixed findings (Hinckson & Curtis, 2013). Valid measures are dependent on placement during constant walking, with the front right hip exhibiting the highest validity for both step count (ICC = .83; 95% CI .76, .88) and activity time (ICC = .99; 95% CI .98, .99), with the back placement showing the least validity ICC = .43 (95% CI .30, .59) and ICC = .65 (95% CI .53, .75), respectively (Beets et al., 2007). However, the validity of pedometers is lower during dynamic movements, with Pitetti, Beets, and Flaming (2009) reporting that the Walk4Life 2505 pedometer overestimates steps by 14% to 16.5% during physical education, against a criterion measure of direct observation.

Pedometers provide a low cost method to measure walking behaviours and are feasible for use in children with intellectual disabilities. However, as pedometers provide little information regarding duration, frequency, and intensity of activity, their use is limited to studies where the primary outcome is walking. Furthermore, considering the dynamic nature of children's physical activity behaviours, the effect of these movements on step count accuracy needs to be considered. Therefore, studies which aim to measure parameters of physical activity rather than walking should consider an alternative method (Corder et al., 2008).

1.4.4.3.3 Accelerometers

Accelerometers are small, lightweight devices which can be worn on various body placements, such as the waist, wrist, and ankle. Accelerometers are the only objective measure which can collect free-living data over multiple days on the frequency, intensity, and duration of physical activity. As illustrated in Figure 1.1, accelerometers provide the optimum balance between feasibility and validity. Accelerometers are relatively non-intrusive and are of little burden to participants and, therefore, are one of the most commonly used measures of free-living physical activity. Not without limitations, accelerometers are generally more expensive than other objective measures. Furthermore, the complexity of these devices makes collecting and translating raw data into physical activity outcomes potentially difficult, with various decisions facing researchers in relation to how to collect, reduce, and interpret accelerometer data.

Accelerometers measure raw biomechanical acceleration of the body on up to three planes (vertical, mediolateral, anterior-posterior) during movement (Chen & Bassett, 2005). Acceleration signals are converted into arbitrary activity counts which can be interpreted by equations or cut points to provide information on energy expenditure or activity intensity (Hinckson & Curtis, 2013; Kim et al., 2012). Therefore, to ensure valid interpretation of accelerometer data, population-specific cut points and equations need to be calibrated. However, a major limitation with the use of accelerometers in children with intellectual disabilities is that no population-specific equations or cut points have been developed, which limits the accuracy of estimating physical activity intensity and energy expenditure (Hinckson & Curtis, 2013).

Nonetheless, against a measure of direct observation, concurrent validity of counts has been investigated for the older ActiGraph AM7164 accelerometer in children with cerebral palsy ($r = .75$, $R^2 = .56$, $p < .001$), the RT3 accelerometer in adolescents and young adults with intellectual disabilities ($r = .76$), and the Actiwatch accelerometer in children with intellectual disabilities ($r = .10–.61$; Capio et al., 2010; Kozub, 2003; Taylor & Yun, 2006). However, due to differences in the internal design between accelerometer brands, there is limited comparability between raw outputs in the form of counts.

In comparison with other devices, accelerometers have the capabilities to provide in-depth data relating to the measurement of physical activity in children with intellectual disabilities. Although accelerometers have been relatively widely used in research involving children with intellectual disabilities, the complexity of accelerometers and the lack of conclusive research regarding validity is a limiting factor (Hinckson & Curtis, 2013). However, if the validity and use of accelerometers could be better understood in this population, accelerometry could be feasible for the measurement of physical activity.

1.4.5 Rationale for choosing accelerometers

Accelerometers provide a feasible and objective method to measure the intensity, frequency, and duration of physical activity. Due to the compact design and memory capacity of accelerometers, activity can be monitored over multiple days with minimal participant burden. Considering that little is known about the physical activity behaviours of children with intellectual disabilities, measuring these dimensions during free-living activity will develop our knowledge of the benefits of activity and increase our understanding of the dose-response relationship.

Accelerometers differ from other measures as they are still in their relative infancy and therefore are advancing regularly, with more user-friendly methods and in-depth outcomes being developed. For example, there an is increasing focus on understanding the raw acceleration signal which has the potential to allow the type of activity being conducted to also be measured using pattern recognition algorithms (Freedson, Bowles, Trioano, & Haskell, 2012). Therefore, promoting and increasing the use of accelerometers in children with intellectual disabilities will not only improve the quality and depth of data collected, but will keep this area of research abreast with emerging measurement technologies and techniques. This is important as physical activity research in children with intellectual disabilities, and our knowledge in this area, lags behind the research and knowledge-base involving typically developing children. Therefore, it is crucial that researchers strive to conduct high quality, relevant research in children with intellectual disabilities, a fundamental aspect of which is the use of a feasible and valid measurement method.

The measurement of physical activity has been the topic of several review articles. These have included general (Corder et al., 2008; Reilly et al., 2008; Warren et al., 2010) and population-specific (Rikli, 1997) reviews of subjective and objective measures. The findings from these reviews suggest accelerometers as the preferred method for measuring free-living physical activity in children. On the other hand, specific to populations with intellectual disabilities, the advocacy of accelerometer use is less. A review by Hinckson and Curtis (2013) notes the complexities of accelerometer use and the lack of established reliability and validity as reasons for the hesitation in promoting accelerometers for use in children with intellectual disabilities. However, these are limitations which can be addressed. Therefore, there is scope for this thesis to address the lack of knowledge surrounding accelerometer use and develop a body of research to increase the validity of accelerometers for use in children with intellectual disabilities.

Chapter 2 – Systematic review: accelerometer use in children with intellectual disabilities

2.1 Chapter overview

Chapter one discussed how accelerometers provide the optimum balance between validity and feasibility for the measurement of free-living physical activity. However, accelerometers are complex devices and pose many methodological use decisions for physical activity researchers. The purpose of this chapter is to expand upon and discuss the use decisions which face researchers when using accelerometers. Furthermore, this chapter will also systematically review how accelerometers are used in research involving children with intellectual disabilities, against best practice guidelines, to identify areas of accelerometer use which need additional focus in future research.

2.2 Introduction

Accelerometers provide a feasible method of measuring physical activity in children. The small and lightweight design of these devices allow physical activity to be measured over multiple days with minimal participant burden. However, the low participant burden is disproportionate to the high researcher burden pre- and post-data collection (McClain & Tudor-Locke, 2009). Prior to collecting data, researchers are faced with multiple use decisions, such as which accelerometer to use and how many days of monitoring are required. Furthermore, the burden on researchers is higher post-data collection due to the substantial volume of data collected by accelerometers. Accelerometer count data is arbitrary, with only a portion of the total data collected relevant to the study outcomes. Therefore, post-data collection, researchers need to decide what data is relevant and identify methods to extract, reduce, and interpret this relevant data.

The most widely cited considerations which face researchers using accelerometers are: device selection, device placement, epoch length, number of monitoring days, what constitutes a valid day, data reduction, and interpreting accelerometer output (Cliff, Reilly, & Okely, 2009; de Vries, Bakker, Hopman-Rock, Hirasing, & van Mechelen, 2006; Reilly et al., 2008). Therefore, the following sections will discuss each of these use decisions.

2.2.1 Device selection

One of the first decisions facing researchers who want to measure physical activity is which accelerometer to use. As there are numerous commercially available accelerometers, choosing a device can be a complex decision. Previous measurement reviews have identified upwards of fifteen different research-grade accelerometers, not accounting for various versions of the same device (Murphy, 2009: Reilly, et al., 2008). These devices can vary greatly, e.g. in size, weight, number of axes measured, price, wear location, integration of other data sources, data processing/storage, and reliability and validity.

An important consideration when deciding upon an accelerometer is validity. However, there is no conclusive evidence on the superiority of one device over another, in terms of reliability and validity (Rowlands, 2007; Trost, McIver, & Pate, 2005). In typically developing children, ActiGraph accelerometers are most commonly used due to the higher volume of validity evidence for these devices (McCain & Tudor-Locke, 2009). In children with intellectual disabilities, however, validity evidence is limited. Therefore, this is preventing researchers from making an evidence-based decision on device selection in relation to validity. As discussed in section 1.4.3.3.3, only three studies have investigated the validity of raw accelerometer output in children with intellectual and developmental disabilities, with the strength of validity evidence device-dependant. The lowest validity was reported for the Actiwatch ($r = .10$-$.61$), with similar validity reported for the ActiGraph AM7164 ($r = .75$, $R^2 = .56$, $p < .001$) and RT3 ($r = .76$); however, these studies included small sample sizes (7 to 31 participants), with participants ranging from children to young adults aged between 6 and 25 years (Capio et al., 2010; Kozub, 2003; Taylor & Yun, 2006).

Another consideration which is important when deciding upon a device is the number of axes the accelerometer measures. In general, devices measure acceleration of the body on either one axis (vertical) or three axes (vertical, medio-lateral, and anterior-posterior) of the body (Chen & Bassett, 2005). Theoretically, triaxial accelerometers should be more valid as the additional inclusion of the medio-lateral and anterior-posterior axes should more accurately capture activity and, in particular, the high-intensity, sporadic movements conducted by children. However, there is little empirical evidence to support this, with numerous review articles suggesting the advantages of triaxial accelerometry over uniaxial accelerometry are negligible (de Vries et al., 2006; Reilly et al., 2008; Rowlands, 2007).

As there is limited evidence to suggest the superiority of one accelerometer over another in terms of validity and number of axes measured, researchers should consider practical differences between devices, such as cost, size, memory capacity, technical support, and devices used in previous research, which will increase comparability (Rowlands, 2007; Trost et al., 2005). A review by de Vries et al. (2006) compared some of the most commonly used accelerometers, with many practical differences between devices highlighted. For example, cost per device ranges from $500 to $2270 (RT3 and Actiwatch, respectively), with the weight of devices ranging from 30g to 170.4g (Tracmor2 and Tritrac-R3D, respectively). Furthermore, another important consideration for multiple day measurements is device storage capacity and the effect this has on monitoring days. Rowlands (2007) compared the storage capacity of the RT3 device and the ActiGraph GT1M, showing that if data is recorded every second, the RT3 can measure activity for 9 hours, whereas the ActiGraph can measure almost 6 days of activity.

In summary, accelerometers generally show similar levels of validity for the measurement of acceleration and movement, with little consensus on which device is most valid. Furthermore, there are many practical differences between devices which will impact on study outcomes and feasibility. Therefore, when deciding upon an accelerometer, researchers need to think about the aims and outcomes of the study and decide which accelerometer features are most important.

2.2.2 Device placement

Device placement refers to the location on the body where an accelerometer is positioned and how the device is attached. Accelerometers can be worn in various placements, such as the waist, wrist, arm, or ankle. However, dependant on the type of accelerometer used, device placement is not always a decision researchers need to make, as some devices are placement specific. For example, the Sensewear accelerometer is placement specific to the upper arm, whereas the ActiGraph can be worn on either the waist, wrist, or ankle. For devices that can be worn on multiple placements, it is important that researchers understand the implications of device placement. It is important to consider the ergonomics of accelerometer placement, i.e. the interactions between the device and human body, and decide upon an unobtrusive placement which captures the movements of interest (Yang & Hsu, 2010).

As accelerometers are most commonly used to measure whole body movement, the most common placement is around the waist, i.e. hip or back. The waist placement is generally most feasible as it does not inhibit movement and is associated with minimal discomfort (Yang & Hsu, 2010). Many commonly conducted movements, such as walking and running, result in trunk acceleration which is more accurately detected by waist-worn accelerometers (Sekine, Tamura, Togawa, & Fukui, 2000; Yang & Hsu, 2009). Devices worn at the waist have shown high validity against a criterion measure of indirect calorimetry in children during treadmill walking (Bouten, Sauren, Verduin, & Janssen, 1997). Furthermore, due to limited validity and the inability to directly measure trunk movement, the authors of this study recommend that the wrist and ankle placements are not used. However, this study only included two participants, therefore the validity and generalisability of these results are questionable. Nilsson, Ekelund, Yngve, and Sjöström (2002) compared the difference between the lower back and right hip placement and found no significant difference in total physical activity estimates. However, the authors discuss that the hip is a more comfortable and unobtrusive placement, which will increase compliance with wearing the device, and therefore recommend the hip placement.

The model or generation of a device, in relation to the internal design, also needs to be considered when deciding upon device placement. Older devices, particularly those which contain a cantilever beam and seismic mass, have to be worn in line with the vertical axis of the body, and as near as possible to the bodies center of gravity, to allow the greatest accuracy of acceleration detection (Chen & Bassett, 2005; McClain & Tudor-Locke, 2009). However, a limitation with the waist placement is the inability of devices to measure upper body movements or certain types of activities, such as cycling (Warren et al., 2010). On the other hand, devices worn on the extremities may not accurately capture all trunk movement and acceleration. Therefore, researchers need to acknowledge the limitations of accelerometry and that the data collected is affected by placement, which will impact on comparability between placements.

There are direct and indirect methods which can be used to attach an accelerometer, which are generally device specific. For example, the ActivPAL accelerometer is attached directly to the skin and not removed by participants; the ActiGraph is secured using a removable elastic belt; and the Actiwatch is attached to the body using a removable strap. However, regardless of the method/equipment used to attach the device to the body, it is important for researchers to ensure that devices can be securely attached, as loose attachment methods will cause extraneous movements to be detected, thus reducing the accuracy of measurements (Yang & Hsu, 2010). Furthermore, researchers should consider the feasibility of attaching the device and aim to ensure participant comfort to increase compliance (Nilsson et al., 2002).

In summary, the hip placement is recommended for use in children, as it provides a relatively unobtrusive measurement of whole-body movement. However, as some devices are placement-specific, researchers should be aware of device-placement, and the subsequent effect on validity and comparability between studies, when choosing an accelerometer.

2.2.3 Epoch length

Accelerometers constantly collect data over the course of the measurement period, with some devices capable of recording acceleration data up to 100 times per second. However, this depth of detail is too vast - and often irrelevant - for the outcomes of most physical activity studies and additionally requires a large amount of data storage capacity. Therefore, data in the form of counts are averaged over a selected time period, or epoch, which can range from 1-second to 60 seconds. Epochs can either be manufacturer determined, which cannot be changed, or selected by the researcher from a specific range of epochs, which varies between devices (McClain & Tudor-Locke, 2009). For example, the ActiGraph enables activity to be measured in as short as 1-second epochs, whereas the shortest epoch available from the Actical device is 15 seconds.

The choice of epoch length has to be chosen based upon the depth of data required, device storage, and the population of interest. In adult populations, 60-second epochs are commonly used due to the more constant patterns of activity conducted by this population. On the other hand, children generally conduct intermittent bursts of short, high intensity activity which may require a shorter epoch. On average, children's high intensity bouts of activity last 3 seconds, with 95% of these bouts lasting less than 15 seconds (Bailey et al., 1995). Similar findings have been reported for children with intellectual disabilities, with the average time of high intensity bouts ranging from 2.0 to 2.5 seconds (Shields et al., 2009; Whitt-Glover et al., 2006). Therefore, an epoch which is averaged over 60 seconds may not detect this type of sporadic activity, with high intensity activity misclassified as a lower intensity. As a result, the use of shorter epochs, such as ≤ 15 seconds, has been recommended for children (Corder et al., 2008; Trost et al., 2005; Welk, Corbin, & Dale, 2000).

However, the effect of epoch length is affected by study outcomes, specifically physical activity intensity. The effect of epoch length is smaller for lower intensity activity but greater for higher intensity activity. Nilsson et al. (2002) applied various cut points ranging from 5 to 60 seconds to the same data set collected over 4 days using the ActiGraph AM7164 device. For the hip placement, the difference in the estimation of physical activity intensity between 5-second and 60-second epochs was 8.0, 27.2, and 10.6 minutes for moderate, vigorous, and very vigorous intensity activity, respectively, with a trend that the amount of time recorded for each intensity decreased as the epoch increased. A similar study by Rowlands, Powell, Humphries, and Eston (2006) compared the use of a 1-second and 60-second epoch using the RT3 accelerometer over a 6-hour period. This study reports significant differences between the epochs for moderate, vigorous, and very hard activity; however, unlike the findings reported by Nilsson et al. (2002), the 60-second epoch resulted in higher time for each intensity compared to the 1-second epoch, except for very hard intensity.

Although both Nilsson et al. (2002) and Rowlands et al. (2006) suggest the use of a shorter epoch, these findings may suggest that a 1-second epoch is too short to capture all activity, therefore the use of a slightly longer epoch, such as 15 seconds, may be more appropriate. Furthermore, an advantage of using a shorter epoch is that data can be reintegrated into a larger epoch, whereas a longer epoch cannot be converted into a shorter epoch (Corder et al., 2008). However, Reilly et al. (2008) disagree with the use of a shorter epoch and conducted a secondary data analysis which notes no significant differences for the estimation of sedentary behaviour between 15, 30, 45, or 60-second epochs, measured using the ActiGraph. For moderate to vigorous intensity activity, significant differences were found between epochs; however, the authors suggest that these are not clinically significant and recommend the use of 60-second epochs, unless the primary outcome of the study is vigorous intensity activity. It is important to note, however, that none of these studies included a criterion measure, so it is not possible to determine which epoch is most valid.

Previous research is inconclusive on the most valid epoch length. Therefore, researchers should consider other factors when deciding upon the most appropriate epoch for their study. More recent accelerometers can measure physical activity in 1-second epochs, which produces 86,400 measurements per participant for a 24-hour measurement period and 604, 800 data points for a seven-day monitoring period. As a result, the advantages of collecting in-depth data using shorter epochs has to be weighed against the processing of this large volume of data. Furthermore, data storage should be considered to ensure the device can store all data collected over the measurement period, and that there is sufficient computer storage capacity available for large-scale studies.

Another consideration for epoch is how the data will be interpreted. For example, if cut points are to be used to classify counts into intensity brackets, many cut points are based on 60-second epochs, although more recent cut points have been calibrated using smaller epochs. Therefore, measuring activity using short epochs would increase the workload for researchers, as data would need to be reintegrated into larger epochs. On the other hand, if data analysis is focussed on the raw acceleration signal, then shorter epochs, i.e. 1- or 2-second epochs, would be necessary (McClain & Tudor-Locke, 2009). A more recent advancement with some accelerometers, such as recent versions of the ActiGraph, is that these devices do not require an epoch length to be selected prior to data collection, instead data is continually recorded. As a result, data can be converted into epochs after data collection, with various epochs applied to the same data set. Therefore, if these devices are used, the epoch length chosen will only affect the interpretation of data and not data collection.

In summary, although there is some debate in the literature surrounding the most suitable epoch length for use in children, there is a higher volume of research recommending the use of a shorter epoch (< 15-seconds). However, if using a shorter epoch, researchers need to ensure they have the sufficient device data storage for the measurement period. The trend with advancing technology, however, suggests that the effect of epoch length may lessen in the coming years with the availability of more flexible epoch application processes.

2.2.4 Number of monitoring days

When designing a study involving free-living physical activity, researchers have to decide how many days physical activity will be measured over. Children's physical activity behaviours exhibit high inter- and intra-individual variability; therefore, the decision facing researchers is how many monitoring days are required to obtain reliable and representative measurements, whilst reducing random error (Baranowski, Masse, Ragan, & Welk, 2008; McMurray et al., 2004). A wide range of monitoring days have been reported in previous studies involving children, ranging from 2 days to 2 weeks (Finn, Johannsen, & Specker, 2002; Hoos, Plasqui, Gerver, & Westerterp, 2003).

Theoretically, it could be assumed that the longer the measurement period the more representative the data will be. However, as the measurement period increases, participant adherence with wearing the accelerometer decreases and, after multiple days, reliability also decreases (Baranowski, Masse, Ragan, & Welk, 2008; Corder et al., 2008; Penpraze et al., 2006; Trost, 2007b). On the other hand, due to reactivity with wearing accelerometers, children record significantly higher levels of activity on the first measurement day, which could reduce the reliability of shorter measurement periods (Mattocks et al., 2008). Therefore, previous research has investigated the minimum number of monitoring days required to obtain a reliable and representative measure of physical activity, with limited participant burden; however, the results have been conflicting.

One of the earliest studies to investigate wear time reports that for the ActiGraph AM7164 in children aged 7 to 15 years, 4 days of measurement is recommended, which gives reliability between .75 and .78 (Janz, Witt, & Mahoney, 1995). Trost, Pate, Freedson, Sallis, and Taylor (2000) additionally note that the minimum number of required days is dependent on age and desired level of reliability. Using the Spearman-Brown analysis - which tests the effect on reliability of increasing or decreasing the monitoring period - for children aged 6 to 11 years, 2 to 3 days and 4 to 5 days are required to achieve reliability of .70 and .80, respectively. However, for children aged 12 to 16 years, 4 to 5 days and 8 to 9 days are required to achieve the same reliability coefficients of .70 and .80, respectively. Furthermore, the authors note significant differences in physical activity levels between weekday and weekend days and therefore recommend that physical activity is measured over 7 days for both children and adolescents to account for this variance.

Treuth et al. (2003) report that in a sample of 8 to 9 year old African-American girls, 4 days of monitoring gives a reliability coefficient of .37, with 7 days required to produce a coefficient of .80. In comparison with Janz et al. (1995), this low coefficient suggests that the required number of monitoring days is potentially affected by sex or ethnicity. A study by Penpraze et al. (2006) investigated the number of monitoring days in younger children (M = 5.6 years). Concurrent with Trost et al. (2000), Penpraze et al. (2006) report that reliable measures can be obtained from shorter measurement periods in younger children, with 2 monitoring days producing 70–73% reliability and 4 days giving 82–84% reliability; however, the authors also recommend the use of a 7-day measurement period to maximise reliability.

Another factor to be considered when deciding on the monitoring period is sample size. Based on a sample of 5595 children, Mattocks et al. (2008)

recommend a 3-day measurement period, as this provides good reliability (r = .70) and reduces the number of participants excluded due to insufficient wear time, thus increasing statistical power. However, this shorter measurement period may not provide reliable measurements in smaller sample sizes.

Deciding upon the number of monitoring days is a difficult decision for researchers. In addition to age, the inclusion of weekday or weekend days, sample size, seasonality, school terms/holidays, and climate will also affect the reliability of findings (Baranowski et al., 2008; Corder et al., 2008). The reliability of monitoring days in previous studies has been estimated from ICC scores, using the Spearman-Brown formula. However, the validity of these methods to establish reliability for wear days is questionable, with Baranowski et al. (2008) suggesting that the use of ICCs will underestimate the required number of days due to the violation of variance-related assumptions. Therefore, practical considerations, such as outcomes and cost could be influential factors in the decision of wear days (Trost et al., 2005).

In summary, researchers need to ensure that physical activity is measured over a sufficient number of days to obtain reliable measures of physical activity. To account for group differences, such as age, there is a consensus that 7 days will provide reliable physical activity estimates. However, the feasibility of this measurement period should be considered in relation to practical considerations, such as cost and participant/researcher burden.

2.2.5 Valid day

Physical activity levels and intensity not only vary from one day to another, but also from minute-to-minute and hour-to-hour (Trost et al., 2000). Therefore, it is important that a sufficient number of hours per day are measured to get a valid representation of daily activity. Although it would be ideal for researchers to constantly measure physical activity 24 hours per day, this adds an increased burden to participants, which may reduce compliance. Based on this, researchers generally ask participants to wear an accelerometer for all waking hours, except during swimming and bathing. However, not all study participants comply with these recommendations, with wear time varying between participants. Therefore, similar to deciding upon the number of monitoring days, researchers need to decide the minimum number of hours per day required to be included in the analysis.

The decision of what constitutes a valid day should be considered in relation to the number of monitoring days. Penpraze et al. (2006) provides a detailed summary of the interactions between daily wear time, number of monitoring days, and participant compliance. The optimum reliability (.80) was 10 hours per day of wear for 7 days, with 75 out of 76 participants achieving these requirements. Interestingly, reliability remained almost constant from 3 to 10 hours of wear across each monitoring day. For example, 4 days of monitoring produced a reliability of .69 for both 3 hours and 10 hours of wear time. For 7 days a reliability of .79 was achieved for 3 hours wear, which only increased to .80 for 10 hours of wear. Furthermore, for daily wear time exceeding 10 hours, reliability and the number of participants returning complete data notably reduces, suggesting daily wear requirements should not exceed 10 hours. These findings are contrary to those by Trost et al. (2000) who noted within-day variation in time spent in moderate to vigorous intensity activity, which suggests that shorter measurement periods, such as 3 hours, are insufficient. However, Trost et al. (2000) did report that daily wear time was lower in younger children, therefore, the drop in reliability and participant compliance >10 hours reported by Penpraze et al. (2006) could be a result of the shorter waking hours in younger children.

Similar to the findings by Penpraze et al. (2006), Mattocks et al. (2008) reports that hours of wear per day has almost negligible effects on reliability, with ICCs increasing from .43 to .45 for 7 hours and 10 hours of wear, respectively. Another interesting finding by Mattocks et al. (2008) is that children who are younger, lighter, and whose mother has a higher level of education return a higher number of valid days. A study by Rich et al. (2013) uses data from the large-scale UK Millennium Cohort study to investigate daily wear time in a sample of 7,704 children, and recommends 10 hours per day over 2 days, which gives a reliability of .86 and an included sample of 6,528 participants. This study additionally supports the findings reported by Mattocks et al. (2008) in Section 2.2.4 that less monitoring time, in relation to wear hours and wear days, is required for large sample sizes.

In summary, although the research relating to what constitutes a valid day is limited, there is a consistent conclusion that daily wear time has an almost negligible effect on reliability, as long as activity is measured over a sufficient number of days. However, if a participant returns days which do not meet the required number of wear time hours, researchers need to decide how to deal with this missing data, specifically, whether it is excluded from the final data set or if the data is imputed – a decision which will impact on the power of the study.

2.2.6 Data reduction

In general, accelerometers will continually record data for the duration of the measurement period, irrespective of whether the participant is wearing the device or not; exceptions being the Sensewear which stops recording when not in direct contact with the skin and the ActivPAL which is not removed. Therefore, after multiple days of measurement, researchers will have a vast amount of data, only some of which is relevant. Data reduction is the process of extracting relevant data, i.e. that which meets the minimum valid days for the monitoring period, identifying and removing spurious data, and dealing with missing data. This is an important stage of the measurement process as it reduces random and systematic error, e.g. identifying spurious data points or checking for device malfunction, and increases the validity of accelerometer measurements (Cliff et al., 2009). However, no standardised criteria exist for cleaning and reducing accelerometer data in children (Rowlands, 2007).

As highlighted in the previous section, accelerometers are generally only worn during waking hours. Therefore, periods where the accelerometer was removed, e.g. during swimming or sleeping, have to be identified in the data set. Due to the high sensitivity of accelerometers, even minimal movements will be recorded and increase the count score per epoch above zero. Therefore, a commonly used method to identify non-wear time is strings of zeros (Cliff et al., 2009). Various definitions of non-wear have been used in previous research, ranging from 10-minute to 180-minute bouts of constant zeros (Rowlands, 2007). However, Esliger, Copeland, Barnes, and Tremblay (2005) note that the average bout of motionless wear time, i.e. a string of zeros when the device is being worn, is 17 minutes in children aged 8 to 13 years, and therefore recommend the use of > 20 minutes of zero counts to represent non-wear. On the other hand, the authors suggest the use of 15,000 counts per minute (cpm) for the ActiGraph as the upper boundary for what is considered biologically plausible, with scores above this threshold deemed not to be a result of physical activity.

Once spurious and missing data have been identified, researchers need to decide whether to exclude these data points from the analysis or impute the data. This decision has to be considered in relation to the effect that excluding data/participants will have on statistical power. However, the imputation of missing data has been recommended as it is effective in reducing bias and increasing the precision of results, regardless of the imputation technique used, whilst retaining statistical power (Catellier et al., 2005). Although, the more stringent the methods used to classify non-wear time, the lower moderate to vigorous intensity activity reported and the lower the statistical power (Masse et al., 2005). Due to the large researcher burden associated with data reduction, when deciding upon which accelerometer to use, the availability of data-reduction and processing programmes available for each device should be considered (McClain & Tudor-Locke, 2009). Furthermore, considering the varying methods which can be employed, it is important that researchers fully describe data reduction techniques used to increase comparability between studies.

In summary, data reduction and identifying/dealing with missing data points is one of the most complex aspects of accelerometer use. Using > 20 minutes of constant zero counts to identify non-wear is recommended, with missing data imputed to retain statistical power and reduce bias. Considering the complexity of data reduction, researchers should consider the availability of processing programmes and technical support when selecting a device.

2.2.7 Interpreting accelerometer output

After the collection and reduction of accelerometer data, the next decision for researchers is how to translate arbitrary count data into a physiologically meaningful outcome. Interpreting accelerometer output is one of the most important decisions facing researchers as this information is used to develop our knowledge of the health benefits of activity, the dose-response relationship, and the attainment of physical activity guidelines (Freedson et al., 2012). The most common methods to do this is through the application of regression equations for the estimation of energy expenditure, or count cut points for the estimation of physical activity intensity.

The use of equations provides energy expenditure output for either specific bouts of activity or total activity. On the other hand, cut points categorise activity into intensities, such as sedentary, moderate, and vigorous, for a specified epoch (usually 60 seconds), which can then be summed to give a total score for time spent in each intensity. The decision of which method to use should be primarily based on the outcomes of the study. However, equations and cut points are device- and population-specific. Therefore, when deciding upon which accelerometer to use, consideration should be given to the availability and validity of cut points or equations for each device, specific to the population of interest (McClain & Tudor-Locke, 2009).

Cut points and equations are fundamentally based on the relationship between energy expenditure and accelerometer output and are calibrated by concurrently measuring physical activity using a criterion measure, such as indirect calorimetry or direct observation, and accelerometry. However, the relationship between these variables is complex, with calibration affected by various factors, such as maturation, sex, and level of cardiorespiratory fitness (Freedson et al., 2012). Therefore, numerous equations and cut points have been developed to account for these population differences, with little consensus to which equations or cut points should be used (Kim et al., 2012).

As a result, there is high variation in intensity cut points calibrated in children, with sedentary cut points ranging from < 101 cpm to < 800 cpm, and moderate to vigorous cut points ranging from > 500 cpm to > 3580 cpm (Freedson, Pober, & Janz, 2005; Mattocks et al., 2007; Puyau, Adolph, Vohra, & Butte, 2002: Treuth et al., 2004). Similar differences are present for energy expenditure equations, which include varying or no participant-related variables, such as age or sex. Therefore, the cut points or equations used will have significant and clinically meaningful effects on results (Reilly et al., 2008; Trost, Loprinzi, Moore, & Pfeiffer, 2011). Furthermore, many of the developed cut points and equations exhibit limited criterion validity, especially when generalised to populations which are different to the original calibration sample (Corder et al., 2007;

McClain, Abraham, Brusseau, & Tudor-Locke, 2008; Trost et al., 2011; Warolin et al., 2012). An additional area of concern for research involving children with intellectual disabilities is that no cut points or equations have been developed specifically for this population, which raises questions of the validity of data interpretation (Frey et al., 2008; Hinckson & Curtis, 2013). Therefore, one method to limit the effect of varying data interpretation techniques is for researchers to also report data in the format of counts.

In summary, accelerometer data can be interpreted using energy expenditure equations or intensity cut points, with the decision of which method to use dependant on study outcomes. However, as multiple cut points and equations have been developed, comparability between studies using different cut points or equations is limited. Furthermore, as these techniques are population-specific, the validity of using cut points or equations in populations different from that in which they were originally calibrated will introduce measurement error and affect validity.

2.2.8 Accelerometer use summary

This section has discussed some of the decisions which face physical activity researchers when using accelerometers. Deciding upon methods of use is complex, yet it is important that researchers understand the effect that these discussions will have on study outcomes and comparability between studies. The use decisions discussed surrounding accelerometers are not independent but interrelated. Therefore, it is important that prior to making any decisions researchers understand the outcomes of their study, and what factors are most important, to ensure that the most appropriate methods are selected. Although ensuring validity is important, for many of the use decisions discussed there is little consensus on which methods of use are most valid. Therefore, in practical terms, researchers should consider the aims of their study and the feasibility of their decisions, in terms of researcher and participant burden. Furthermore, it is also important that researchers fully report and justify their use decisions to give clarity to readers as to why specific decisions were made.

2.2.9 Accelerometer use in children with intellectual disabilities

As interest in physical activity in children with intellectual disabilities increases, so does the number of studies reporting the use of accelerometers in this population. A brief literature search for accelerometer and intellectual disabilities related terms highlights that studies citing accelerometry has almost doubled in the past decade (2005 to 2014). However, very little of this research is focussed on measurement. Therefore, almost all of the research discussed in this introduction section is specific to typically developing children as no studies have investigated how accelerometers are used in children with intellectual disabilities. There is a lack of research focussing on standards of practice for using accelerometers to measure physical activity in both typically developing children and children with intellectual disabilities (Freedson et al., 2012; Hinckson & Curtis, 2013). Furthermore, there is a need for more systemic reviews to increase and consolidate the knowledge-base relating to the measurement of physical activity in populations with disabilities (Cervantes & Porretta, 2010). Therefore, it is important to increase our knowledge of accelerometer use in children with intellectual disabilities, which will allow areas for future research to be identified.

2.2.10 Rationale for a systematic review

2.2.10.1 Definition and purpose of research synthesis

Research synthesis is the process of reviewing independent studies on a similar topic and culminating this existing research to form new conclusions. Reviewing existing literature can serve many purposes, such as a precursor to a new research area, to gain a greater understanding of existing knowledge from which a research agenda can be formulated, or as a stand-alone piece of work which aims to answer specific research questions (Badger, Nursten, Williams, & Woodward, 2000). This has many benefits for individual researchers and the research community as a whole as it prevents the over replication of research - or “re-inventing the wheel” - and instead enables previous research to be built upon and new knowledge generated.

Reviewing existing literature has become more important in recent years due to the growth of available resources and the ease with which literature can be accessed (Chalmers, Hedges, & Cooper, 2002). This is in part due to technological advances, such as the internet and electronic databases, and increasing numbers of conferences (Badger et al., 2000). The advancement of electronic databases has increased the accessibility of research, with policy makers now relying more on research synthesis to keep abreast of research developments and to inform policy, e.g. the development of physical activity guidelines (Bull et al., 2010; Chalmers et al., 2002). However, a distinction can be made between unstructured reviews of the literature and systematic reviews.

2.2.10.2 Limitations of an unstructured review

Unstructured reviews of existing literature rely on subjectivity and individual judgement in relation to which studies should be included in a review, strengths and weaknesses of studies, and overall conclusions drawn (Pillemer, 1984). This unstructured approach therefore creates a situation in which researchers can draw different conclusions based on a review of the same studies. Furthermore, the lack of structured inclusion criteria increases the likelihood that a review will only include, or focus more attention on, studies which reported positive and significant results (Chalmers et al., 2002). In addition, small-scale studies, those with negative results, and theses are less likely to be published or as easy to identify without a systematic search and inclusion criteria; therefore, valuable information may be excluded from an unstructured review (Badger et al., 2000). As a result, an unstructured review may be detrimental as a starting point for identifying areas for future research due to the lack of validity and reliability surrounding the conclusions drawn from this methodology.

2.2.10.3 Advantages of a systematic review

A systematic review utilises a structured methodology which minimises bias and enables a valid synthesis and critical appraisal of research (Egger, Smith, & Altman, 2001). Detailed guidelines have been developed which provide extensive methodological guidelines on conducting high quality systematic reviews (Higgins & Green, 2011). Furthermore, to ensure complete and transparent reporting of systematic reviews, specific guidelines have been developed (preferred reporting items for systematic reviews and meta-analyses; PRISMA; Moher, Liberati, Tetzlaff, & Altman, 2009). Due to this structured methodology, systematic reviews are regarded as one of the most rigorous forms of research.

Based on these guidelines, a systematic review should be conducted using the following five processes: 1) identify all relevant research, 2) select studies based on predefined inclusion/exclusion criteria, 3) assess study quality, 4) synthesise findings, 5) report and interpret results in an unbiased summary (Hemingway & Brereton, 2009).

The use of a systematic review methodology is not only relevant for the synthesis of results and findings relating to what is already “known,” but is effective in highlighting areas where an additional research focus or improvement is needed (Chalmers et al., 2002). Therefore, considering the lack of research investigating the use of accelerometers in children with intellectual disabilities, utilising this methodology will increase the knowledge-base relating to the measurement of physical activity in this population. Furthermore, this will allow specific gaps in existing literature to be identified which require further investigation in future research.

2.2.11 Research questions

This purpose of this study is to systematically review accelerometer use during field-based physical activity research involving children with intellectual disabilities. The research question to be examined within this chapter is:

RQ 1: Do the methods of accelerometer use employed in field-based physical activity research involving children with intellectual disabilities meet best practice guidelines?

2.3 Method

The PRISMA statement was used as the basis for this review. These guidelines include a flow diagram illustrating the four phases to be included and reported within a systematic review (identification, screening, eligibility, and included reviews) and a 27-item checklist which describes specific items to be reported in each section.

2.3.1 Search strategy

Relevant studies were identified from three sources: 1) published articles identified through a systematic search of electronic databases; 2) reference list/bibliography search; and 3) unpublished theses search.

1. (developmental adj (disab$ or disorder or difficult$)).tw.
2. (intellectual adj (disab$ or disorder or difficult$)).tw.
3. (learning adj (disab$ or disorder or difficult$)).tw.
4. (mental$ adj (retard$ or deficiency)).tw.
5. Acceleromet*.tw.
6. Accelerometry-based monitors.tw.
7. Physical activity measurement.tw.
8. Activity monitor.tw.
9. ActiGraph.tw.
10. MTI.tw.
11. CSA.tw.
12. Actical.tw.
13. Actiheart.tw.
14. Activpal.tw.
15. Tritrac.tw.
16. Uniaxial.tw.
17. Dualaxial.tw.
18. Triaxial.tw.
19. MVPA.tw.
20. Or/1-4
21. Or/5-19
22. 20 and 21

Limit 22 to (human and English language and child <unspecified age>

Figure 2.1. Embase search strategy used to identify studies which measured free-living physical activity using accelerometers in children and adolescents with intellectual disabilities

An electronic literature search was conducted to identify papers that used accelerometers to quantify physical activity in children with intellectual disabilities. Six electronic databases specific to biomedical and life sciences topics were searched (Medline, Embase, Cochrane Library, Web of Knowledge, PsycINFO, and PubMed) from 1990 up to, and including, May 2013. These databases were searched as they covered several topic areas which could include relevant articles, such as life sciences, health, biomedicine, and psychology. This search was limited to 1990 onwards as accelerometers were still in a developmental stage during the 1990's and were not widely used prior to this time (Troiano, 2005). The search strategy focussed on truncated population terms (e.g. developmental disability, intellectual disability, learning disability, and mental retardation) and accelerometer terms (e.g. accelerometry, activity monitor, ActiGraph, ActivPAL). Searches were limited to children (≤ 18 years), English language, and human. The full Embase search strategy, which was adapted for other databases, is presented in Figure 2.1.

Reference lists of relevant studies were hand searched for additional studies. Furthermore, the reference list of all review articles identified were also searched. Unpublished/grey literature were searched for using the Proquest Dissertations and Theses database and a Google Scholar search, with a search strategy adapted from that presented in Figure 2.1.

2.3.2 Study selection

An inclusion and exclusion criteria were developed as the basis for the study selection process to ensure that the included studies were appropriate to the aims of this review. These criteria were developed to identify studies which: 1) included a sample of ambulatory children with intellectual disabilities; and 2) measured free-living physical activity using accelerometers.

The inclusion criteria were studies which:

- used accelerometers to measure physical activity
- aimed to quantify levels of physical activity based on intensity, duration, or frequency in free-living settings
- included participants with intellectual disabilities

- included participants aged ≤ 18 years
- full-text articles.

The exclusion criteria were studies which:

- aimed to calibrate/validate accelerometers
- included a population with developmental disabilities, such as autism, but did not specify if intellectual disabilities were present
- included a non-ambulatory population.

As this study focussed on children with intellectual disabilities, it was important to develop appropriate disability-related criteria to ensure the inclusion of appropriate study samples. This was specifically relevant to studies which included samples with developmental disabilities or autism; although these are associated with intellectual disabilities, not all children with developmental disabilities or autism have intellectual disabilities. Therefore, unless studies explicitly reported that the study sample included children with intellectual disabilities, these studies were excluded. However, due to the limited research focussing on free-living physical activity in children with intellectual disabilities, studies which met the design and measurement criteria, but only a portion of the sample met the population criteria – e.g. only some participants had intellectual disabilities or the sample included children and adults with intellectual disabilities - were included. Accelerometers have been used in various types of physical activity research; however, many of the use decisions which this reviews aimed to investigate are specific to free-living measurements, e.g. wear days, valid days, and compliance strategies. Therefore, studies which used accelerometers in structured settings, such as organised activity classes, were excluded as these study designs do not require as in-depth use protocols.

After studies were identified using the search strategy, duplicate articles were removed. Studies for inclusion in the review were then identified using a 3-step selection process which involved a title and abstract review by one researcher

(AM), a full-text review by one researcher (AM), with the final decision on including/excluding articles based on discussion between two researchers (AM & CM).

2.3.3 Data extraction

2.3.3.1 Development of review criteria

The complexities of using accelerometers has been the focus of many previous studies in the measurement literature, which have been summarised in the previous sections of this chapter. However, for many years there was a lack of standardised guidelines to inform researchers, particularly those who were inexperienced in using accelerometers, on the best methods of use and the implications of use decisions. To address these gaps in knowledge, a scientific meeting was held in 2004, entitled “Objective Monitoring of Physical Activity: Closing the Gaps in the Science of Accelerometry” (Ward, Evenson, Vaughan, Rodgers, & Troiano, 2005). At this meeting, experts in the field of accelerometry presented a total of nine papers which covered various aspects of accelerometer use and suggested specific best practice recommendations; all these papers are presented in a supplement in Medicine and Science in Sports and Exercise (volume 37, supplement 11, 2005). The findings and recommendations from these papers were subsequently integrated into a single paper by Ward et al. (2005) to provide best practice guidelines for the following five areas:

1. Monitor selection, quality, and dependability
2. Monitor use protocols
3. Monitor calibration
4. Analysis of accelerometer data
5. Integration with other data sources.

Within these five general areas, guidelines were presented for 18 specific components of accelerometer use which should be considered when designing and conducting free-living physical activity measurements. The depth and detail within these best practice guidelines therefore could be utilised as criteria to further understand how accelerometers are used within physical activity research involving children with intellectual disabilities. Furthermore, although more recent best practice guidelines were published by Matthews, Hagströmer, Pober, and Bowles (2012), these guidelines focused on understanding and reporting accelerometer use for different population-based designs. As a result, this paper included less comprehensive guidelines than Ward et al. (2005), and did not include recommendations specific to interpreting accelerometer output, which has previously been noted as an area requiring further investigation in children with intellectual disabilities (Frey et al., 2008; Hinckson & Curtis, 2013). Therefore, the guidelines developed by Ward et al. (2005) were most relevant to the aims of this review.

2.3.3.2 Data extraction form

A data extraction form based on these best practice guidelines was developed by one researcher (AM) to obtain relevant data for review. This novel approach enabled the methods of accelerometer use in research involving children with intellectual disabilities to be assessed against the best practice recommendations from experts in the field of physical activity measurement. Although the original guidelines included 18 components, one of these, “analysing data”, within the theme of monitor calibration was not included as it was deemed too specific to calibration and therefore not relevant to the aims of this review. Based on this, 17 research components were used; these are presented in Table 2.1 with a summary of each specific research component and review criteria. The summary and criteria were formulated primarily from the information presented by Ward et al. (2005), however, the specific research articles were used for clarification and to ensure the accuracy of the guidelines reported. Table 2.2 is adapted from Ward et al. (2005) and summarises which of the five areas of accelerometer use the nine papers contributed to.

Table 2.1. Best practice guidelines, summary, and review criteria

Best practice guidelines	Summary	Review Criteria
Monitor selection, quality, and dependability		
1. Selecting instruments	With the variability of available accelerometers, e.g. in size, cost, data processing and data storage, and with no device viewed as superior, choice of device should be based on research purpose.	Rationale provided for choice of device
2. Assessing instrument quality and dependability	Instrument variance (i.e. coefficient of variability), reliability and validity should be tested before and after use	Population-appropriate coefficient of variability, validity and/or reliability evidence provided for device used
Monitor use protocols		
3. Using multiple monitors	Additional information and accuracy of using multiple monitors should be considered in relation to study population and participant burden	Rationale provided for number of monitors used

Table 2.1. Continued

Best practice guidelines	Summary	Review Criteria
4. Defining wearing days	Days of monitoring required should be based on population-specific calculations, e.g. based on ICC, with setting, resources, and research question considered	Seven days of monitoring required or appropriate justification for a shorter monitoring period provided
5. Determining monitor placement	Placement should be decided based on existing calibration equations, e.g. which placements do pre-existing equations exist for the study population, participant comfort, and manufacturer recommendations should be considered	Rationale provided for monitor placement

Table 2.1. Continued

Best practice guidelines	Summary	Review Criteria
6. Establishing field practices	Quality control measures should be employed throughout the period of accelerometer use, and during distribution and collection Investigator- or participant-based compliance strategies can encourage accelerometer wear	Inter-unit variation controlled and/or face-to-face accelerometer distribution and collection employed
7. Ensuring compliance	Investigator- or participant-based compliance strategies can encourage accelerometer wear	At least one of the following compliance techniques employed: log diary, reminder calls, information on proper wear, relapse prevention model, visual prompts, wear information given to teachers, coaches, etc., participant shown example of data output which indicates when device is not worn, or incentives offered

Table 2.1. Continued

Best practice guidelines	Summary	Review Criteria
Monitor calibration		
8. Predicting energy expenditure	Equations employed should be calibrated against a gold standard measure and should account for different patterns of activity	Cut points or equations calibrated against gold standard measure of energy expenditure (calorimetry, doubly-labelled water or direct observation) during various patterns of activity
9. Using individual calibration equations	In small scale studies, individual calibration equations should be used	Individual calibration equations developed for each participant
10. Constructing group calibration equations	In larger-scale studies, equations should be used which were calibrated in similar population with a representative sample, with population appropriate activities conducted.	Equations calibrated in population matched for participant characteristics during various activities

Table 2.1. Continued

Best practice guidelines	Summary	Review Criteria
11. Determining epoch length	Epoch length should be determined based on the study population and their activity characteristics, e.g. children generally engage in short bouts of high intensity activity	Rationale provided for epoch length used
Analysis of accelerometer data		
12. Defining a day	The time period of monitoring required to constitute a day can vary, e.g. from 12 to 24 hours, with participant age, weekday/weekend monitoring, and activities to be considered	Definition of what constitutes a day of measurement for inclusion in analysis
13. Handling incomplete data	Activity is not always measured over a consistent time period and data can be missing, therefore decisions should be made to try to prevent under/overestimation of activity	Method specified for dealing with missing or incomplete data, e.g. imputation.

Table 2.1. Continued

Best practice guidelines	Summary	Review Criteria
14. Creating reporting standards	Due to a lack of standardised procedures for data processing and reduction, decision rules need to be clearly stated	Clarity of data collected and assumptions made, e.g. identifying wearing period, minimum wear required for valid day, spurious data, computing variables and aggregating days, extracting bouts of MVPA
15. Determining bouts	Reporting bouts of activity can allow for more accurate estimations of MVPA and understanding/comparison of activity patterns	Measurement of total activity duration and number and length of bouts per day
16. Handling spurious data	Data should be cleaned for implausible data points with accelerometers checked for malfunction, error, or participant tampering	Comment made on how spurious data is dealt with, e.g. setting data points to “missing”

Table 2.1. Continued

Best practice guidelines	Summary	Review Criteria
Integration with other data sources		
17. Integration with other data sources	Use of multiple technologies can increase the quality and breadth of data collected	Use of other data source, e.g. HR or GPS

Adapted from McGarty et al. (2014)
Global positioning system (GPS); Heart rate (HR); Moderate to vigorous physical activity (MVPA)

Table 2.2. Contribution of the nine supplement papers to the best practice guidelines

Study	Monitor selection, quality, & dependability	Monitor use protocol	Monitor calibration	Analysis of accelerometer data	Integration with other data sources
Catellier et al. (2005)				X	
Chen & Bassett (2005)	X			X	
Freedson et al. (2005)			X		
Masse et al. (2005)				X	
Matthews (2005)	X	X	X		
Rodriguez et al. (2005)					X
Strath et al. (2005)	X	X			X
Trost et al. (2005)	X	X	X		
Welk (2005)	X	X	X	X	

Adapted from Ward et al. (2005)

All data were extracted by one researcher (AM) and collated in an Excel spreadsheet (Version 14.0; Microsoft, 2010). Descriptive statistics were extracted for each study relating to participants (sample size, age, and level of intellectual disabilities), study characteristics, and type of accelerometer used. Review data relating to accelerometer use were extracted based on the 17 components presented in Table 2.1.

2.3.4 Data analysis

Data were analysed by one researcher (AM) based on a dichotomised criteria of whether the study did or did not meet the best practice criteria, as presented in Table 2.1. Accelerometer use was investigated on a study-by-study and component-by-component basis. The total percentage of review criteria met by each study was calculated, with the data synthesised to calculate the percentage of studies which met each of the 17 review components. To account for studies which did not include aspects of the review criteria, as denoted by "n/a" in Table 2.3, data were presented as percentages to allow comparison between individual studies and comparison between components of the best practice guidelines.

2.4 Results

2.4.1 Study selection

A total of 429 articles were initially identified using the three sources; 428 from the electronic search (Medline, n = 16; Embase, n = 14; Cochrane Library, n = 0; Web of Knowledge, n = 341; Psycinfo, n = 37; PubMed, n = 20) and one from the thesis database search. The full study selection process is presented in Figure 2.2. After the removal of duplicate articles, title and abstract screening was conducted by one researcher (AM) on the remaining 367 articles. Thirty articles were identified from the title and abstract screen, which then underwent a full-text review by the same researcher (AM). Based on the full-text review, an initial list of 10 possible inclusion and 20 possible exclusion studies was developed. This list was discussed with a second researcher (CM) to ensure that studies were correctly included and excluded. After discussion, it was agreed that the 20 possible exclusion articles did not meet the inclusion criteria due to inappropriate population or the collection of physical activity data to measure validity and reliability and not to quantify activity levels during field-based research. Based on this, the inclusion of 10 studies for review was finalised.

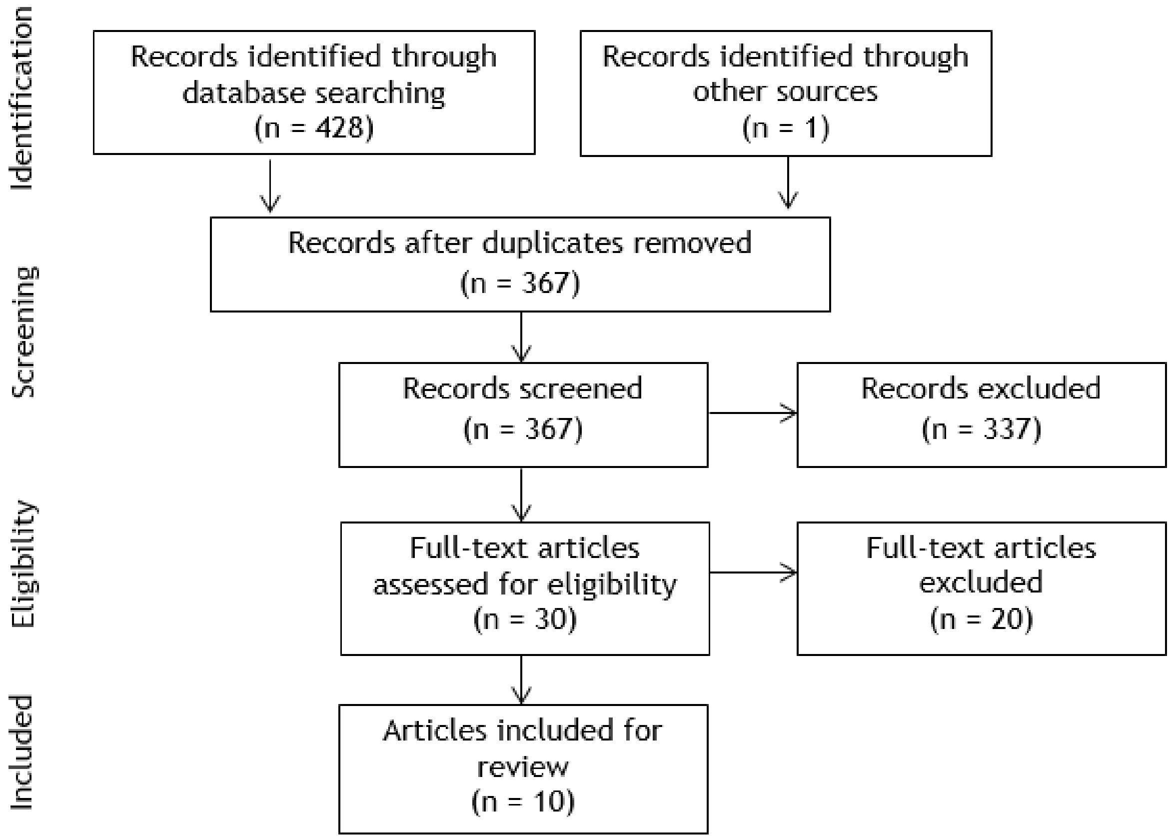


Figure 2.2. Flow chart of the study selection process

2.4.2 Study characteristics

Eight studies were carried out in the USA, one in Australia, and one in the UK. Participants (n = 677) aged from 3 to 70 years, with reported intellectual disabilities ranging from mild to severe. All studies, with the exception of Phillips and Holland (2011; 12 to 70 years), included only child and adolescent participants (3 to 17 years). Seven studies quantified total physical activity during daily free-living (Esposito et al., 2012; Foley, Bryan, & McCubbin, 2008; Lloyd, 2008; Phillips & Holland, 2011; Shields et al., 2009; Ulrich, Burghardt, Lloyd, Tiernan, & Hornyak, 2011; Whitt-Glover et al., 2006), two studies quantified school recess and classroom based physical activity (Horvat & Franklin, 2001; Lorenzi et al., 2000), and one quantified after school physical activity (Foley & McCubbin, 2009). Seven accelerometer brands were used: ActiGraph (Phillips & Holland, 2011), Actical (Esposito et al., 2012; Lloyd, 2008; Ulrich et al., 2011), Actitrac (Whitt-Glover et al., 2006), Actiwatch (Foley & McCubbin, 2009; Foley et al., 2008), Caltrac (Lorenzi et al., 2000), TriTrac R3D (Horvat & Franklin, 2001), and RT3 (Shields et al., 2009).

2.4.3 Monitor selection, quality, and dependability

Five studies clearly reported selecting an accelerometer based on study design or outcomes; Table 2.3, which is adapted from McGarty, Penpraze, and Melville (2014), reports the attainment of each component of the review criteria on a study-by-study basis, with Figure 2.3 illustrating results on a component-by-component basis.

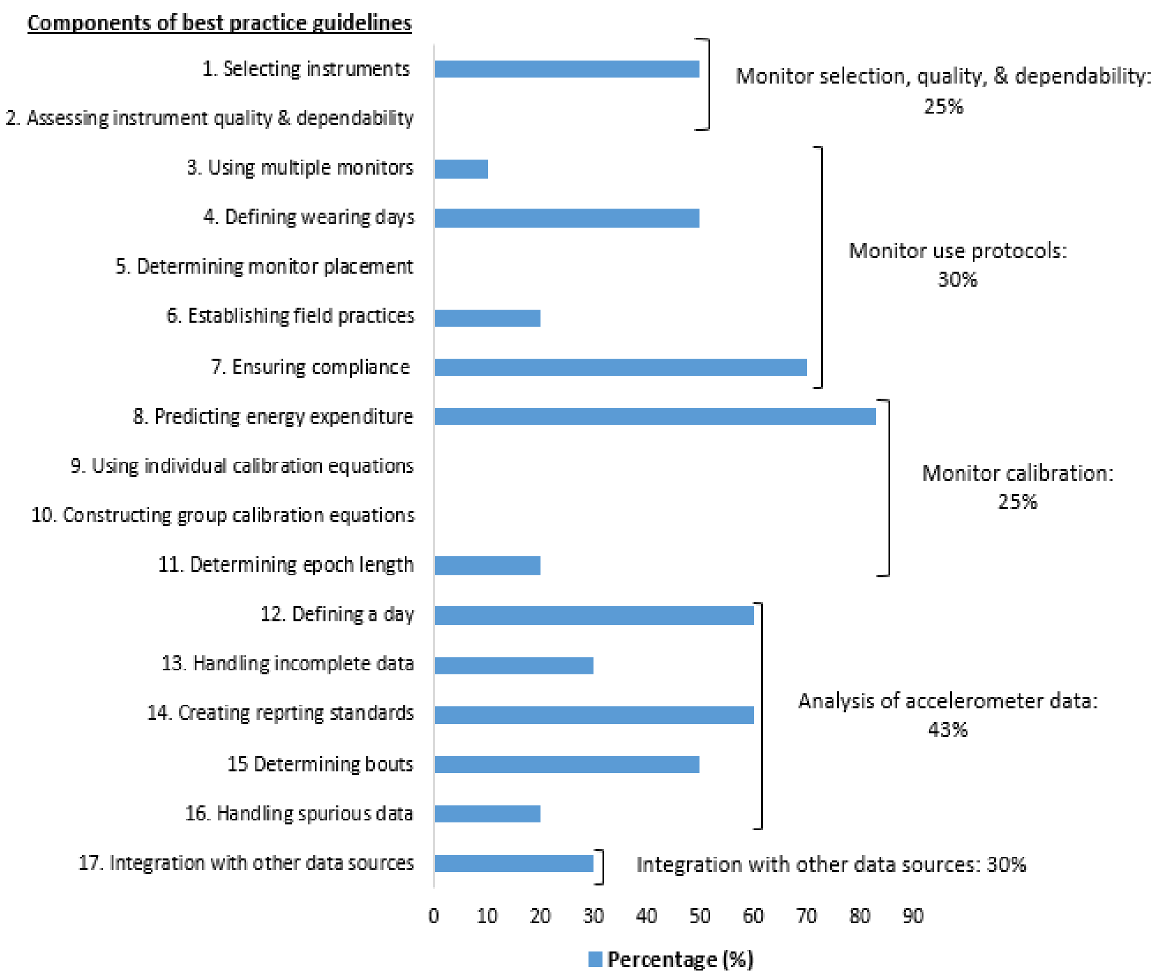


Figure 2.3. Total percentage of best practice guidelines achieved by review studies

Three studies presented a general rationale for the benefits of accelerometer use, citing the capability of this method to detect movement across various planes (Esposito et al., 2012; Horvat & Franklin, 2001) and advantages of accelerometry in comparison with other methods of physical activity assessment, e.g. measurements not affected by stress or self-report bias, and non-compliance more easily quantified (Shields et al., 2009). Two studies referenced previous accelerometer use in children with intellectual disabilities as a rationale for device selection (Foley & McCubbin, 2009; Foley et al., 2008). However, no study stated any measure of instrument quality or dependability that was specifically established in children with intellectual disabilities.

2.4.4 Monitor use protocols

Nine studies used a single accelerometer, without providing a rationale for doing so. Lorenzi et al. (2000) used two monitors to enable comparisons between constant and individualised methods of device programming relating to participant characteristics, although only one device was used for data analysis. Monitoring periods were: 16 minutes (Horvat & Franklin, 2001; Lorenzi et al., 2000), 4 days (Lloyd, 2008), 5 days (Foley & McCubbin, 2009), and 7 days (Esposito et al., 2012; Foley et al., 2008; Phillips & Holland, 2011; Shields et al., 2009; Ulrich et al., 2011, Whitt-Glover et al., 2006), with only two studies providing a rationale for wear days. These rationales were based on previous use (Foley et al., 2008) and reliability (Shields et al., 2009).

The hip was the most popular placement, with four studies attaching the accelerometer to the right hip (Esposito et al., 2012; Lloyd, 2008; Phillips & Holland, 2011; Ulrich et al., 2011), one study using the right and left hip (Lorenzi et al., 2000), and three studies not specifying right or left hip (Horvat & Franklin, 2001; Shields et al., 2009; Whitt-Glover et al., 2006). Two studies placed the device on the wrist of the participant’s non-dominant hand, which was in line with the device manufacturer’s recommendations (Foley & McCubbin, 2009; Foley et al., 2008).

Foley and McCubbin (2009) and Foley et al. (2008) reported establishing field practices by replacing the accelerometer twice during the monitoring period to reduce systematic error associated with individual devices. Furthermore, Foley et al. (2008) reported testing device calibration before and after data collection. Strategies employed to aid compliance were the use of daily monitoring logs (Esposito et al., 2012; Phillips & Holland, 2011; Shields et al., 2009; Ulrich et al., 2011), familiarisation and orientation sessions (Horvat & Franklin, 2001; Lorenzi et al. 2000), and instructing parents to regularly check monitor wear and placement (Whitt-Glover et al., 2006).

2.4.5 Monitor calibration

Of the studies which aimed to predict activity intensity, five studies used cut points that had been calibrated against indirect calorimetry, with only one study (Whitt-Glover et al., 2006) not providing sufficient evidence of a gold standard measure used for calibration. Five different intensity cut points were used (Puyau et al., 2002; Rowlands, Thomas, Eston, & Topping, 2004; Trost et al., 2002; Strauss, Rodzilsky, Burack, & Colin, 2001; Pfeiffer, McIver, Dowda, Almeida, & Pate, 2006). However, none of these had been calibrated specifically for children with intellectual disabilities. The studies which only reported physical activity data in vertical axis counts (Foley & McCubbin, 2009; Foley et al. 2008; Lorenzi et al., 2000) and vector magnitude counts (Horvat & Franklin, 2001) all stated the error associated with energy expenditure or intensity calculations as a rationale.

The most commonly used epoch length was 15 seconds (Esposito et al., 2012; Foley & McCubbin, 2009; Foley et al., 2008; Lloyd, 2008; Ulrich et al., 2011), with the use of 5-second (Phillips & Holland, 2011), 30-second (Whitt-Glover et al., 2006), and 60-second epochs also reported (Horvat & Franklin, 2001; Lorenzi et al., 2000; Shields et al., 2009). With the exception of two studies (Esposito et al., 2012; Lloyd, 2008), which both chose a 15-second epoch based on children’s activity tempo and behaviours, no other study stated a rationale for epoch length used.

Table 2.3. Standard of accelerometer use based on best practice

Best Practice Guidelines	Study									
	Esposito et al. (2012)	Foley & McCubbin (2009)	Foley et al. (2008)	Horvat & Franklin (2001)	Lloyd (2008)	Lorenzi et al. (2000)	Phillips & Holland (2011)	Shields et al. (2009)	Ulrich et al. (2011)	Whitt-Glover et al. (2006)
Monitor selection, quality, and dependability										
1. Selecting instruments	yes	yes	yes	yes	-	-	-	yes	-	-
2. Assessing instrument quality and dependability	-	-	-	-	-	-	-	-	-	-
Monitor use protocols										
3. Using multiple monitors	-	-	-	-	-	yes	-	-	-	-
4. Defining wearing days	yes	-	yes	-	-	-	yes	yes	yes	-
5. Determining monitor placement	-	-	-	-	-	-	-	-	-	-
6. Establishing field practices	-	yes	yes	-	-	-	-	-	-	-

Table 2.3. Continued

Best Practice Guidelines	Study									
	Esposito et al. (2012)	Foley & McCubbin (2009)	Foley et al. (2008)	Horvat & Franklin (2001)	Lloyd (2008)	Lorenzi et al. (2000)	Phillips & Holland (2011)	Shields et al. (2009)	Ulrich et al. (2011)	Whitt-Glover et al. (2006)
7. Ensuring compliance	yes	-	-	yes	-	yes	yes	yes	yes	yes
Monitor calibration										
8. Predicting energy expenditure	yes	n/a	n/a	n/a	yes	n/a	yes	yes	yes	-
9. Using individual calibration equations	-	n/a	n/a	n/a	-	n/a	-	-	-	-
10. Constructing group calibration equations	-	n/a	n/a	n/a	-	n/a	-	-	-	-
11. Determining epoch length	yes	-	-	-	yes	-	-	-	-	-
Analysis of accelerometer data										
12. Defining a day	yes	-	-	yes	-	yes	yes	yes	yes	-

Table 2.3. Continued

Best Practice Guidelines	Study									
	Esposito et al. (2012)	Foley & McCubbin (2009)	Foley et al. (2008)	Horvat & Franklin (2001)	Lloyd (2008)	Lorenzi et al. (2000)	Phillips & Holland (2011)	Shields et al. (2009)	Ulrich et al. (2011)	Whitt-Glover et al. (2006)
13. Handling incomplete data	-	-	yes	-	-	-	yes	-	-	yes
14. Creating reporting standards	yes	yes	-	-	-	-	yes	yes	yes	yes
15. Determining bouts	-	n/a	n/a	n/a	-	n/a	yes	yes	-	yes
16. Handling spurious data	-	yes	-	-	-	-	yes	-	-	-
Integration with other data sources										
17. Integration with other data sources	-	yes	-	yes	-	yes	-	-	-	-
Percentage of review criteria met (%)	41	31	31	31	12	31	47	41	29	24

Yes; Study met the best practice guidelines for specific research component
-; Study did not meet the best practice guidelines for specific research component
n/a; Research component not applicable to study
Adapted from McGarty et al. (2014)

2.4.6 Analysis of accelerometer data

To be included in the analysis, three studies specified a minimum of 10 hours of wear per day (Esposito et al., 2012; Phillips & Holland, 2011; Ulrich et al., 2011), with Esposito et al. (2012) and Ulrich et al. (2011) further specifying that this had to achieved on 4 out of 7 days, including one weekend day. Shields et al. (2009) did not report a required number of hours of wear per day, but specified that 6 out of 7 days of monitoring was required. For dealing with missing or incomplete data, two studies (Foley et al., 2008; Whitt-Glover et al., 2006) excluded the measurements from any participants who returned incomplete data, whereas one study (Phillips & Holland, 2011) asked participants to wear the accelerometer for an additional seven days.

Of the studies which met the criteria for creating reporting standards, four noted wear time (Esposito et al., 2012; Phillips & Holland, 2011; Shields et al., 2009; Ulrich et al., 2011), which ranged from 12.38 to 14.23 hours per day. Two studies also noted the average number of wear days as 6.3 days (Whitt-Glover et al., 2006) and 6.8 days (Shields et al., 2009) for a 7-day monitoring period. Two studies (Foley & McCubbin, 2009; Phillips & Holland, 2011) reported checking for spurious data, with Phillips and Holland (2011) specifying that non-wear time was defined as ≥ 10 minutes of zero counts.

Three studies presented information relating to bouts. Shields et al. (2009) reported that for children with a mean age of 11.7 years, the average bouts of moderate to vigorous and vigorous intensity activity lasted 2.8 minutes (SD = 0.6 minutes) and 2.0 minutes (SD = 0.6 minutes), respectively. Whitt-Glover et al. (2006) reported higher bout lengths in a sample of younger children (M = 6.6 years, SD = 2.1 years), with the average daily bouts of moderate to vigorous and vigorous intensity activity lasting 10.7 minutes (SD = 2.7 minutes) and 2.5 minutes (SD = 2.3 minutes), respectively. Furthermore, the longest bouts of moderate to vigorous and vigorous intensity reported by Whitt-Glover et al. (2006) were 22.1 minutes (SD = 9.6 minutes) and 6.9 minutes (SD = 4.6 minutes), respectively. A different method of including bouts was reported by Philips and Holland (2011), who aimed to investigate the attainment of physical activity guidelines, therefore only included data which was in ≥ 10 minute bouts of moderate to vigorous intensity and therefore deemed sufficient to be health-enhancing.

2.4.7 Integration with other data sources

To increase the breadth of data collected, two studies combined accelerometry with the SOAL observational checklist and heart-rate measurements (Horvat and Franklin, 2001; Lorenzi et al., 2006) and one used a proxy activity log based on the Activitygram to determine which activities children engaged in and when they did so (Foley & McCubbin, 2009).

2.5 Discussion

The purpose of this study was to systematically review how accelerometers are used in free-living physical activity research involving children with intellectual disabilities. Accelerometers are complex devices with many decisions facing accelerometer users, which can put a high burden on researchers, particularly those with minimal experience of using accelerometers. Therefore, if current use can be better understood in children with intellectual disabilities, research areas can be identified which can improve how accelerometers are used in this population.

This systematic review illustrates the variance of accelerometer use in physical activity research involving children and adolescents with intellectual disabilities. A variety of methods were reported in each of the five themes, with an array of use decisions described within each research component. However, the majority of these did not meet the criteria based on best practice guidelines (Ward et al., 2005). A lack of measurement specific research involving children with intellectual disabilities was also highlighted. This is limiting the scope for decisions to be made based on population-specific research and, therefore, negatively impacting on the reliability and validity of results.

2.5.1 Methods of use

No studies within this review met all the criteria for best practice guidelines. The percentage of review criteria met by individual studies ranged from 12% (Lloyd, 2008) to 47% (Phillips & Holland, 2011). There was also high variability in the attainment of the 17 components of the best practice guidelines, which ranged from 0% to 83%. The highest percentage of review criteria met was within the "analysis of accelerometer data" theme (43%), with the themes of "monitor selection, quality, and dependability" and "monitor calibration" the lowest, with 25% of review criteria met. This highlights the variability and overall lack of research being conducted in line with best practice guidelines.

The most common reason for studies not meeting the review criteria was a failure to fully describe or provide a rationale for accelerometer use decisions. As discussed in the introduction to this chapter, there are no areas of accelerometer use where a universal consensus has been reached on how to collect, reduce, or interpret data, with methods of use affected by various factors, including study outcomes, age of participants, and sample size. It is therefore the responsibility of researchers to choose the most appropriate device and methods of data collection and analysis for their study (Trost et al., 2005). Furthermore, to allow comparison between studies and to enable critical evaluation on the validity of the conclusions made, it is important that researchers fully describe and justify methodological decisions (Cliff et al., 2009; Freedson et al., 2012).

Providing in-depth reporting of accelerometer use in publications may not always be feasible though, as many journals impose length limitations on articles. Despite article limitations, the availability of online appendices for journals is increasing, which should be utilised to overcome word limits and allow a full description of methodological details (Freedson et al., 2012). That said, however, Lloyd (2008) was a doctoral thesis and the only study in this review not affected by publication word limits; yet, Lloyd (2008) achieved the lowest percentage of criteria met, which suggests that publication word limits is not the only factor contributing to the poor reporting of accelerometer use identified within this review.

Another possible reason for the low standards of use/reporting is that Ward et al. (2005) developed the best practice guidelines to address a gap in the literature, specifically the availability of combined and standardised guidelines. Therefore, it could be assumed that studies published prior to these guidelines, i.e. pre-2005, would achieve fewer of the review criteria due to the lower number of studies related to accelerometer use available to researchers at that time. However, the two studies published pre-2005 both achieved 31% of the review criteria, which is only slightly below the combined average of 32% (Horvat & Franklin, 2001; Lorenzi et al., 2000). On the other hand, the apparent lack of effect on studies published post-2005 could also suggest that these guidelines have had limited effects on improving the standard of accelerometer use and reporting, at least in research which is conducted in children with intellectual disabilities.

The lack of description of accelerometer use somewhat limits the scope for discussion on how researchers approach the measurement of physical activity using accelerometers in children with intellectual disabilities. Although various methods of use were employed, few of these decisions were reported to be specifically related to children with intellectual disabilities. This could be in part due to the lack of measurement research conducted in children with intellectual disabilities. As highlighted in Figure 2.3, however, no studies achieved the criteria related to the reliability, validity, and calibration of devices. Considering these are fundamental principles of measurement, the use of accelerometers with no information provided on validity raises questions on the validity of the data collected, which needs consideration.

2.5.2 Effect of use decisions

Considering the varying accelerometer use reported within this review, this section will discuss some of the possible effects on reliability, validity, and overall study outcomes as a result of use decisions.

2.5.2.1 Monitor selection, quality, and dependability

Monitor selection, quality, and dependability is focussed on selecting a device which exhibits intra- and inter-unit reliability, validity, and is suitable for the research aims and study sample. However, as no device is deemed superior to another, accelerometer selection is generally based upon device characteristics, such as: cost, software, technical support, user-friendliness, or participant reactivity and tampering (Trost et al., 2005; Ward et al., 2005).

Seven different accelerometer devices were used within the studies reported in this review. However, only five studies provided a rationale for choosing a device. The rationales provided were based upon general advantages of accelerometry over other measurement methods, rather than the benefits of one device over another. Furthermore, only two studies (Foley & McCubbin, 2009; Foley et al., 2008) considered the feasibility of the device in children with intellectual disabilities when selecting an accelerometer.

Although it has been widely cited that no device is superior to another, in terms of unit reliability and validity, this finding is not consistent across all studies. Of the previous research which investigated the technical reliability of the devices used within these review studies, i.e. reliability measured using mechanical shakers or oscillating devices, high variability has been reported within and between devices. Numerous studies report varying ICC reliability and/or coefficients of variation (CV) for the ActiGraph (ICC = .84 to .93 and <1.0% to 4.4%), TriTrac R3D (ICC = .97), and RT3 devices (0.2% to 56.2%; Brage, Brage, Wedderkopp, & Froberg, 2003; Kochersberger, McConnell, Kuchibhatla, & Pieper, 1996; Metcalf, Curnow, Evans, Voss, & Wilkin, 2002; Powell, Jones, & Rowlands, 2003). This wide variance has also been noted for inter-instrument reliability during laboratory-based walking between the ActiGraph (r = .80 and 8.9%), TriTrac R3D (r = .73 and 9.4%), and Actical (r = .62 and 20%), which further illustrates the varying reliable between devices which were used in studies included in this review (Welk, Schaben, & Morrow, 2004).

The findings from these studies suggest that researchers need to consider the reliability of devices to limit the within- and between-device effects on the validity of study outcomes. However, the devices reported within this review have since, in some cases, undergone several updates and changes in the fundamental accelerometer mechanism, therefore the intra- and inter-unit reliability discussed may be less relevant to newer devices. This has empirical support, as newer versions of the ActiGraph device have higher inter-unit reliability than older versions (Rothney, Apker, Song, & Chen, 2008). This improvement may not be consistent across all devices though, as the rate of device redevelopment in line with emerging technology varies between device brands; for example, the Actical accelerometer was last updated in 2007 and the

RT3 was only recently updated to the RT6, although currently no published studies report using the RT6.

In addition to reliability, another consideration should be the accuracy of the device to measure movement. Although the research is limited, there is evidence showing that the validity of accelerometer counts against a criterion measure varies between devices for children with intellectual disabilities, i.e. reported validity ranges from r = .10−.76, depending on device (Kozub, 2003; Taylor & Yun, 2006). Therefore, researchers need to consider the validity of devices for the study population in addition to reliability and feasibility. However, no studies in this review provided sufficient or population-specific validity evidence related to the device used. As a result, this raises questions on the validity of the data collected. It is important to note, however, that reliability- and validity-related research has not been conducted specifically for children with intellectual disabilities for all devices used within the review studies. Therefore, it was not possible for some studies to achieve the review criteria. On the other hand, no studies within this review discussed any measurement or feasibility-related effects due to device selection. This suggests that researchers are selecting devices which are suitable to both participants and study outcomes.

In summary, the varying validity and reliability evidence between devices in children with intellectual disabilities is in contrast to the general consensus in research involving typically developing children that no device is superior to another in terms of validity/reliability. Therefore, it is important that future research, especially review articles, highlight rather than de-emphasise the importance of reliability and validity to help ensure that data is collected using the most valid, as well as feasible, methods. Furthermore, the lack of existing literature affected the ability of researchers to report population-appropriate validity evidence, and lowered the number of studies achieving the review criteria.

2.5.2.2 Monitor use protocols

The theme of monitor use protocols covers data collection and the accelerometer use decisions associated with this period. The best practice guidelines within this theme recommend how to collect the most relevant and valid data for the study outcomes, as well as techniques to increase participant compliance and increase the quality of the data collected.

The use of multiple monitors increases the depth of data collected and the accuracy of estimating energy expenditure and intensity from accelerometry (Strath, Brage, & Ekelund, 2005). However, for most study designs, the advantages of multiple monitors does not outweigh the additional participant burden (Trost et al., 2005). Only one study in this review used two monitors. This decision was based on the outcomes of the study, which aimed to compare different methods of data processing, and therefore was an appropriate use decision.

As discussed in the introduction to this chapter, accelerometer placement can affect results. Yet, as some devices are placement-specific, this is not always a decision that researchers need to make. With the exception of two studies which used devices that were placement specific to the wrist, all other studies used the hip placement, which provides the highest level of validity and feasibility for the chosen devices. Therefore, the only effect of placement will be the direct comparability between studies using different placements. Nevertheless, although these placements were most suitable, no studies specified a rationale as to why the placement was chosen and therefore no studies met the review criteria. For this reason, interpretation of these results requires caution as the placements used were the most valid and feasible for each study but the lack of reporting was the reason for not meeting the review criteria. This highlights the importance of full and accurate reporting of the methods of use. However, it also highlights a limitation with the developed review criteria, as the negative effects of the use decisions are limited, even though no studies met the criteria.

The monitoring times reported in this study varied greatly, from 16 minutes to 7 days. This large discrepancy is due to the different aims of each study and the environment in which physical activity was measured. For the studies which reported measuring physical activity for multiple days, the aims included quantifying total free-living physical activity. On the other hand, Horvat and Franklin (2001) and Lorenzi et al. (2000) only aimed to measure free-living activity during school-time and therefore the monitoring periods were shorter.

Of the studies aiming to measure total free-living activity, only Foley and McCubbin (2009) and Lloyd (2008) did not meet the 7-day wear criteria and did not provide a rationale for the reliability of a shorter monitoring period. Therefore, the data collected may not be a valid representation of the physical activity behaviours of the study sample.

As the number of monitoring days increase, participant compliance with wearing the device decreases (Penpraze et al., 2006). Therefore, the use of compliance strategies is important in increasing the number of participants who return sufficient data to be included in the analysis. However, the most comprehensive compliance strategies were reported by Horvat and Franklin (2001) and Lorenzi et al. (2000), both of which used the shortest monitoring periods of 16 minutes. Both these studies held orientation sessions on how and when to wear the device. Although this is an additional research burden, both these studies report that this was an effective compliance strategy, therefore this strategy should be considered in future studies. The other compliance strategy reported was the use of parental log diaries to note any times when the device was removed and the reason for removal (Esposito et al., 2012; Phillips & Holland, 2011; Shields et al., 2009; Ulrich et al., 2011). However, none of these studies reported any data obtained from the log diaries. Phillips and Holland (2011) was the only study to provide details on how the log diary data was used, specifying that it was not included in the analysis due to the subjective interpretation required. Although, the authors did report that comparison between the log diary and accelerometer counts suggested that physical activity was only minimally underestimated, but did not specify which method estimated the lower level of activity. Therefore, as the data obtained from log diaries is not used in the analysis, further consideration is needed to decide if this additional parental burden is necessary. Furthermore, proxy-report measures have limited validity, which is due to the subjective design and that it is not always feasible for a parent to capture all the required wear or activity behaviours of children (Corder et al., 2008).

Compliance strategies are participant-focussed and aim to increase accelerometer wear and therefore decrease missing data. On the other hand, field practices are device-focussed and are used to increase the quality and reliability of accelerometer data, thus reducing the likelihood of spurious or missing data due to device malfunction or loss. This review identified three studies which employed field practices. Foley et al. (2008) and Foley and McCubbin (2009) replaced devices twice during the 7-day measurement period to reduce measurement error, with Foley et al. (2008) additionally calibrating devices prior to use. This will increase the validity and reliability of results by reducing measurement error caused by intra-unit variation. However, this adds an increased burden on researchers, although the smaller sample sizes in these studies may have made this a feasible field-practice. Whitt-Glover et al. (2006) aimed to increase the reliability of data collected by asking parents to fit the device and regularly check it was being worn correctly, which instead increases the burden on parents during data collection. Lloyd (2008) was the only study in this review which specified that data were excluded from the analysis due to device malfunction and loss. This study did not employ any compliance or field-practices, which was a limitation of the accelerometer methods used; however, the full reporting of errors encountered was an advantage of this study. Freedson et al. (2012) discussed the importance of fully reporting difficulties encountered with accelerometer compliance, malfunction, and loss. This could be additionally important in research involving children with intellectual disabilities to increase our understanding of accelerometer wear and effective and feasible field practices.

In summary, a difficulty encountered by researchers when deciding upon monitor use protocols is the lack of standardised guidelines for accelerometer use (Freedson et al., 2012). However, considering accelerometer use when designing a study is important to ensure that a high standard of accelerometer data is collected. This review highlights that various use protocols have been used in research involving children with intellectual disabilities. However, it is not possible to determine the effect of more stringent use protocols, such as compliance strategies and field practices, as data related to these components was either not used in the analysis or the effects on data collected was not fully reported.

2.5.2.3 Monitor calibration

The theme of monitor calibration is focussed on the decisions and methods used to translate raw accelerometer count data into a physiologically meaningful outcome, such as energy expenditure or physical activity intensity. Calibrating accelerometers for the estimation of physiological outcomes based on biomechanical accelerations is a complex process. Due to the interactions between physiological factors and accelerometry, cut points and equations are not valid when generalised to populations different from the original calibration sample (Freedson et al., 2012).

When using cut points or equations to interpret accelerometer output, the validity of these data interpretation methods is dependent on the use of an appropriate criterion measure during calibration (Ward et al., 2005). The only study within this review that did not provide sufficient evidence that the cut points used were established against an appropriate criterion measure was Whitt-Glover et al. (2006), which did not reference an original calibration study for the chosen cut points. Of the studies in this review which applied group intensity cut points, none of these cut points were established in children with intellectual disabilities. On closer examination of the literature, however, it is apparent that population-specific intensity cut points or energy expenditure prediction equations have not been established for children with intellectual disabilities.

This lack of population-specific cut points could have significant effects on results. As children with intellectual disabilities have lower levels of cardiorespiratory fitness, generalising cut points that were established in a population with higher maximal oxygen uptake ($\dot{V}O_{2max}$) will underestimate the activity levels of children with intellectual disabilities (Pitetti, Yarmer, & Fernhall, 2001). Furthermore, as numerous child-specific cut points have been developed, there is no guidance as to which cut points are most valid for use in children with intellectual disabilities. The choice of cut points can have significant and clinically meaningful effects on intensity estimates, and therefore further research is needed to investigate the effect of cut points (Reilly et al., 2008).

For the six studies within this review which estimated intensity, five different sets of cut points were used (Puyau et al., 2002; Rowlands et al., 2004; Trost et al., 2002; Strauss et al., 2001; Pfeiffer et al., 2006). However, due to different devices used, there is little scope within this review to investigate the effect that the cut points used has on the estimation of time spent in each physical activity intensity. Further investigation into the effect of cut points, specifically the effect on intensity estimations, could be used to increase the awareness of researchers on the effects of cut points and highlight the need for consideration of use decisions. Previous reviews on physical activity in children and adolescents with intellectual disabilities discuss this lack of population-specific validity evidence and recommend that further measurement research is conducted for this population (Cervantes & Porretta, 2010; Hinckson & Curtis, 2013). Therefore, to increase the validity of activity intensity estimates in children with intellectual disabilities, it is important that population-specific cut points are calibrated.

Another method of ensuring validity would be to individually calibrate accelerometer output for each participant in a study (Welk, 2005). Theoretically, this would eliminate the error caused by individual behavioural and physiological differences which influence calibration, such as age, body fatness, or fitness level, and would prevent the need for a representative sample to be recruited. In practice, however, this method would only be feasible in small-scale studies. A more practical, or quasi, approach to individual calibration could be to develop cut points in a sub-set of a larger study sample, similar to that done by Pulsford et al. (2011), as part of the Millennium Cohort Study.

An alternative method employed by four studies within this review was not to employ cut points or equations but instead report raw counts as the primary study outcome. The use of this outcome was appropriate for these studies as they aimed to compare activity between children with and without intellectual disabilities (Foley & McCubbin, 2009; Foley et al., 2008; Lorenzi et al., 2000) or compare the activity levels in different environments (Horvat & Franklin, 2001). An advantage of reporting counts is that study results are not affected by the error associated with cut points or equations. Not without limitations, however, as previously discussed, accelerometer counts are non-comparable between devices, which limits the consolidation of research and comparison between studies. It would therefore be beneficial for studies to report counts and intensity/energy expenditure estimates.

Another important consideration when estimating activity intensity is the epoch length used. There is no consensus on the optimal epoch length for children, with practical considerations, such as device storage, being important considerations when deciding upon the epoch length to be used. However, numerous studies suggest that 60-second epochs are too long and instead advocate for the use of shorter epochs to more accurately capture the sporadic activity patterns of children and prevent physical activity intensity being underestimated (Corder et al., 2008; Nilsson et al., 2002; Rowlands et al., 2006; Trost et al., 2005; Welk et al., 2000). Due to the lack of consensus, the review criteria was that a rationale was provided for the chosen epoch, rather than specifying the epoch length to be used. Only two studies provided a rationale for the use of a 15-second epoch, which was that this shorter epoch will more accurately capture the sporadic nature of children's activity (Esposito et al., 2012; Lloyd, 2008). Furthermore, only two studies used a 60-second epoch, with the remaining epochs ranging from 1-second to 30 seconds.

The wide range of epochs used could affect intensity estimates, with shorter epochs capturing more high intensity activity than longer epochs (Corder et al., 2008; Trost et al., 2005; Welk et al., 2000). However, due to the different aims and environments in which activity was measured between studies, it is not possible to make inferences regarding the effect of epoch on intensity estimations within this review. Furthermore, the studies which used 60-second epochs only reported data in the form of raw counts, therefore the use of a longer epoch in these studies will not affect intensity estimates (Horvat & Franklin, 2001; Lorenzi et al., 2000). Another limitation with this review is that the attainment of the review criteria for epoch length, which was 20%, suggests that the epoch lengths used were inappropriate. However, as all studies which report intensity estimates used ≤ 15-second epochs, with the exception of Whitt-Glover et al. (2006) which used 30-second epochs, these epochs are short enough to capture children's sporadic activity. Therefore, the results relating to the attainment of the review criteria underestimates the amount of research being conducted in line with best practice guidelines, and more specifically refers to the lack of full reporting of accelerometer use. Furthermore, as newer devices enable epoch length to be chosen and altered after data is collected, the use of shorter epochs is becoming more feasible, with many practical considerations, such as device storage, no longer relevant.

In summary, monitor calibration requires researchers to make many complex accelerometer use decisions. One of the most important aspects of accelerometer use is the translation of arbitrary count data into physical activity intensity. However, this review identified that the lack of population-specific cut points is limiting the ability of researchers to use valid data interpretation methods, therefore this is an important area for future research. In addition, it is also important to highlight the effect that cut points has on the estimation of physical activity intensity to make researchers aware of the importance of this use decision. As highlighted with epoch length, advancing technology is starting to help researchers by reducing the effect of some accelerometer use decisions. The importance of full reporting was also highlighted in this review, with the simplicity of the review criteria underestimating the attainment of some of the components within the theme of monitor calibration.

2.5.2.4 Analysis of accelerometer data

Prior to collecting field-based accelerometer data, it is important that researchers consider how the large amount of data collected will be cleaned, reduced, and analysed. These decisions will have important implications regarding the data which are included in the final analysis and therefore requires careful consideration.

As many study outcomes are reported in relation to activity per day, it is important to consider how many hours of wear per day is required to accurately represent daily activity behaviours and account for within-day variations (Trost et al., 2000). However, as highlighted in the introduction to this chapter, deciding upon the number of hours required to define a day is affected by the number of monitoring days, with longer monitoring periods requiring fewer hours of daily wear to achieve reliable measurements (Mattocks et al., 2008; Penpraze et al., 2006).

Within this review, of the studies which measured activity over multiple days, three specified 10 hours of wear per day, with two of these studies additionally requiring this on at least 4 out of 7 days, with one day being a weekend (Esposito et al., 2012; Phillips & Holland, 2011; Ulrich et al., 2011). Therefore, this suggests that this definition of a day, combined with the days of wear, will provide reliable data, based on previous estimates in typically developing children (Mattocks et al., 2008; Penpraze et al., 2006). For Horvat and Franklin (2001) and Lorenzi et al. (2000), which measured activity during specified periods of the day, it is difficult to conclude if these short periods of measurement provide reliable measures of physical activity. However, as school-based activity was measured on 2 and 3 separate occasions (Lorenzi et al., 2000 and Horvat & Franklin, 2001, respectively) and considering that longer monitoring periods are required, these shorter measurement days may limit the reliability of results.

Required wear time can depend on many population-specific factors, such as age, weekend/weekday variances, within- and between-day variances (Rich et al., 2013). Considering previous research in children with intellectual disabilities has identified within- and between-day differences in activity levels, a longer wear time, such as 10 hours, may be required (Foley et al., 2008; Horvat & Franklin, 2001). Furthermore, Tudor-Locke and Myers (2001) suggest that for more sedentary populations, a lower wear time is required. This could be relevant for children with intellectual disabilities as, in general, the studies within this review report children to be an inactive population group.

In addition to specifying the number of hours required for daily wear, Esposito et al. (2012) and Ulrich et al. (2011) note that to be included in the analysis participants had to return complete wear time data 4 out of 7 days, with Shields et al. (2009) specifying 6 out of 7 days. Three studies did not make this specification but instead discussed if participants returned data which did not meet the wear requirements, the data was excluded from the analysis (Foley et al., 2008; Whitt-Glover et al., 2006) or participants were asked to wear the accelerometer for another 7 days (Phillips & Holland, 2011).

Imputation is regarded as the most valid method of dealing with missing data, therefore the studies which excluded missing data will have introduced error into the results (Catellier, 2005). If days which do not meet the required wear time are included, this will likely underestimate the true activity level; in contrast, if these days are eliminated, the likelihood of overestimating activity is increased (Ward et al., 2005). As Foley et al. (2008) and Whitt-Glover et al. (2006) excluded data which did meet their criteria, the results of these studies could overestimate physical activity. Repeating the measurement to account for missing measurements, as employed by Phillips & Holland (2011), could prevent the effects of including or excluding data below the required wear time, however, it requires additional participant and researcher burden which needs to be considered.

Another use decision to increase the accuracy of results is to check for spurious data and decide upon methods to deal with this data. Foley and McCubbin (2009) and Phillips and Holland (2011) were the only studies to report checking for spurious data. However, only Phillips and Holland (2011) discussed how they dealt with irregular counts, which was by excluding outliers. The remaining eight studies within this review did not report checking for spurious data, which could result in the inclusion of data which was not the result of activity or which was due to device malfunction. Furthermore, as checking for and dealing with spurious data increases the validity of the results, the studies in this review which did not report checking for spurious data reduce the level of validity of the data used in the analysis (Cliff et al., 2009).

In the best practice recommendations, Ward et al. (2005) discuss the need to create standards of reporting relating to accelerometer use, specifically the methods and decisions used when analysing data. The decisions employed to analyse data have important implications on results and, therefore, it is important to fully report and justify these decisions (Freedson et al., 2012; Masse et al., 2005). Six studies in this review (60%) met the criteria for creating reporting standards by presenting information which clarified the data collected and the assumptions made, such as criteria for non-wear time, how data was checked for spurious data, and reporting the average number of hours and days the device was worn for. This information is important for understanding the specifics of how data was cleaned and reduced. Furthermore, reporting the average wear times could increase our knowledge of children with intellectual disabilities compliance with accelerometer use and can help develop reliable and population-specific wear time recommendations, which are important areas for future research.

Three studies reported average hours and daily wear, which ranged from 12.37 to 14.23 hours, suggesting that the previously discussed 10 hours per day is achievable for children with intellectual disabilities. Furthermore, Whitt-Glover et al. (2006) and Shields et al. (2009) report that for the 7-day monitoring period, children wore the device for an average of 6.3 and 6.8 days, respectively. This also suggests that children with intellectual disabilities are able to comply with the requirements of a 7-day monitoring period. These two studies also presented information relating to bouts of activity, which ranged from 2.8 to 10.7 minutes for moderate to vigorous intensity and 2.0 to 2.5 minutes for vigorous intensity activity (Shields et al., 2009; Whitt-Glover et al., 2006).

Presenting data in the form of bouts provides valuable information on how the physical activity data used in the analysis was accumulated over the measurement period. Furthermore, it adds to our knowledge of the physical activity patterns of children with intellectual disabilities which provides many advantages from a measurements perspective, such as better understanding the optimal choice of epoch to capture bouts of activity. Although Phillips and Holland (2011) did not report data relating to bouts, e.g. the average duration of bouts, only bouts which exceeded 10 minutes were included in the analysis, as this is the suggested duration required for health benefits. This information is important for children with intellectual disabilities as little is known about the dose-response relationship and the volume or intensity required to gain various physical and mental health benefits.

In summary, there were various use decisions and standards of reporting identified within this review. If researchers are to fully understand the data which results and conclusions are based on, it is important that all aspects of accelerometer use are fully reported. Furthermore, this section also highlights some of the inferences which can be made from well reported use decisions, such as the wear time and compliance with accelerometer use.

2.5.2.5 Integration with other data sources

Integration of other data sources involves measuring another dimension of physical activity, concurrent with accelerometry. This can add breadth and depth to the accelerometer data, but the added value of this additional information has to be considered in relation to the increased participant and researcher burden.

Only three studies within this review report using other data sources. Horvat and Franklin (2001) and Lorenzi et al. (2000) additionally used direct observation and heart rate. The rationale for these measurements was that heart rate provided additional information on intensity whereas direct observation gave added detail to the accelerometer data by describing the type of activity behaviours being conducted. Foley and McCubbin (2009) used a parental activity log to report children's sedentary behaviours. The rationale for the parental log was to gain a better understanding of the types of sedentary behaviours that children participated in. The use of additional measures was therefore appropriate to the research aims of these studies. However, as these studies were based in schools and had shorter monitoring periods, the integration of other data sources was more feasible using these designs. This is most notable within the studies conducted by Horvat and Franklin (2001) and Lorenzi et al. (2000), which used a secondary measure of direct observation, as this method requires measurement in a confined area, such as a school.

Integration with other data sources therefore needs to be considered in relation to the purpose of a study. For example, if the study outcomes rely on objective scores, e.g. changes in activity after an intervention, an accelerometer alone would provide this data. In contrast, if the outcomes of the study rely on contextual details, such as the environments where children are more active, e.g. physical education, recess, sports clubs, the use of multiple data sources could provide this information (Ward et al., 2005). Considering the limitations of using accelerometers during field-based research, such as not providing direct data on activity type or intensity, the use of another data source could provide additional relevant data. The integration of direct observation, heart-rate monitoring, and proxy activity logs reported within this review could enrich the accelerometer data. However, as discussed in Section 1.4, all these methods have limitations. Furthermore, as compliance is important, the use of another worn device, such as a heart rate monitor, may not be appropriate. Instead, additional data sources which are researcher- or parent-reliant may be more feasible.

In summary, the integration of other data sources can provide valuable information in addition to accelerometer data. However, the additional data may not outweigh the extra burden on participants, parents, or researchers, and therefore should be considered in relation to research aims. Although only three studies within this review included other methods of measurement, these were the only studies where other data sources were required to achieve the study aims and which were feasible to the study design. Therefore, overall, this review highlights that the use of additional data sources was appropriate, which is underestimated in the 30% attainment of the review criteria.

2.5.3 Impact on end users

Within physical activity research, a distinction can be made between the measurement researcher and the end user (Freedson et al., 2012). Measurement researchers focus on measurement science and the specifics of accelerometry, such as validation and calibration, whereas end users are researchers who are interested in the outcome measures of accelerometry in relation to surveillance, intervention, and epidemiological research. A novel aspect of this review was investigating how measurement research, i.e. research relating to the best practice guidelines, is translated into practice by end users.

Although this review focuses on accelerometer use specific to end users, many of the shortcomings noted are due to a lack of measurement research conducted in children with intellectual disabilities. This lack of research has an effect across all aspects of accelerometer use investigated within this review, but is most apparent in relation to monitor calibration. No studies met the review criteria for the use of group calibration equations, as accelerometer output has not been calibrated for children with intellectual disabilities. However, all studies - to varying extents - provided a rationale for the cut points used, such as referencing previous use in children with intellectual disabilities or validity established in typically developing children. Therefore, this highlights that end users are aware of the need for valid data interpretation methods and are fully reporting use decisions, but are hindered as no calibration research has been conducted in children with intellectual disabilities by measurement researchers. Therefore, this raises questions on the validity of data interpretation and the inferences made. Although reporting raw counts is one method to limit the impact on end users due to the lack of calibration research, this provides no information on the dose-response relationship, which further limits the ability of end users to investigate and increase our knowledge-base relating to the health benefits of physical activity in children with intellectual disabilities. To enable end users to improve the standard of accelerometer use, the interpretation of data, and the inferences made, measurement researchers need to conduct accelerometer calibration for children with intellectual disabilities.

Another possible impact of this lack of measurement-specific research is that end users may not be aware of the effect use decisions have on research quality, outcomes, and replication. Although it may not be feasible to meet all of these guidelines, end users should be aware of the implications of accelerometer use decisions and ensure the most appropriate methods are used, where possible. Furthermore, these use decisions need to be accurately reported in publications. For readers to fully understand the research and make critical decisions on quality and validity, researchers need to provide a clear rationale for why decisions were made. However, a lack of clear reporting is apparent in several components within this review, with researchers making appropriate use decisions but not fully reporting or justifying these decisions and, therefore, not meeting the review criteria.

It is accepted within measurement research that the use of accelerometers is in part a trade-off between the most valid and the most feasible methods of use. If physical activity research in children with intellectual disabilities is to advance, it is important that end users are aware of the limitations of devices and that reporting whether a decision was based on validity or feasibility is an important aspect of accelerometer use. On the other hand, from a measurement perspective, it is important that end users fully report accelerometer use and discuss the pros and cons of the use decisions, which will highlight areas which require further focus from measurement researchers (Freedson et al., 2012).

In summary, the lack of measurement-specific research for children with intellectual disabilities is limiting the ability of end users to produce results based on valid measures, and is resulting in methods of use which may have a negative impact on the reliability of findings described in the literature. Therefore, this highlights the need for measurement researchers to address these gaps in the research and to provide much needed validity evidence to support end users within the field of physical activity in children with intellectual disabilities.

2.5.4 Strengths and limitations

This systematic review is the first to focus on accelerometer use in children with intellectual disabilities. To ensure the quality of reporting and increasing the validity of conclusions, this review was conducted in accordance with the PRISMA statement (Moher et al., 2009). This review addresses gaps in our knowledge as to how accelerometers are used in research involving children with intellectual disabilities and, as a result, has allowed future areas of research to be identified. The novel approach of translating existing best practice guidelines into a review criteria ensured the inclusion of components of accelerometer use which are considered most important by experts in the field of physical activity measurement. The criteria developed within this review provides end users with simple recommendations on the many methodological decisions relating to accelerometer use and could guide researchers on the importance of decision making and reporting. Although more recent recommendations have been published which account for technological advances in accelerometry (Matthews et al., 2012), these updated guidelines lack the specificity of accelerometer use presented within Ward et al. (2005) and therefore were less relevant to the aims of this review.

Although the developed review criteria has many strengths, it is not without limitations. The review criteria has simplified complex use decisions into a dichotomous method of assessment, i.e. whether the study did or did not achieve the criteria for each component. Although this was necessary to make the review feasible in this format, it diluted the guidelines to an extent that it underestimated the number of studies which were conducting research at least partially in line with the guidelines. This was due to the distinction which can be made between how the accelerometer is used and how its use is being reported, specifically if a rationale was provided for the use decision. As discussed in the introduction to this chapter, many use decisions are based on study outcomes and feasibility with no consensus on methods of use; therefore, where no specific methods of use were presented in the guidelines, the review criteria were based on full reporting and justification for the use decision.

2.5.5 Conclusions

Accelerometer use decisions can have important implications on study results. However, a lack of measurement research specific to children with intellectual disabilities is preventing definitive recommendations being made regarding the most appropriate methods of accelerometer use. To limit the effect these decisions could have on the accuracy of results, the use of appropriate and valid methods is vital.

Many studies in this review, however, failed to report appropriate validity evidence as justification for decisions, e.g. in relation to monitor selection, placement, epoch, and group calibration equations employed. Of the studies which did report validity evidence, none of this was established in children and adolescents with intellectual disabilities. This subsequently raises questions regarding the accuracy of study findings, as validity evidence needs to be established in the situation and population of interest (Lincare, 2000; Yun & Ulrich, 2002). It is important that calibration and validation research is conducted in children with intellectual disabilities; however, to date, the validity of intensity cut points or equations in children with intellectual disabilities has not been investigated. As a result, this limits the ability of physical activity researchers to make informed decisions on accelerometer use.

Therefore, it is important that measurement researchers address these shortcomings, specifically in relation to the development of valid methods for the interpretation of accelerometer output. With the abundance of methodological questions facing the physical activity researcher and variance of accelerometer use reported, investigation into the effect of these use decisions - although problematic - is vital for informing future accelerometer use and monitoring protocols.

Chapter 3 – Thesis aims and research questions

3.1 Thesis aims

The first aim and research question of this thesis were addressed in Chapter 2. Based on the existing literature discussed in Chapter 1 and the findings of Chapter 2, the following four broad research aims were developed to be addressed within the following chapters of this thesis:

Aim 2: To examine the effect of accelerometer cut points on the estimation of physical activity intensity in children with intellectual disabilities.

Aim 3: To develop an effective and feasible accelerometer calibration protocol for children with intellectual disabilities.

Aim 4: To calibrate population-specific accelerometer cut points for the estimation of physical activity intensity in children with intellectual disabilities.

Aim 5: To cross-validate the developed accelerometer cut points in children with intellectual disabilities.

3.2 Research questions

To meet the four research aims described above, 18 research questions were developed. The research questions, and chapter in which each research question (RQ) is investigated, are described below:

<u>Chapter 4</u>

RQ 2: Does the use of different cut points result in significant differences in the estimated time spent sedentary in children with intellectual disabilities?

RQ 3: Does the use of different cut points result in significant differences in the estimated time spent in moderate intensity activity in children with intellectual disabilities?

RQ 4: Does the use of different cut points result in significant differences in the estimated time spent in vigorous intensity activity in children with intellectual disabilities?

RQ 5: Does the use of different cut points result in significant differences in the estimated time spent in moderate to vigorous intensity activity in children with intellectual disabilities?

Chapter 5

RQ 6: Is it feasible to recruit children with intellectual disabilities from additional support needs schools to participate in a calibration study?

RQ 7: Are activities conducted in a calibration protocol designed for typically developing children feasible for children with intellectual disabilities?

RQ 8: Is it feasible to measure resting energy expenditure (REE) in children with intellectual disabilities?

RQ 9: Is a treadmill-based graded exercise test to measure $\dot{V}O_2max$ feasible in children with intellectual disabilities?

RQ 10: Is the use of respiratory gas exchange equipment feasible in children with intellectual disabilities?

RQ 11: Does altering Ultima CPX breath-by-breath system threshold settings have an effect on $\dot{V}O_2$ in children with intellectual disabilities and typically developing children?

RQ 12: Is there a significant difference in the relationship between $\dot{V}O_2$ and accelerometer counts between children with intellectual disabilities and typically developing children?

Chapter 6

RQ 13: Does heart rate provide acceptable criterion validity for the measurement of total physical activity in children with intellectual disabilities?

RQ 14: Does the ActiGraph wGT3X+ provide acceptable criterion validity for the measurement of total physical activity in children with intellectual disabilities?

RQ 15: What are the optimal ActiGraph wGT3X+ vector magnitude cut points for the classification of sedentary, moderate, and vigorous intensity activity for children with intellectual disabilities?

RQ 16: What are the optimal ActiGraph wGT3X+ vertical axis cut points for the classification of sedentary, moderate, and vigorous intensity activity for children with intellectual disabilities?

Chapter 7

RQ 17: Do the developed vertical axis cut points provide a valid estimation of physical activity intensity in a sub-sample of children with intellectual disabilities?

RQ 18: Do the developed vector magnitude cut points provide a valid estimation of physical activity intensity in a sub-sample of children with intellectual disabilities?

RQ 19: Do the developed vertical axis cut points provide a more valid estimation of physical activity intensity in children with intellectual disabilities than existing cut points?

Chapter 4 – Application and effects of use decisions

4.1 Overview of this chapter

The systematic review reported in Chapter 2 details the many use decisions facing researchers when using accelerometers. Chapter 2 also discusses possible effects of these use decisions, with the interpretation of accelerometer output an area which could have clinically meaningful effects on study conclusions. In line with the guidelines presented in Chapter 2, the purpose of this chapter is to discuss and justify the accelerometer use decisions described within this thesis; specifically, the selection of a device and method for data interpretation. Furthermore, this chapter will empirically examine the effect that data interpretation methods, specifically cut points, has on the estimation of physical activity intensity in children with intellectual disabilities.

4.2 Introduction

To increase the comparability between studies and increase our understanding of the feasibility and validity of physical activity measurement in children with intellectual disabilities, it is important that accelerometer use decisions are fully described and justified. Therefore, the following sections will discuss and justify selecting a device and deciding upon the method used to interpret accelerometer output.

4.2.1 Use decision 1: selecting a device

There are numerous commercially available accelerometers, which can make choosing a device a difficult decision for physical activity researchers. With the growing interest in measuring activity behaviours, there has been a steep increase in the number of available devices, with upwards of 15 devices available, not accounting for different versions of the same device (Murphy, 2009; Reilly et al., 2008). Furthermore, as these devices can vary greatly in size, weight, number of axes measured, price, wear location, integration of other data sources, data processing/storage, and validity, it is important that researchers give careful consideration to which device is most appropriate to their study population and intended outcomes. In accordance with the guidelines described in Chapter 2, device selection should be based on several considerations, including population of interest, data processing, and storage capacity. Furthermore, it also important to consider the empirical evidence-base relating to device feasibility and validity for the population of interest.

4.2.1.1 Rationale for the ActiGraph

Of the available accelerometers, ActiGraph devices are most commonly used in research to measure physical activity in children and are regarded as a valid measure of physical activity (Cain, Sallis, Conway, Van Dyck, & Calhoon, 2013; McClain & Tudor-Locke, 2009; Reilly et al., 2008). Therefore, an advantage of using these devices in children with intellectual disabilities is that it will increase the scope of comparability with research involving typically developing children. The wide use of ActiGraph devices is partially due to their practicality and being user friendly, both in relation to the device and the associated software, with free and very accessible technical support available. From a technical perspective, ActiGraph devices are frequently updated in line with emerging technology and in response to specific measurement functions requested by researchers (Welk, McClain, & Ainsworth, 2012). Although this is a great advantage of the ActiGraph, a limitation with the more frequent developments is that the internal components and processing methods vary between different generations of device, which can limit comparability. That said, however, these technical advances put ActiGraph devices at the forefront of physical activity research and the development of new data handling and analysis techniques (Freedson et al., 2012; Welk et al., 2012). As a result, the use of the ActiGraph in children with intellectual disabilities will enable research in this area to progress in line with research involving typically developing populations, rather than continuing the research lag which is currently present in this field of research.

ActiGraph devices have also been used in research involving children with intellectual disabilities, with no feasibility issues identified (Hinckson & Curtis, 2013). Although, due to the small number of studies which have used accelerometry in children with intellectual disabilities, discerning between devices in relation to feasibility is difficult. Furthermore, due to the limited studies investigating the validity of accelerometers for use in children with intellectual disabilities, it is not possible to make accurate conclusions on the most valid device either. Therefore, the decision to use the ActiGraph device is based on feasibility and the device-specific factors discussed.

4.2.1.2 Evolution of the ActiGraph

With the frequent updates seen for ActiGraph devices, researchers need to additionally decide which generation of device to use and consider the effect this could have on the data collected and the scope for comparison with previous research. Therefore, the following sections will discuss the evolution of ActiGraph accelerometers and the comparability between different generations of device.

4.2.1.2.1 AM7164

The AM7164, or CSA/MTI, is the original ActiGraph device which was released in 1999. Although the development of the AM7164 has long been discontinued, its wide use in calibration and validation studies makes this device an integral part of physical activity measurement research. Relatively small (51 × 41 × 15 mm) and lightweight (43 g), the AM7164 is a uniaxial device which measures acceleration on the vertical axis. This device has the capability to store up to 64 KB of data (John, Tyo, & Bassett, 2010). The acceleration sensor within the AM7164 is the traditional bimorph piezoelectric cantilever beam and seismic mass. With acceleration in the dynamic range of 0.05–2.13 g, the seismic mass forces the sensor to bend in the direction of the acceleration and produce a proportional electric charge. This charge is filtered using a hardware-based band-pass filter (0.21–2.28 Hz) and digitalized using an 8-bit analogue-to-digital convertor at a sampling rate of 10 Hz. The signal is subsequently converted to an absolute acceleration value (full-wave rectification) and converted to activity counts using a proprietary algorithm for the predetermined epoch (Ried-Larsen et al., 2012).

4.2.1.2.2 GT1M

The GT1M replaced the AM7164 in 2005. Prior to the release of the GT1M, ActiGraph released an intermediate update of the AM7164, named the AM71256 (John et al., 2010). However, this device was rarely used in research due to the subsequent release of the GT1M, therefore in-depth discussion of this device is not deemed necessary.

The GT1M was the first ActiGraph device to use a capacitive accelerometer, instead of the formally used piezoelectric sensor. This small, lightweight device (3.8 × 3.7 × 1.8 cm, 27 g) contains an ADXL320 acceleration sensor (Analog Devices, MA), specifically a micromachined, monolithic circuit chip, dual-axis Microelectro-Mechanical-System (MEMS) accelerometer (John & Freedson, 2012). This accelerometer has the capabilities to measure static and dynamic accelerations in a range of ± 5 g; however, ActiGraph restricts this to 0.05 to 2g. Furthermore, the GT1M has a substantially greater memory capacity (1 MB) than the preceding AM7164 device (John et al., 2010).

The internal mechanism of this device contains two fixed plates which act as electrodes, between which is a moveable third plate, resulting in two back-to-back capacitors. Together, the three plates form a differential capacitor. Acceleration causes variances in the capacitance, which results in a change in voltage of the analogue signal, which is proportionate to the acceleration. This signal is amplified and converted into a digital output for the vertical and mediolateral axes using a 12-bit analogue-to-digital convertor, at a sampling rate of 30 Hz. This signal is filtered at a bandwidth of 0.25 to 2.50 Hz to eliminate frequencies which are not deemed a result of human movement, and finally converted into the output of activity counts (John & Freedson, 2012). Unlike the AM7164, however, the GT1M allows data to be viewed in the pre-filtered raw acceleration format. Data is outputted in the form of gravitational force (g), using the following formula (John & Freedson, 2012):

Raw g-force = 2.022V (voltage signal from accelerometer) – 1.5V (zero-g offset) ÷ 174 millvotls/g (sensitivity of accelerometer)

An advantage of output in this pre-filtered format is that it has the depth of detail to allow researchers to further understand physical activity behaviours, as acceleration is measured 30 times per second (30 Hz). As a result, data in this format has the potential to be used for new methods of data interpretation, such as pattern recognition. Furthermore, it theoretically increases the scope of comparability between devices, as differences in filtering and processing of this raw data into counts affects the equivalency of output between devices (Welk et al., 2012).

4.2.1.2.3 GT3X

The GT3X was released in 2009 and is the first of the current “third generation” of ActiGraph devices. Although very similar to the GT1M in terms of specifications - including, size, weight, data filtering, and digital conversion - the primary difference is the updated ADXL335 internal accelerometer (Analog Devices, MA). This includes a triaxial capacitive MEMS sensor which enables the measurement of acceleration in the range of ± 3 g across all three planes (John & Freedson, 2012). The inclusion of a triaxial accelerometer has important implications for physical activity measurement, primarily as this enables output data in the form of vector magnitude. Vector magnitude utilises raw measurements from all axes and is derived using the following formula (ActiGraph, 2012a):

$$\text{Vector magnitude} = \sqrt{[(\text{axis } 1)^2 + (\text{axis } 2)^2 + (\text{axis } 3)^2]}$$

Another important addition to the GT3X is the inclusion of the low frequency extension filter. The MEMS accelerometer within this device is very sensitive and has the capabilities to detect even slight movements. Therefore, to exclude accelerations which are deemed not to be representative of human activity, the acceleration signal must cross a threshold to be recorded (ActiGraph, 2012b). However, a limitation of this is that physical activity data recorded by populations with a slow walking speed or low acceleration output may not reach the required threshold and be excluded. Therefore, the low frequency extension allows researchers to reduce the lower filter threshold and expand the bandwidth of data recorded. As a result, the likelihood of low intensity activity being excluded is reduced.

4.2.1.2.4 GT3X+

In 2010, the most recent third generation version of the ActiGraph was released – the GT3X+ (46 × 33 × 15 mm, 19 g). Although this device is very similar to the GT3X in relation to its internal components, it has various functional improvements which make it more user-friendly. This device is available in a wireless option (wGT3X+) which allows wireless interface with other devices enabled with ANT+ technology. This is an advantage for both researchers and participants as it can make the use of multiple data sources easier and less of a burden for participants; for example, with ANT+ enabled heart rate monitors, the wGT3X+ can wirelessly record heart rate data without the need for a separate heart rate device receiver to be worn. The GT3X+ also contains a much greater storage capacity (256 MB), in comparison to the GT3X (16 MB), which allows longer duration and increased depth of data to be recorded. Furthermore, this device is water resistant, which allows the device to be worn during water-based activities.

In terms of physical activity measurement, the most substantial changes to this device are in relation to data sampling and reduction. Unlike previous versions of the ActiGraph, the GT3X+ records data in raw acceleration only. Using ActiLife software, researchers have the flexibility to choose the sampling frequency with which data are recorded, which ranges from 30 Hz to 100 Hz, and edit data filtering and reduction techniques, such as choosing epoch length - which ranges from 1 second to 1 minute - after data have been collected. This allows measurement researchers to gain in-depth data (up to 100 measurements per second) which can aid in our understanding of physical activity and accelerometer measurement. However, as the GT3X+ requires additional decisions from end users, in relation to processing and reducing raw data, it also reinforces the need for clear guidelines on how to handle accelerometer data and the effects of use decisions.

4.2.1.3 Comparability between devices

From a technical perspective, as all versions of the ActiGraph contain an accelerometer, data on acceleration should be interchangeable between devices (Welk et al., 2012). However, as each version filters and processes the raw acceleration signal differently, this could lead to non-comparable output. As comparison between studies and data consolidation are integral aspects of research, it is important to understand the extent of equivalency between these devices. Table 4.1 summarises the internal specifications of each ActiGraph device, as discussed in Section 4.2.1.2.

Previous studies have used mechanical oscillators to test the inter-generational differences between ActiGraph devices. Devices are subjected to a known frequency of oscillations, with frequency referring to the number of oscillations per second (Hz). An advantage of this methodology is that multiple devices can be monitored simultaneously at various frequencies without the inter- and intra-individual error associated with human trials (Rothney et al., 2008).

Rothney et al. (2008) report that the AM7164 and GT1M have a similar response to increasing acceleration, in that the relationship between acceleration and counts is linear at lower frequencies but nonlinear at higher frequencies. However, the actual output between these devices is significantly different at all frequencies, except 120 Hz, with the AM7164 recording higher count values. These findings are supported by Ried-Larsen et al. (2012) who report the AM7164 to record significantly lower counts at a lower frequency (0.8–2.0 Hz; -2.5–9.0 counts/sec, $p < 0.001$), and significantly higher counts at higher frequencies (3.3 counts/sec at 0.7 Hz, $p < .017$; 5.0–14.0 counts/sec at 2.5–3.0 Hz, $p < .017$), in comparison with the GT1M. In addition, Ried-Larsen et al. (2012) also examined the differences between the GT3X and GT3X+. The authors report no significant differences between the third generation devices. However, in comparison with the GT1M, at higher frequencies ranging from 1.8–3.0 Hz, the GT3X and GT3X+ record significantly lower output (-1.0–3.0 counts/sec, $p < .017$).

In a laboratory setting involving free-living and treadmill-based activities, Robusto and Trost (2012) examined differences between the GT1M, GT3X, and GT3X+ in a sample of 29 children aged 7 to 18 years. Almost perfect agreement was found between the devices for the vertical axis ($r = .994$), vector magnitude ($r = .981$), and moderate to vigorous intensity activity ($r = .996$).

Table 4.1. Summary of the internal specifications and processing differences between ActiGraph devices

Device	Axes	Band pass (Hz)	Dynamic range (g)	Sampling frequency (Hz)	Accelerometer	Storage
AM7164	Uniaxial	0.21-2.28	0.05-2.13	10	Piezoelectric	64 KB
GT1M	Dualaxial	0.25-2.50	0.05-2.50	30	MEMS	1 MB
GT3X	Triaxial	0.25-2.50	± 3	30	MEMS	16 MB
GT3X+	Triaxial	0.25-2.50	± 6	30-100	MEMS	256 MB

For the vertical axis, the GT1M records 1.5% higher counts than the GT3X, with the GT3X+ recording 1.3% higher count values than the GT3X. A similar trend was found for vector magnitude, with the GT3X recording 1.7% lower counts than the GT3X+. In a more recent study, Grydeland, Hansen, Ried-Larsen, Kolle, and Anderssen (2014) compared the output from the AM7164, GT1M, and GT3X+ in a sample of 16 children aged 9 years during free-living activity. Similar to the results by Robusto and Trost (2012), Grydeland et al. (2014) report almost perfect agreement between devices for mean vertical axis counts ($r = .985$). In relation to time spent in physical activity intensities, the AM7164 records less sedentary time than the GT1M and GT3X+, but more time in vigorous intensity activity. This in part contradicts the mechanical assessment findings by Ried-Larsen et al. (2012) who found the AM7164 to be more sensitive to lower frequency movement. However, this supports previous findings that the AM7164 records higher counts during more vigorous movement (Ried-Larsen et al., 2012; Rothney et al., 2008).

Although the findings of these studies are somewhat contradictory, there is a general consensus that output between the GT1M, GT3X, and GT3X+ is comparable. However, output from the AM7164 is generally significantly different from the later devices. In terms of physical activity measurement, the AM7164 produces higher count values for activity in comparison with newer models. Considering that the AM7164 contains a piezoelectric sensor, these findings suggest that the MEMS accelerometer produces a more reliable output. However, from the data reported, it is difficult to fully understand the causes for variances between the GT1M, GT3X, and GT3X+, although this could be attributed to sensitivity and filtering differences between these devices. Furthermore, this research area is limited by the lack of child-specific studies investigating comparability of devices, particularly in comparison to the greater number studies which include an adult sample (Cain et al., 2013). Therefore, these conclusions need to be interpreted with caution.

One possible approach to increasing device comparability is using raw data rather than counts, which could limit the effect of internal data processing on device output. That said, however, the GT1M, GT3X and GT3X+ measure different ranges of acceleration (3, 5, and 6 g, respectively) which could produce different output. Furthermore, as described in Section 4.2.1.2, output in the form of raw acceleration (g) is a result of processing the acceleration signal. Therefore, as sensitivity between the internal ADXL320 and ADXL335 accelerometers is different (174 milivolts/g and 270 milivolts/g, respectively), this could also result in raw output which is not comparable between the GT1M and GT3X/GT3X+ (John & Freedson, 2012). Although, to date, these possible differences in raw output have not been investigated. Another method to increase the comparability between accelerometers is to calibrate output for the estimation of physiological variables. If output is calibrated for each device and population in which it is used, this will not only allow comparison between versions of the same device, but between different devices.

4.2.2 Use decision 2: interpreting accelerometer output

Accelerometer output is generally calibrated to measure energy expenditure or activity intensity. As the methods employed to interpret accelerometer output are device-specific, it is important to make this decision in conjunction with device selection. Also, as data interpretation methods are population-specific, theoretically, this decision should additionally be based on the available population-specific validity evidence. However, as no population-specific methods have been developed for children with intellectual disabilities, the following sections will broadly discuss the advantages and disadvantages of using energy expenditure and activity intensity data interpretation methods for the ActiGraph device in children.

4.2.2.1 Estimating energy expenditure

For researchers who want to measure physical activity in relation to energy expenditure, five ActiGraph regression equations have been developed to estimate this parameter of physical activity in children (Trost, 2007a). These regression equations allow energy expenditure to be estimated in a free-living setting on a minute-by-minute basis over multiple monitoring days. These equations have been developed to estimate various parameters of energy expenditure, such as METs, activity energy expenditure, and total energy expenditure. To account for the complexities of estimating energy expenditure in children, these equations - to varying degrees - include population-specific independent variables known to influence energy expenditure. For example, the equations developed by Freedson et al. (2005) and Mattocks et al. (2007) include the independent variables of age, and age and sex, respectively; in comparison, however, Puyau et al. (2002) and Treuth et al. (2007) include no participant-specific variables.

There are many limitations with the use of energy expenditure equations, with low predictive validity found for typically developing children. Equations generally overestimate sedentary behaviours, underestimate moderate and vigorous intensity activity, and cannot accurately estimate energy expenditure for multiple types of activity (Corder et al., 2007; Trost, Way, & Okely., 2006; Warolin et al., 2012). However, these finding are not consistent across all studies, with Trost (2007a) reporting that the equations developed by Freedson et al. (2005) and Puyau et al. (2002) underestimate sedentary and light intensity activity and overestimate over-ground walking, against a criterion measure of indirect calorimetry.

A possible cause of this limited validity is the nonlinear relationship between counts and energy expenditure, which introduces activity- and intensity-related measurement errors (Rothney et al., 2008). Furthermore, the methods employed in these original calibration studies vary considerably in relation to study sample, activity protocol, and criterion measure. For example, sample size ranges from 26 to 163 participants, includes large (6 to 18 years) and narrow (12.4 ± 0.2 years) age ranges, single and mixed sex samples, free-living and treadmill-based activities, and direct and indirect criterion measures of energy expenditure. This has important implications for the generalisation of equations, therefore, energy expenditure equations have limited validity in populations with different characteristics to original calibration sample (Warolin et al., 2012).

As a result, the use of regression equations in children requires great caution. Furthermore, Frey et al. (2008) also recommend that regression equations developed in typically developing children are not used in children with intellectual disabilities due to movement and metabolic variability between these populations. However, when MET thresholds are applied to energy expenditure output, these data can be used to develop accurate cut points to discriminate between activity intensities (Trost et al., 2006). More recently, instead of regression equations, studies have employed receiver operating characteristic (ROC) curve analysis to estimate activity intensity to limit the effects associated with regression equations.

4.2.2.2 Intensity cut points

An alternative method to derive physiological meaning from counts is the use of intensity cut points. Although cut points are affected by similar measurement error as regression equations, for example as a result of different BMI or age, the magnitude of this effect is less in comparison with regression equations (Rothney et al., 2008). This is primarily because intensity is categorised within wider count boundaries. Therefore, only counts which are close to the cut point thresholds are expected to be misclassified (Rothney et al., 2008).

Understanding physical activity intensity is important, particularly in relation to the attainment of physical activity guidelines and understanding dose-response relationships. As a result, cut points are the most commonly used method to interpret accelerometer output in children with intellectual disabilities (Hinckson & Curtis, 2013). Although the magnitude of error associated with cut points is smaller than that of energy expenditure equations, cut points are still calibrated based on the relationship between accelerometer counts and a physiological/behavioural criterion measure, with many cut points derived from energy expenditure regression equations.

Multiple sets of cut points have been developed for children, each with different count boundaries for the classification of activity intensities. The methodologies employed within these studies vary considerably in relation to protocol, criterion measure, accelerometer device, and study sample, which will affect the cut points calibrated. With no consensus on which cut points to use, researchers wanting to measure physical activity intensity are left with what is known as the "cut point conundrum" (Trost, 2007a; Trost et al., 2006). This has resulted in many issues with the use of cut points, such as the misuse of adult cut points in child samples, and the comparison of results between studies which use notably different cut points (Guinhouya et al., 2006; Nilsson et al., 2002). Therefore, it is important for researchers to be aware that intensity cut points are also very population- and accelerometer device-specific.

In typically developing children, the choice of ActiGraph cut points can have significant and clinically meaningful differences in physical activity intensity estimations. In a free-living setting, Anderson, Hagstromer, and Yngve (2005) measured physical activity for four days using the AM7164 device and report significantly higher moderate to vigorous estimates using the cut points developed by Freedson et al. (2005) in comparison to those developed by Puyau et al. (2002; 65.20 ± 43.20 minutes and 17.50 ± 18.50 minutes, respectively). These findings are concurrent with a study by Guinhouya et al. (2006), which found the Freedson cut points report significantly higher daily moderate to vigorous intensity activity (114 ± 39 minutes) in comparison to the Puyau cut points (28 ± 18 minutes), which illustrates a mean error bias of 113 minutes per day. Considering physical activity guidelines recommend 60 minutes of activity per day, the choice of cut point results in clinically meaningful differences, with the percentage of children in the study by Guinhouya et al. (2006) meeting the guidelines ranging from 8.7% to 100%, depending on cut points used. Similarly, Reilly et al. (2008) conducted a secondary data analysis on 7-day accelerometer data to investigate the effect of cut points on daily time spent in moderate to vigorous intensity activity. This study reports that the cut points used results in significant and clinically meaningful effects, with Puyau et al. (2002), Treuth et al. (2004), and Freedson et al. (2005) estimating 28 minutes (95% CI 27, 33 minutes), 41 minutes (95% CI 33, 48 minutes), and 266 minutes (95% CI 254, 281 minutes) of moderate to vigorous intensity activity per day, respectively.

In addition to comparing the effect of cut points, more recent studies have additionally included a criterion measure to empirically test the criterion validity of cut points. A laboratory-based study by Trost, Loprinzi, Moore, and Pfeiffer (2011) reports that the combined level of agreement for the estimation of total activity, against a criterion measure of indirect calorimetry, is substantial for the cut points developed by Evenson, Catellier, Gill, Ondrak, and McMurray (2008; κ = .68), Freedson et al. (2005; κ = .66), and Treuth et al. (2004; κ = .62), and moderate to fair for Mattocks et al. (2007; κ = .54) and Puyau et al (2002; κ = .36). The authors conclude that the Evenson et al. (2008) and Freedson et al. (2005) cut points provide valid intensity estimations for field-based research, but recommend the use of the Evenson et al. (2008) cut points due to the higher accuracy shown for all intensities. On the other hand, the authors also suggest use of the Treuth et al. (2004), Puyau et al. (2002), and Mattocks et al. (2007) cut points is discontinued. These findings are concurrent with McClain et al. (2008), who used direct observation as a criterion measure, and found moderate to vigorous activity estimates based on the Freedson et al. (2005) cut point were not significantly different from the criterion, although the cut points developed by Mattocks et al. (2007) and Treuth et al. (2004) significantly underestimated activity by as much as 39–74%.

There is limited research investigating the validity of cut points in children with intellectual disabilities. However, a small number of studies have included a sample of children with cerebral palsy, which is associated with intellectual disabilities in some children. In a laboratory-based study, Clanchy, Tweedy, Boyd, and Trost (2011) tested the validity of cut points developed by Freedson et al. (2005), Evenson et al. (2008), Puyau et al. (2002) and Trueth et al. (2004) in 29 children. For sedentary behaviour, the Freedson, Evenson, Puyau, and Treuth cut points had excellent classification accuracy (area under the curve; AUC = 90.00 to 91.60) against a criterion measure of indirect calorimetry. However, for moderate to vigorous intensity activity, only the cut point developed by Evenson exhibited excellent classification accuracy (AUC = 90.90), with the other cut points showing only fair classification accuracy. An interesting aspect of this study was the calibration of a population-specific cut point for moderate to vigorous activity. The developed cut point of ≥ 2012 cpm produced the highest sensitivity for detecting this intensity of activity (91.40%, AUC = 94.00), and was 284 cpm lower that the Evenson et al. (2008) cut point. A possible cause of this lower cut point boundary is that the energy costs of walking in children with cerebral palsy is higher than that of typically developing children, which is primarily associated with gait abnormalities (Johnston, Moore, Quinn, & Smith, 2004; Thomas, Buckon, Russman, Sussman, & Aiona, 2011).

A more recent calibration study conducted by Oftedal, Bell, Davies, Ware, and Boyd (2014) aimed to calibrate and test the predictive validity of uniaxial and triaxial cut points using the GT1M, GT3X and GT3X+ devices for sedentary activity in toddlers with cerebral palsy. Although the cut points developed in toddlers are not comparable with children due to the different modes of ambulation seen in toddlers, such as crawling, rolling, and shuffling, the methodology and results of this study are still noteworthy. This study calibrated sedentary cut points of 24 cpm and 240 cpm for the vertical axis and vector magnitude, respectively. When cross-validated, the vector magnitude cut points recorded minimal and non-significant bias, however, the vertical axis cut points were significantly different from estimated sedentary time, based on the criterion measure of direct observation. Therefore, this illustrates that the newer triaxial ActiGraph devices have the potential to limit the bias associated with cut points for estimating sedentary time. However, these cut points were calibrated at the rarely used back placement which, theoretically, prevents their use in studies which utilise the more commonly used hip placement.

In summary, intensity cut points reduce the bias associated with energy expenditure equations, with some cut points showing high criterion validity in typically developing children. However, as with energy expenditure equations, intensity cut points are population-specific, yet no cut points have been calibrated specifically for children with intellectual disabilities. As a result, researchers wanting to measure physical activity in children with intellectual disabilities are reliant on generalising cut points, thus raising questions on validity. With little evidence on the most valid cut points in children with intellectual disabilities, there is little empirical evidence to help researchers choose the most appropriate and valid cut points. Therefore, the choice of cut points will affect the estimation of physical activity intensity in children with intellectual disabilities.

4.2.3 Summary

ActiGraph devices provide a feasible method of measuring physical activity in children with intellectual disabilities. From a measurement perspective, more recent devices give researchers a greater level of control of how data is collected and processed. As a result, using ActiGraph devices in children with intellectual disabilities provides a feasible measure, with the potential to develop data interpretation techniques in line with emerging technology and the research developments seen in typically developing populations. However, as advanced data interpretation techniques, such as pattern recognition, are still in the early stages, researchers are currently reliant on the traditional techniques of energy expenditure equations and intensity cut points to interpret data.

Cut points are the most commonly used method to interpret accelerometer data and provide valuable intensity-related outcomes, which are relevant to the attainment of physical activity guidelines and can increase our understanding of the dose-response relationship. Furthermore, intensity cut points are prone to less bias and exhibit a higher degree of validity than energy expenditure equations. However, with no population-specific cut points for children with intellectual disabilities, researchers have to use cut points validated in typically developing children, with little information on the most appropriate cut points. As a result, the cut points used will result in differences in the estimated time spent in various activity intensities, thus affecting the validity and comparability of results.

Previous studies have investigated the effect of cut point used on the estimation of physical activity intensity in typically developing children, with significant and clinically meaningful differences reported. However, since these studies were conducted, new cut points have been developed, updated ActiGraph devices released, and different statistical techniques used, which have not been investigated. Furthermore, these effects have not been investigated in children with intellectual disabilities.

4.2.4 Research questions

The purpose of this study is to 1) update the existing literature on the effect of cut points on the estimation of physical activity intensity, and 2) provide an empirical rationale for the calibration of accelerometer cut points in children with intellectual disabilities. Specifically, the research questions to be examined in this study are:

RQ 2: Does the use of different cut points result in significant differences in the estimated time spent sedentary in children with intellectual disabilities?

RQ 3: Does the use of different cut points result in significant differences in the estimated time spent in moderate intensity activity in children with intellectual disabilities?

RQ 4: Does the use of different cut points result in significant differences in the estimated time spent in vigorous intensity activity in children with intellectual disabilities?

RQ 5: Does the use of different cut points result in significant differences in the estimated time spent in moderate to vigorous intensity activity in children with intellectual disabilities?

4.3 Method

4.3.1 Design

A secondary data analysis design was used within this study. This design was chosen as a data set was identified which contained the relevant data required to examine the research questions within this study.

4.3.2 Data: Get Active, Be Healthy study

Data from the Get Active, Be Healthy study was used for this secondary data analysis. Get Active, Be Healthy was a multi-component intervention conducted in 2010 in Glasgow, Scotland, by researchers at the University of Glasgow. Two of the researchers involved in the research reported within this thesis (CM & VP) were investigators in the Get Active, Be Healthy study, with CM being the principal investigator. Therefore, this data set was easily accessible.

Two additional support needs schools participated in this study, with a combined participation rate of 59 children. The 10-week intervention consisted of three components (physical activity sessions, classroom material, and home-based material), with 7-day accelerometer measures conducted pre- and post-intervention. The physical activity component of this intervention consisted of twice-weekly activity sessions, which were developed by the original research team. The sessions lasted approximately 60-70 minutes and consisted of a warm up, non-skill based exercises and games, and a cool-down, and aimed to get participants active at a moderate to vigorous intensity. All sessions were conducted by a trained researcher. One of the aims of the Get Active, Be Healthy study was to investigate how physically active children were during the physical activity session component of the intervention. To investigate this, ten children were randomly selected to wear an accelerometer during one of the sessions; the accelerometer data from this activity session was used for the analyses reported within this chapter.

The accelerometer data from the activity session has many advantages which make it suitable for the aims of the present study. Firstly, activity was measured using an ActiGraph accelerometer and included a sample of children with intellectual disabilities. Secondly, due to the structured session, it was assumed that the data would include a range of behaviours and intensities of activity. This was important to the aims of the present study as a thorough investigation into the effect of cut point use decisions requires sufficient data within each intensity category. Thirdly, as activity was completed in a controlled environment, there were no missing data points. As a result, there was no need to impute or exclude data, thus making the data representative of children with intellectual disabilities physical activity behaviours. Fourthly, the data set was deemed to be of sufficient volume required for the analyses and also one that was feasible to use, in terms of research burden. Therefore, the data from the physical activity sessions within this intervention was deemed more relevant to the aims of the present study, in comparison with using the 7-day free-living physical activity measurement data.

4.3.3 Measurement of physical activity

Physical activity was objectively measured using ActiGraph GT1M accelerometers (ActiGraph, Pensacola, FL). Participants wore one device on the right hip at the iliac crest for the duration of the activity session.

These small, lightweight devices (3.8 × 3.7 × 1.8 cm, 27 g) convert measured acceleration signals into digitized output signals at a rate of 30 Hz, which are filtered at a frequency rate of 0.25–2.5 Hz. The GT1M measures accelerations ranging in magnitude from 0.05–2.0 g for the vertical and mediolateral axis (Robusto & Trost, 2012). These devices have shown high technical reliability for the measurement of movement (Rothney et al., 2008; Santos-Lozano et al., 2012). A full description of the GT1M is presented in Section 4.2.1.2.2.

4.3.4 Identification of relevant cut points

A systematic search and inclusion methodology were used to identify cut points calibrated for ActiGraph accelerometers in children. As multiple cut points have been developed, with various terms used to describe cut points, the use of a systematic search strategy helped identify all cut points which were relevant for this study. The search strategy presented in Figure 4.1 was used to identify relevant studies within Ovid MEDLINE and Embase databases. This search was conducted in March 2015 and included studies from 1990 to March 2015. The search was limited to post-1990 as the development of cut points began in this decade (Troiano, 2005). Furthermore, the ActiGraph website, which includes a knowledge-base of research relating the development of ActiGraph-specific cut points, was also hand searched (ActiGraph, 2012a).

The widely accepted standard age ranges for calibration are: infant = < 1 year, toddler = 1 to 2 years, preschool = 3 to 5 years, children = 6 to 18 years, and adult = ≥ 19 years (ActiGraph, 2012a). Therefore, to be included in the analyses, cut points had to be calibrated in children within the age range of 6 to 18 years. Furthermore, with the technological advances of the third generation ActiGraph accelerometers, more recent cut points have utilised triaxial measurements and calibrated cut points for vector magnitude. However, as the device used within the Get Active, Be Healthy study was an older generation GT1M, which was worn at the hip placement, only cut points which were calibrated for the vertical axis and hip placement were included. The final inclusion criterion was cut points which classify activity intensity, with regression equations to classify energy expenditure excluded. In summary, the inclusion criteria were calibration studies which:

- included a sample aged 6 to 18 years
- calibrated cut points for an ActiGraph device
- calibrated cut points for the vertical axis
- calibrated cut points for the hip placement

- calibrated intensity-related cut points.

The search strategy identified an initial 222 studies. After studies which did not meet the inclusion criteria were removed, nine calibration studies were identified which were subsequently included in the analysis. Table 4.2 gives a brief overview of the cut points developed from these studies, the calibration methodology used, and the study sample. All studies included were original calibration studies, except for Freedson et al. (2005).

1. Acceleromet*.tw
2. ActiGraph*.tw.
3. CSA.tw.
4. MTI.tw.
5. GT1M.tw.
6. GT3X.tw.
7. Cut points.tw.
8. Cut offs.tw.
9. Accelerometer thresholds.tw.
10. Calibration.tw.
11. Validation.tw.
12. Physical activity.tw.
13. Children.tw.
14. Adolescents.tw.
15. 1 or 2 or 3 or 4 or 5 or 6
16. 7 or 8 or 9
17. 10 or 11
18. 13 or 14
19. 12 and 15 and 16 and 17 and 18
20. Limit 19 to “all children”
21. Limit 20 to “full text”

Figure 4.1. Embase search strategy used to identify existing ActiGraph cut points calibrated in children

Table 4.2. Summary of existing ActiGraph cut points and the specific calibration methods of each study

Study	Puyau (2002)	Treuth (2004)	Freedson (2005)	Mattocks (2007)	Evenson (2008)	Pulsford (2011)	Vanhelst (2011)	Mackintosh (2012)	Jimmy (2013)
Cut points (cpm)									
Sedentary	0-799	0-100	0-500	0-100	0-100	0-99	0-400	0-372	n/a
Light	800-3199	101-2999	n/a	101-3580	101-2295	100-2240	401-1900	373-2160	n/a
Moderate	3200-8199	3000-5200	501-4000	3581-6129	2296-4011	2241-3840	1901-3918	2161-4806	1596-2315
Vigorous	≥ 8200	≥ 5201	4001-7600	≥ 6130	≥ 4012	≥ 3841	≥ 3919	≥ 4807	≥ 2316
Very Vigorous	n/a	n/a	≥ 7601	n/a	n/a	n/a	n/a	n/a	n/a
Participants									
Total	26	74	50	246	33	53	40	28	21
Male	14	0	27	110	12	29	20	13	Not specified
Female	12	74	23	136	21	24	20	15	Not specified
Age (years)	6 to 16	13 to 14	6 to 17	12.4 (0.2)	5 to 9	7 to 8	13.2 (0.9)	10 to 11	5 to 9

Table 4.2. Continued

Study	Puyau (2002)	Treuth (2004)	Freedson (2005)	Mattocks (2007)	Evenson (2008)	Pulsford (2011)	Vanhelst (2011)	Mackintosh (2012)	Jimmy (2013)
Calibration protocol									
Device	AM7164	AM7164	AM7164	AM7164	AM7164	GT1M	GT1M	GT1M	GT3X
Criterion	Whole room calorimetry	Indirect calorimetry	Indirect calorimetry	Indirect calorimetry	Indirect calorimetry	Indirect calorimetry	Indirect calorimetry	Direct observation	Indirect calorimetry
Protocol type	Treadmill & free-living	Free-living	Treadmill	Free-living	Treadmill & free-living	Free-living	Treadmill & free-living	Free-living	Free-living
Analysis	Regression equation[1]	Regression equation[2]	Regression equation[3]	Regression equation[4]	ROC curves	LDA/ROC curves*	ROC curves	ROC curves	ROC curves

* Cut points developed using linear discriminant analysis and validated using ROC curves
[1]Activity energy expenditure (kcal/kg/min) = 0.0183 + 0.000010 cpm
[2]METs = 2.01 + 0.00171 × counts per 30-seconds
[3]METs = 2.757 + (0.0015 × cpm) − [0.08957 × age (yr)] – [0.000038 × cpm × age (yr)]
[4]Energy expenditure (kcal/kg/min) = − 0.933 + [0.000098 × cpm) + [0.091 × age (yr)] – [0.04 × sex (M=0, F1)]

Freedson et al. (2005) developed intensity cut points based on a previously developed regression equation by Trost, Ward, Moorehead, Watson, Riner, and Burke (1998) with MET thresholds of 3, 6, and 9 METs used to identify cut points for sedentary/light, moderate, and vigorous intensity, respectively. However, a discrepancy was identified between the Freedson et al. (2005) cut points reported in the original article and the cut points reported on the ActiGraph website. Freedson had originally developed three cut points (sedentary/light, moderate, and vigorous) based on the aforementioned MET thresholds; however, ActiGraph (2012a) report the Freedson cut points with the inclusion of an additional threshold to discriminate between sedentary and light intensity activity, which was not a result of the initial Freedson et al. (2005) study. ActiGraph were contacted in June 2015 for clarification on how the sedentary cut point was established. They confirmed that the cut point of 149 cpm to discriminate between sedentary and light activity was not established by Freedson, but by persons at ActiGraph, although they had no records of how this cut point was established, e.g. the MET threshold applied or the sample used. As there was no clarification on how the sedentary cut point was established, only the cut points from the original Freedson et al. (2005) article were included in this study.

Jimmy, Seiler, and Mäder (2013) calibrated two vigorous intensity cut points, based on a 5 and 6 MET threshold, and investigated the validity of the derived cut points. The cross-validation shows the vigorous cut point established using the 5 MET threshold to be more valid (sensitivity = 74%, specificity = 79%, κ = .50) than the cut point established using the 6 MET threshold (sensitivity = 53%, specificity = 85%, κ = .35). Therefore, only the vigorous cut point of $\geq$ 2316 cpm, calculated using the 5 MET threshold, was included in the present study.

4.3.5 Data processing

ActiGraph data for the vertical axis were initially downloaded using ActiLife 5 software in 15-second epochs and transferred to an Excel file by the original research team. This is the unedited format in which data was received by the researchers in the present study. Subsequently, data were converted into counts per minute by summing four consecutive 15-second epochs. Data in the Excel file were manually screened by one researcher (AM) for spurious data to ensure that all included data were a result of activity and not due to error. The criteria used to identify spurious data were epochs containing > 15,000 cpm, with epochs above this threshold deemed not to be a result of physical activity (Esliger et al., 2005). No spurious data points were identified.

4.3.6 Statistical analysis

All statistical data were analysed using SPSS 21 IBM statistical package (SPSS IBM, New York, NY, USA).

Normality was assessed for all variables to ensure the use of an appropriate statistical test. Each variable was plotted using a histogram with normal distribution curves and a boxplot to produce a visual representation of the data distribution. Skewness and kurtosis were tested using z-scores, with < 1.96 representing normal distribution. Normality was additionally assessed using the Shapiro-Wilk test. For data that were not normally distributed, logarithmic and square root transformations were separately applied to the data and normality was retested. Descriptive statistics (mean ± SD) were calculated for age, sex, height, weight, and BMI of participants. The estimated time spent in each intensity for the nine sets of cut points were plotted on a bar chart with error bars showing 95% confidence intervals.

A one-way repeated measures ANOVA was used to examine differences in estimated time spent in each activity intensity (dependant variable) between the nine sets of cut points (independent variables). Four separate tests were conducted for sedentary, moderate, vigorous, and moderate to vigorous intensity cut points. Mauchly's Test of Sphericity was used to test for a significant difference between the variances of the differences between the cut points. A Mauchly's test score of $p < .05$ indicates that the data violates the assumption of sphericity, which increases the probability of Type II error. For data which violated this assumption, Greenhouse-Geisser estimates of sphericity were used to choose an appropriate correction for the interpretation of within-subjects effects. The Huynh-Feldt or Greenhouse-Geisser corrections were used for scores of $\varepsilon > .75$ or $\varepsilon < .75$, respectively (Field, 2011). Post-hoc pairwise comparisons with Bonferroni adjustments were additionally used to identify where significant differences occurred.

Data are also presented as percentages of the difference between the largest cut point estimate (E_L) and smallest cut point estimate (E_S) of time (minutes) spent in each intensity. Percentages were calculated for the effect of cut points on the percentage of the session conducted at a specific intensity using the following formula:

$$[(E_L - E_S) / \text{Total session time}] \times 100$$

4.4 Results

This section will present the results on the effect that cut points used has on the estimated time spent in activity intensities. Unless otherwise stated, results are presented in the format of mean (± SD).

4.4.1 Participants

Data for ten children with mild to moderate intellectual disabilities aged 10 to 12 years were included in these analyses. Participant descriptive statistics are presented in Table 4.3.

Table 4.3. Descriptive statistics (mean ± standard deviation) for participant variables

Characteristic	Boys (n = 6)	Girls (n = 4)	All (n = 10)
Age (yrs)	10.33 ± .52	10.75 ± .96	10.50 ± .71
Height (m)	1.42 ± .04	1.35 ± .11	1.39 ± .08
Weight (kg)	40.43 ± 5.39	37.43 ± 17.53	39.23 ± 11.00
BMI (kg/m^2)	19.96 ± 2.07	19.80 ± 6.69	19.89 ± 4.16

4.4.2 Activity session

The physical activity session lasted for a duration of 71 minutes. Subsequently, 710 epochs of data were included in the analyses. The mean and standard

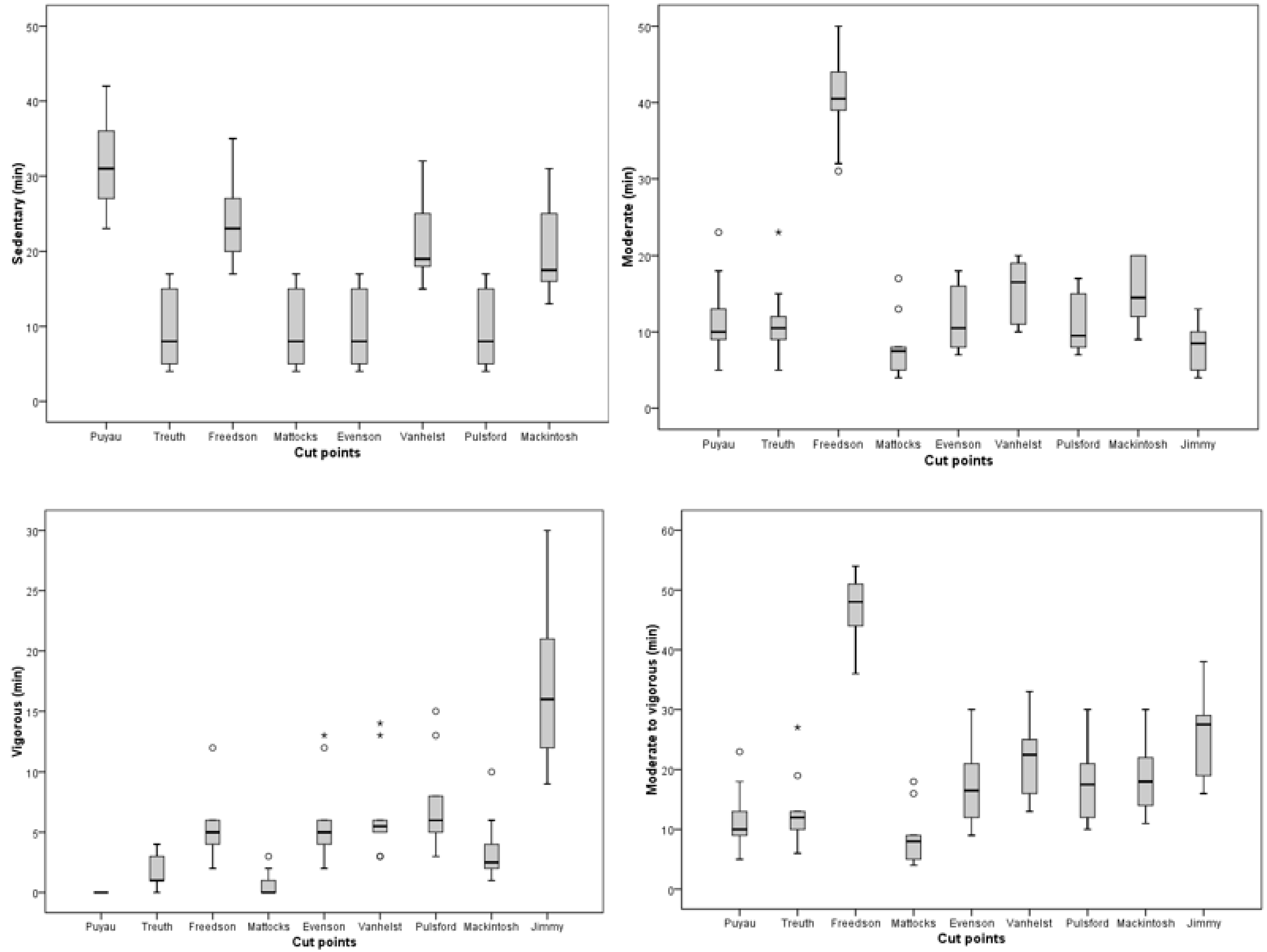


Figure 4.2. Effect of cut points on the estimation of physical activity intensity

deviations of time spent in each intensity for the seven sets of cut points are presented in Figure 4.2.

4.4.3 Effect of cut points

The mean differences between cut points for each intensity are presented in Tables 4.4–4.7. Results are presented in relation to each of the intensity cut points: sedentary, moderate, vigorous, moderate to vigorous.

4.4.3.1 Sedentary cut points

Data for the sedentary cut points were normally distributed, with non-significant Shapiro-Wilk scores ($p > .05$) for all variables, and did not violate the assumption of sphericity.

The results of the ANOVA show that the sedentary cut points used had a significant effect on the estimated time spent sedentary, $F(7, 63) = 201.60$, $p < .0001$. Estimated time spent sedentary ranged from 9.50 (± 4.97) to 31.90 (± 6.77) minutes.

The cut points developed by Treuth et al. (2004), Mattocks et al. (2007), and Pulsford et al. (2011) all estimated the study sample were sedentary for 9.50 (± 4.97) minutes. These cut points derived the lowest estimates of time spent in sedentary behaviour and were significantly different ($p < .0001$) from the other sets of cut points within this study; full statistics presented in Table 4.4. The cut points developed by Puyau et al. (2002) gave the highest estimate of time spent sedentary (31.90 ± 6.77 minutes). In comparison with the other cut points, this resulted in mean differences ranged from 7.40 (± .75) minutes to 22.40 (± 1.40) minutes.

In summary, the choice of cut points used had a significant effect on the estimated time spent sedentary. From the cut points included in this study, mean differences between cut points was as high as 22.40 minutes (± 1.40, $p < .0001$). Considering the session duration was 71 minutes, this equates to as much

Table 4.4. Differences between cut points for estimated time (minutes) spent sedentary
Data are presented as mean ± standard error and 95% confidence intervals of cut points 1 minus cut points 2

Cut points 1	**Cut points 2**							
	Treuth	Freedson	Mattocks	Evenson	Pulsford	Vanhelst	Mackintosh	Jimmy
Sedentary								
Puyau	22.40 ± 1.40** (16.55, 28.25)	7.40 ± .75** (4.27, 10.53)	22.40 ± 1.40** (16.55, 28.25)	22.40 ± 1.40** (16.55, 28.25)	22.40 ± 1.40** (16.55, 28.25)	10.30 ± .84** (9.44, 13.56)	11.50 ± .91** (9.44, 13.56)	---
Treuth	---	-15.00 ± .92** (-18.84, -11.16)	0.00 ± .00	0.00 ± .00	0.00 ± .00	-12.10 ± .85 (-14.02, -10.18)	-10.90 ± .96** (-14.91, -6.89)	---
Freedson	---	---	15.00 ± .91** (11.16, 18.84)	15.00 ± .91** (11.16, 18.84)	15.00 ± .91** (11.16, 18.84)	-9.50 ± .64** (-10.94, -8.06)	4.10 ± .35** (2.65, 5.55)	---
Mattocks	---	---	---	0.00 ± .00	0.00 ± .00	-12.10 ± .85** (-14.02, -10.18)	-10.90 ± .96** (-14.91, -6.89)	---
Evenson	---	---	---	---	0.00 ± .00	-12.10 ± .85** (-14.02, -10.18)	-10.90 ± .96** (-14.91, -6.89)	---
Pulsford	---	---	---	---	---	-12.10 ± .85** (-14.02, -10.18)	-10.90 ± .96** (-14.91, -6.89)	---
Vanhelst						---	1.20 ± .25* (.63, 1.76)	---

Difference between cut points significant at *p < .05 or **p < .001
Note: all results presented are based on original scores per minute

as a 30.99% higher estimate of time spent sedentary during the session between cut points.

4.4.3.2 Moderate intensity cut points

Data for the moderate intensity cut points were normally distributed, with non-significant Shapiro-Wilk scores ($p > .05$) for all variables. However, Mauchly's test of sphericity indicated that this assumption had been violated, $X^2(35) = 109.10$, $p < .05$, therefore the degrees of freedom were corrected using the Greenhouse-Geisser estimate of sphericity ($\varepsilon = .28$).

The ANOVA showed that the choice of moderate cut point had a significant effect on the estimated time spent in moderate intensity activity, $F(2.22, 19.96) = 93.47$, $p < .0001$. Estimated time spent in moderate intensity activity ranged from 8.10 (± 4.07) to 40.40 (± 5.74) minutes (Table 4.5).

There were no significant differences between the cut points developed by Puyau et al. (2002), Treuth et al. (2004), Evenson et al. (2008), and Pulsford et al. (2011), which estimated 11.40 (± 5.44), 11.10 (± 5.07), 11.60 (± 4.22), and 11.00 (± 3.89) minutes of moderate intensity activity, respectively. Furthermore, there were no significant differences between the Evenson et al. (2008), Mattocks et al. (2007) and Jimmy et al. (2013) cut points, which estimated 11.60 (± 4.22), 8.10 (± 4.07) and 8.40 (± 3.27) minutes, respectively. Conversely, the cut points developed by Freedson et al. (2005) estimated 40.40 (± 5.74) minutes of moderate intensity activity, which was significantly higher than all other cut points at $p < .0001$. The mean differences between cut points in comparison with the higher Freedson et al. (2005) estimate ranged from 25.10 (± 1.87) to 32.30 (± 2.29) minutes.

In summary, the cut points used had a significant effect on the estimated time spent in moderate intensity activity. Mean differences between cut points was as high as 32.30 minutes (± 2.29, $p < .0001$), which represents a 45.49% higher estimate of moderate intensity activity during the session.

Table 4.5. Differences between cut points for estimated time (minutes) spent in moderate intensity activity
Data are presented as mean ± standard error and 95% confidence intervals of cut points 1 minus cut points 2

Cut points 1	**Cut points 2**							
	Treuth	Freedson	Mattocks	Evenson	Pulsford	Vanhelst	Mackintosh	Jimmy
Puyau	.30 ± .60 (-2.04, 2.64)	-29.00 ± 2.67** (-40.13, -17.87)	3.30 ± .65* (.58, 6.02)	-.20 ± 1.49 (-6.42, 6.02)	.40 ± 1.57 (-6.14, 6.94)	-3.80 ± 1.30 (-9.70, 2.10)	-3.90 ± 1.35 (-9.52, 1.72)	3.00 ± 1.97 (-5.94, 11.94)
Treuth	---	-29.30 ± 2.48** (-39.65, -18.95)	3.00 ± .52* (.84, 5.16)	-.50 ± 1.06 (-4.92, 3.92)	.10 ± 1.13 (-4.62, 4.82)	-4.10 ± 1.15 (-9.33, 1.13)	-4.20 ± 1.20 (-8.46, -.06)	2.70 ± 2.04 (-6.57, 11.97)
Freedson	---	---	32.30 ± 2.29** (22.73, 41.78)	28.80 ± 1.74** (21.51, 36.09)	29.40 ± 1.75** (22.08, 36.72)	25.20 ± 1.71** (17.45, 32.96)	25.10 ± 1.87** (17.31, 32.89)	32.00 ± 1.67** (24.42, 39.58)
Mattocks	---	---	---	-3.50 ± 1.06 (-7.92, .92)	-2.90 ± 1.10 (-7.50, 1.70)	-7.10 ± 1.05* (-11.87, -2.33)	-7.20 ± .93** (-11.08, -3.32)	-.30 ± 1.71 (-8.06, 7.46)
Evenson	---	---	---	---	.60 ± .22 (-.32, 1.52)	-3.60 ± .83 (-7.39, .19)	-3.70 ± .54* (-5.95, -1.45)	3.20 ± 1.78 (-4.87, 11.27)
Pulsford	---	---	---	---	---	-4.20 ± .92* (-8.37, -.03)	-4.30 ± .68* (-7.16, -1.44)	2.60 ± 1.76 (-5.40, 10.60)
Vanhelst	---	---	---	---	---	---	-.10 ± .69 (-3.24, 3.04)	6.80 ± 1.27* (1.02, 12.59)
Mackintosh	---	---	---	---	---	---	---	6.90 ± 1.63 (-.51, 14.31)

Difference between cut points significant at *p < .05 or **p < .001
Note: all results presented are based on original scores per minute

4.4.3.3 Vigorous intensity cut points

The data for vigorous intensity cut points violated the assumptions of normality, therefore analyses were conducted on square root transformed data. Furthermore, Mauchly's test indicated that the assumption of sphericity had also been violated, $X^2(35) = 73.89$, $p < .05$, therefore the degrees of freedom were corrected using the Greenhouse-Geisser estimate of sphericity ($\varepsilon = .42$).

The results of the ANOVA show that cut points used had a significant effect on the estimated time spent in vigorous intensity activity, $F(3.40, 30.56) = 102.36$, $p < .0001$. During the sessions, participants spent the least amount of time at a vigorous intensity, with cut point estimates ranging from 0.00 (± .00) to 17.40 (± 6.54) minutes (Table 4.6).

The cut points developed by Puyau et al. (2002), which included a vigorous cut point of ≥ 8200 cpm, estimated that none of the study sample were active at a vigorous intensity during the session. On the other hand, the lowest cut point of ≥ 2316 cpm, which was developed by Jimmy et al. (2013), estimated that the participants were, on average, active at a vigorous intensity for 17.40 (± 6.54) minutes. Subsequently, the highest mean difference recorded was 17.40 minutes (± 2.07, $p < .001$) between Puyau et al. (2002) and Jimmy et al. (2013). For the remaining cut points, estimated time in vigorous intensity was 0.70 (± 1.01; Mattocks et al., 2007), 1.60 (± 1.51; Treuth et al., 2004), 3.50 (± 2.68; Mackintosh, Fairclough, Stratton, & Ridgers, 2012), 5.30 (± 2.67; Freedson et al., 2005), 6.10 (± 3.60; Evenson et al., 2008), 6.60 (± 3.81; Vanhelst, Beghin, Turck, & Gottrand, 2011), and 7.00 (± 4.00; Pulsford et al., 2011) minutes.

In summary, the choice of cut points used had a significant effect on the estimated time spent in vigorous intensity activity. The shorter duration of time that participants spent in vigorous activity during the session, as measured by all cut points, resulted in statistical differences between various cut points, with the greatest mean difference representing a 24.08% higher estimate of vigorous intensity activity.

Table 4.6. Differences between cut points for estimated time (minutes) spent in vigorous intensity activity
Data are presented as mean ± standard error and 95% confidence intervals of cut points 1 minus cut points 2

Cut points 1	Cut points 2							
	Treuth	Freedson	Mattocks	Evenson	Pulsford	Vanhelst	Mackintosh	Jimmy
Puyau	-1.60 ± .48 (-3.60, .39)	-5.30 ± .84* (-8.83, -1.77)	-.70 ± .34 (-2.10, .70)	-6.10 ± 1.14* (-10.86, -1.34)	-7.00 ± 1.27* (-12.29, -1.71)	-6.60 ± 1.20* (-12.08, -1.13)	-3.50 ± .85* (-7.04, .04)	-17.40 ± 2.07** (-26.80, -8.00)
Treuth	---	-3.70 ± .63* (-6.35, .1.05)	.90 ± .28 (-.26, 2.06)	-4.50 ± .76* (-7.69, -1.31)	-5.40 ± .86* (-8.99, -1.81)	-5.00 ± .83* (-8.78, -1.23)	-1.90 ± .50 (-4.01, .21)	15.80 ± 1.79** (-23.93, -7.67)
Freedson	---	---	4.60 ± .67* (1.80, 7.40)	-.80 ± .59 (-3.28, 1.68)	-1.70 ± .68 (-4.56, 1.16)	-1.30 ± .67 (-4.34, 1.74)	1.80 ± .84 (-1.71, 5.31)	-12.10 ± 1.66* (-19.67, -4.54)
Mattocks	---	---	---	-5.40 ± .96* (-9.40, -1.40)	-6.30 ± 1.04* (-10.66, -1.94)	5.90 ± 1.02* (-10.52, -1.28)	-2.80 ± .73** (-5.84, .24)	-16.70 ± 1.97** (-25.67, -7.73)
Evenson	---	---	---	---	-.90 ± .23 (-1.88, .08)	-50 ± .17 (-1.26, .26)	2.60 ± .60* (.09, 5.11)	-11.30 ± 1.25** (-16.98, -5.63)
Pulsford	---	---	---	---	---	.40 ± .22 (-.61, 1.41)	3.50 ± .72* (.50, 6.50)	-10.40 ± 1.19** (-15.83, -4.97)
Vanhelst	---	---	---	---	---	---	3.10 ± .66* (.11, 6.09)	-10.80 ± 1.15** (-16.04, -5.56)
Mackintosh	---	---	---	---	---	---	---	-13.90 ± 1.53** (-20.86, -6.94)

Difference between cut points significant at *p < .05 or **p < .001
Note: all results presented are based on original scores per minute

4.4.3.4 Moderate to vigorous cut points

Data for the moderate to vigorous intensity cut points were normally distributed, with non-significant Shapiro-Wilk scores ($p > .05$) for all variables. However, Mauchly's test of sphericity indicated that this assumption had been violated, $X^2(35) = 95.33$, $p < .05$, therefore the degrees of freedom were corrected using the Greenhouse-Geisser estimate of sphericity ($\varepsilon = .33$).

The moderate to vigorous intensity cut points included within this study resulted in significantly different estimates of time spent in moderate to vigorous intensity activity, $F(2.65, 23.80) = 200.57$, $p < .0001$ (Table 4.7).

Estimated time in this intensity ranged from 8.80 (± 4.64) to 46.50 (± 6.02) minutes, for the Mattocks et al. (2007) and Freedson et al. (2005) cut points, respectively. The estimated time of moderate to vigorous intensity activity based on the Freedson et al. (2005) cut points was notably, and significantly ($p < .001$), higher than the other cut point estimates. Conversely, the Mattocks et al. (2007) cut point estimated mean moderate to vigorous activity that was significantly ($p < .001$) lower in comparisons with the alternative cut points. The Puyau et al. (2002) and Treuth et al. (2004) cut points produced similar and non-significant estimates (11.40 ± 5.44 and 12.70 ± 6.24 minutes, respectively). Similarly, there were no significant differences between the cut points derived by Evenson et al. (2005), Pulsford et al. (2011), and Mackintosh et al. (2012) which estimated 17.70 ± 6.62, 18.00 ± 6.45, and 18.90 ± 5.93 minutes, respectively.

In summary, cut points used had a significant effect on the estimated time spent in moderate to vigorous intensity activity, although these significant differences were not present between all sets of cut points. The greatest mean difference between cut points (37.70 ± 1.65, $p < .0001$), represented a 53.10% higher estimate of the session spent in moderate to vigorous intensity activity.

Table 4.7. Differences between cut points for estimated time (minutes) spent in moderate to vigorous intensity activity
Data are presented as mean ± standard error and 95% confidence intervals of cut points 1 minus cut points 2

Cut points 1	Cut points 2							
	Treuth	Freedson	Mattocks	Evenson	Pulsford	Vanhelst	Mackintosh	Jimmy
Puyau	-1.30 ± .42 (-3.07, .47)	35.10 ± 1.94** (26.99, 43.21)	2.60 ± .52* (.42, 4.78)	-6.30 ± 1.08* (-10.80, -1.81)	-6.60 ± 1.01* (-10.83, -2.37)	24.70 ± 1.66** (17.15, 32.25)	-7.50 ± .93** (-11.40, -3.60)	20.70 ± 1.51** (13.82, 27.58)
Treuth	---	33.80 ± 1.96** (25.61, 41.99)	3.90 ± .71* (.95, 6.85)	-5.00 ± .82* (-8.41, -1.59)	-5.30 ± .78* (-8.54, -2.06)	-10.40 ± 1.04** (-15.11, -5.69)	-6.20 ± .74** (-9.30, -3.10)	-14.40 ± 1.27** (-20.16, -8.64)
Freedson	---	---	37.70 ± 1.65** (30.79, 44.61)	28.80 ± 1.74** (21.51, 36.09)	28.50 ± 1.68** (21.50, 35.50)	-9.10 ± .95** (-13.41, -4.79)	27.60 ± 1.56** (21.09, 34.11)	-13.10 ± 1.24** (-18.75, -7.45)
Mattocks	---	---	---	-8.90 ± 1.04** (-13.24, -4.56)	-9.20 ± .98** (-13.28, -5.13)	-13.00 ± 1.07** (-17.84, -8.16)	-10.10 ± .86** (-13.70, -6.50)	-17.00 ± 1.22** (-22.55, -11.45)
Evenson	---	---	---	---	-.30 ± .15 (-.94, .34)	-4.10 ± .80* (-7.72, -.48)	-1.20 ± .33 (-2.57, .17)	-8.10 ± 1.01** (-12.67, -3.53)
Pulsford	---	---	---	---	---	-3.80 ± .74* (-7.18, -.42)	-.90 ± .31 (-2.21, .41)	-7.80 ± .93** (-12.02, -3.58)
Vanhelst	---	---	---	---	---	---	2.90 ± .71 (-.31, 6.11)	-4.00 ± .49** (-6.25, -1.75)
Mackintosh	---	---	---	---	---	---	---	-6.90 ± .97* (-11.32, -2.48)

Difference between cut points significant at *p < .05 or **p < .001
Note: all results presented are based on original scores per minute

4.5 Discussion

The purpose of this chapter is to discuss the use decisions used within this thesis and to update and add to the existing literature on the effect of cut points in children with intellectual disabilities. The effect of cut points was examined by comparing the different estimations of time spent in various physical activity intensities between cut points using an intellectual disabilities-specific data set.

The large number of intensity-related accelerometer cut points which have been developed for typically developing children is impeding research efforts to quantify and compare physical activity levels. This is limiting researchers' understanding of children's physical activity behaviours and hindering the development of interventions (Trost et al., 2011). This problem is amplified in physical activity research relating to children with intellectual disabilities due to the lack of validity surrounding methods of data interpretation. Although the need for population-specific cut points for children with intellectual disabilities has been widely recognised, no cut points have been developed specifically for this group (Frey et al., 2008; Hinckson & Cutis, 2013). Researchers in this field therefore need to interpret accelerometer output by generalising typically developing cut points, which has been shown to introduce systematic error into results (Freedson et al., 2005).

The results of the present study show that for all intensities the choice of cut points results in significantly different estimates of physical activity intensity. The magnitude of the effect was largest for the sedentary and moderate to vigorous cut points, which resulted in a difference of up to 22.40 minutes and 37.70 minutes, respectively. Considering the duration of the session was 71 minutes, the variance between cut points represent a different classification of activity for up to 53% of the measurement period. This also highlights the important clinical effects of cut points, as reducing sedentary time and increasing moderate to vigorous intensity activity are integral aspects of the physical activity guidelines. Furthermore, as it is important to increase our understanding of the dose-response relationship in children with intellectual disabilities, this level of discrepancy between cut points will hinder future research in this field.

The following sections will discuss these findings in relation to previous research. Specifically, the wider effects of cut points on the consolidation of research and on our understanding of the physical activity behaviours of children with intellectual disabilities will be discussed.

4.5.1 Potential causes of cut point variance

Understanding possible causes of the differences between established cut points is important in advancing the field of physical activity measurement research in children with intellectual disabilities. The cut points examined within this study were calibrated using different devices, protocols, and criterion measures, which could all have attributed to the differences identified between cut points established for the same population. The following sections will discuss each of these possible causes of cut point variance in more detail.

4.5.1.1 Device

The majority of cut points examined within this study were calibrated using the AM7164 device, with more recent studies using the GT1M and GT3X. In relation to the effect of the device on the cut points calibrated, previous studies which have examined the comparability between devices, as discussed in Section 4.2.1.3, note that the AM7164 device records higher count values for the same movement in comparison with the GT1M. This is concurrent with the cut point values derived for moderate and moderate to vigorous intensity activity, as the cut points established using AM7164 are higher than those established using the GT1M device, with the exception of the Freedson et al. (2005) cut points. This suggests that the different internal mechanisms between devices used for calibration has an effect on the derived cut point thresholds. Therefore, it could be assumed that applying cut points derived using the AM7164 to data measured using the GT1M will underestimate time spent in the physical activity intensity of interest. However, as many non-significant differences were found between cut points derived using the AM7164 and the GT1M, the effect of device on the estimation of physical activity intensity may be limited. Furthermore, as many significant differences were found between cut points derived using the AM7164, this would also suggest that the device used for calibration does not have a consistent effect on intensity estimations.

As the development of the AM7164 and GT1M have been discontinued, the effects found for these devices may be less relevant for current and future research. However, numerous studies suggest the Evenson et al. (2008) cut points, which were derived using the AM7164, are most valid and recommend the use of these cut points (Clanchy et al., 2011; Reilly et al., 2008; Trost et al., 2011). Furthermore, recent studies which measured physical activity in children with intellectual disabilities used older generations of the ActiGraph, such as Boddy et al. (2015) which used the GT1M and applied the Evenson et al. (2008) cut points, therefore the effect of generalising cut points between devices is still an important issue.

The cut points developed by Jimmy et al. (2013), which were the only cut points calibrated using the GT3X device, produced significantly different estimates of time spent in vigorous and moderate to vigorous intensity activity, in comparisons with all other cut points. Although it is not possible to draw firm conclusions based on one study, this suggests that cut points calibrated using the GT3X are not comparable with cut points calibrated using older generations of the ActiGraph device. Previous research which investigated the comparability between devices found that the AM7164 and GT1M record higher count values for the same movement in comparison with the GT3X (Ried-Larsen et al., 2012; Rothney et al., 2008). This finding is concurrent with the developed cut points, as the count boundaries for the GT3X are notably lower and, as a result, the physical intensity estimates are significantly higher for the cut points developed by Jimmy et al. (2013). With the current recommendation being that accelerometer data is interpreted using the Evenson et al. (2008) cut points, the application of these cut points to data using the GT3X could introduce systematic error into the results, thus affecting the validity of conclusions (Clanchy et al., 2011; Reilly et al., 2008; Trost et al., 2011).

The differing internal mechanisms between devices and the effect that this has on the count output, as discussed in Section 4.2.1.3, is also apparent in many of the developed cut points. This suggests that cut points are affected by the device which was used for calibrated and, therefore, generalising cut points between different generations of the same ActiGraph device will introduce systematic error. However, as many significant differences were found for cut points calibrated using the same device, the influence of other factors has to be considered.

4.5.1.2 Protocol

Another possible effect on cut points and intensity estimations is the type of protocol used for calibration. The cut points discussed within this study were calibrated using various free-living and treadmill-based activities. The most notable difference between these protocol types, and the subsequent activities conducted, is that free-living activities are deemed to be more representative of children’s natural movement and play behaviours, compared to the constant walking/running associated with treadmill-based activities. Therefore, there is debate in the literature surrounding the comparability between treadmill and overground walking and the effect of the protocol on calibration (Lee & Hidler, 2008; Trost et al., 2006).

Although research involving children is limited, studies including an adult population report that treadmill walking results in lower vertical hip displacement and vertical ground forces in comparison to overground walking. Therefore, at the same speed, treadmill walking produces a significant and systematically lower count output in comparison with overground walking, although this effect is almost negligible for higher intensity running (Trost et al., 2006). This has important implications for calibration as the application of cut points derived during treadmill walking to overground, free-living activities will overestimate physical activity intensity. In the present study, only the cut points developed by Freedson et al. (2005) were based on a treadmill-only protocol. However, in comparison with the other cut points, the trend of a lower cut point for moderate intensity, but no clear difference for vigorous intensity could be an effect of the treadmill protocol. As a result, the physical activity estimates for moderate intensity activity are significantly higher for the Freedson cut points, which is concurrent with the finding that generalising treadmill-derived cut points to overground activity will overestimate time spent in the moderate intensity activity (Trost et al., 2011).

Another protocol-related factor which could affect the cut points calibrated is the criterion measure used. Two primary criterion measures were used to calibrate the cut points used within this study: calorimetry methods and direct observation. The most notable difference between these methods is that they measure different dimensions of activity, i.e. energy expenditure and activity type, respectively, which could affect the cut points calibrated. However, as there was only one set of cut points included which were calibrated against direct observation, there is little scope for discussion on possible effects, with no previous research identified which specifically investigated the effect of criterion measure on the calibration of intensity cut points.

4.5.1.3 Participant factors

There was a wide age-range of participants included in the samples reported within this study, which could affect the calibration and generalisation of cut points. Trost et al. (2011) conducted the most in-depth analysis of possible effects of cut points and found that age introduces the greatest bias into results. However, prior to the discussion of age-related effects, it is important to note that the effect of age is specific to the estimation of physiological outcomes based on accelerometer counts. Reilly et al. (2008) investigated the effect of age on raw accelerometer output for the ActiGraph, in the form of counts, and found no systematic variation in accelerometer output as a result of age. Therefore, the effect of age specifically relates to the classification of activity into intensity-related categories.

The resting metabolic rate of children decreases with age, therefore the use of cut points in a population with a different age from the calibration sample will introduce systematic error. Trost et al. (2011) investigated the effect of age on classification accuracy and found no significant differences from the criterion measure for the cut points developed by Evenson et al. (2008), Freedson et al. (2005), or Treuth et al. (2004); however, the Mattocks et al. (2007) and Puyau et al. (2002) cut points were affected by age for the classification of moderate to vigorous activity. When Trost et al. (2011) grouped children into ages 5 to 8 years, 9 to 10 years, and 11 to 12 years, classification accuracy significantly increased with each age increment for the Mattocks et al. (2007) and Puyau et al. (2002) cut points, with the AUC for ROC curve analysis ranging from AUC = .68−.82. With reference to the original calibration studies, the mean age of the study sample for Puyau et al. (2002) was 10.7 ± 2.9 years and 11.1 ± 2.9 years for boys and girls, respectively, with the mean age of the sample in Mattocks et al. (2007) 12.40 ± 0.02 years. This illustrates that cut points exhibit the highest classification accuracy when used in children with a similar age to the calibration sample, therefore generalising cut points between age groups will introduce systematic error.

Another age-related issue is with the interpretation and classification of energy expenditure-related data into intensity categories, which is generally done by applying MET thresholds to regression analysis output. Metabolic rate decreases with age, with resting O_2 corresponding with 1 MET decreasing from 6 to 3.50 mL/O_2/kg between the ages of 5 to 18 years (Schofield, 1985). Therefore, the following thresholds are currently recommended for classifying activity intensity in children; sedentary = < 1.5 METs, light = ≥ 1.5 and < 4 METs, moderate = ≥ 4 and < 6 METs, and vigorous activity = ≥ 6 METs (Trost et al., 2011). In comparison, physical activity intensity in adults is generally categorised using MET thresholds of 3, 6, and 9 METs for sedentary/light, moderate, and vigorous intensity, respectively.

Therefore, the classification of intensity in children based on adult thresholds creates substantial bias in energy expenditure estimations. Sallis, Buono, and Freedson (1991) report that the use of adult MET thresholds in children will underestimate energy costs by approximately 40%, 20%, and 5% for children aged 5 to 9 years, 10 to 15 years, and 16 to 17 years, respectively. Of the calibration studies reported in the present study, all cut point which were based on energy expenditure equations used child-specific thresholds, except for Freedson et al. (2005), which used adult MET thresholds of < 3, ≥ 3 to < 6, and ≥ 6 and < 9 METs for resting/light, moderate, and vigorous intensity, respectively. Considering the effects of using adult MET thresholds in children, it would be expected that the Freedson et al. (2005) cut points would underestimate time spent in each intensity. In contrast, however, the Freedson et al. (2005) cut points produced the highest estimate of moderate to vigorous intensity activity, and second highest sedentary time estimates in the present study.

Rather than only being an effect of the MET thresholds used, an alternative reason for the higher estimates produced by the Freedson et al. (2005) cut points is that a specific threshold for light intensity activity was not established.

As a result the upper boundary for sedentary activity is higher. Therefore, the results of the present study relating to the estimation of sedentary behaviour using the Freedson et al. (2005) cut points should be interpreted with caution, as the lower MET threshold corresponds to sedentary/light intensity in adults, and not solely sedentary activity in children. As previously discussed in Section 4.3.4, to address this limitation, ActiGraph now provides an additional cut point of 149 cpm to discriminate between sedentary and light intensity activity but, due to the lack of information on the calibration of this cut point, it was not included in this study. However, a limitation with this additional cut point is that the lower count boundary for sedentary could limit comparison between studies which used the original Freedson et al. (2005) cut point and the ActiGraph cut point. On the other hand, with the use of adult MET thresholds to derive the Freedson et al. (2005) cut points, it would be expected that time spent in each intensity would be underestimated (Freedson et al. 1991). However, the significantly higher estimates for moderate and moderate to vigorous intensity suggests that this significant difference in the cut points and intensity estimates is not an effect of the MET thresholds applied.

4.5.1.4 Analysis methods

To address the issues associated with applying MET thresholds to energy expenditure regression equations, more recent calibration studies, i.e. 2008 onwards, have analysed data using ROC curves. The output from this analysis does not require the application of MET thresholds, as specific activities and movements are classified into activity intensities prior to analysis, which is generally based on the compendium of energy expenditure for youth (Ridley, Ainsworth, & Olds, 2008). Although previous studies have discussed factors affecting the development of cut points, none of these studies have examined the possible effects of statistical analysis.

The current general consensus within the literature is that the cut points developed by Evenson et al. (2008), which were the first cut points calibrated using ROC curve analysis, are most valid (Clanchy et al., 2011; Reilly et al., 2008; Trost et al., 2011). ROC curve analysis reduces bias in intensity estimations in comparison with regression equations (Rothney et al., 2008). Therefore, the analysis used for calibration could have attributed to the higher validity reported for the Evenson et al. (2008) cut points by Trost et al. (2011) for typically developing children and Clanchy et al. (2011) for children with cerebral palsy. In the present study, the only cut points which exhibited no significant differences for the estimation of any intensity activity were the Evenson et al. (2008) and Pulsford et al. (2011) cut points, which were both calibrated using ROC curve analysis. Furthermore, with the exception of Puyau et al. (2002) and Freedson et al. (2005), the cut points calibrated by Evenson et al. (2008), Pulsford et al. (2011), and Mackintosh et al. (2012) using ROC curves were the only cut points in which no significant differences were found for the estimation of moderate to vigorous intensity activity.

As no studies were identified which discussed the effect of analysis, there is little scope for comparison with previous research. Although, as this study found fewer significant differences between cut points established using ROC curves in comparison with regression equations, this in concurrent with Rothney et al. (2008) who discussed ROC curves to be less prone to bias. Therefore, there is a growing amount of evidence to support the use of this method of analysis for accelerometer calibration in children.

4.5.1.5 Summary

Translating raw accelerometer counts into valid intensity cut points is a complex process, with many factors affecting calibration (Freedson et al., 2005). This section has highlighted the effects of generalising cut points between participants and ActiGraph devices, and discussed the possible effects of the protocol used, which will require consideration when designing a calibration protocol for children with intellectual disabilities. As choosing cut points is a use decision which faces many researchers using accelerometers, it is important to increase awareness of the significant effects that cut points can have on the estimation of time spent in each intensity.

It is important to note, however, that the effects found in the present study are, to an extent, study specific and should not be assumed the same across different studies. For example, if children spend the majority of the measurement period sedentary, such as during classroom time, the effect on moderate and vigorous cut points may be smaller, as there will be fewer measurement epochs which fall in the error area, i.e. the difference in cut point thresholds. Furthermore, as children with intellectual disabilities spend less time in vigorous intensity activity, the likelihood of epochs falling in the error area between cut points is less and therefore the effect of cut points may subsequently be less. To illustrate this further, for vigorous intensity, there is an 806 cpm difference between the cut points developed by Freedson et al. (2005) and Mackintosh et al. (2012). However, this difference in cut point boundaries did not result in significant differences in vigorous intensity activity estimates using the Get Active, Be Healthy data. In comparison, for moderate intensity activity a significant difference was found between the Evenson et al. (2008) and Mackintosh et al. (2012) cut points, although the actual count difference was only 135 cpm.

This highlights the difficulties in understanding and comparing the effect of cut points between studies. However, there is sufficient evidence to conclude that cut points used do have a significant effect on physical activity estimations in children with intellectual disabilities.

4.5.2 Effect of cut points on comparing existing research

When the results of the present study are considered in the context of comparison and consolidation of field-based research in children with intellectual disabilities, the cut points used will impact on results. The purpose of this section is to provide a “real-world” example of how the wide range of available cut points for typically developing children, and lack of population-specific cut points for children with intellectual disabilities, is limiting the comparison of research in this area.

Phillips and Holland (2011) and Boddy et al. (2015) used ActiGraph GT1M devices to measure free-living activity in children with intellectual disabilities. Therefore, as both studies include a similar sample and measured free-living physical activity using the same device, there should be a large scope for comparison between these studies. However, Boddy et al. (2015) used the moderate to vigorous Evenson et al. (2008) cut point of > 2295 cpm, whereas Phillips and Holland (2011) used a cut point of > 2802 cpm. This therefore equates to a discrepancy of 507 cpm for the estimation of moderate to vigorous intensity activity for data collected using the same accelerometer device and in similar study samples. However, Phillips and Holland (2011) did not provide an original calibration study reference for the cut points and therefore it is not clear how these cut points were established.

To further highlight the effect of cut points, the previous ANOVA analysis using Get Active, Be Healthy data was re-run with the inclusion of the > 2802 cut point for moderate to vigorous physical activity used by Phillips and Holland (2011) to allow comparison on intensity estimates with the Evenson et al. (2008) cut point used by Boddy et al. (2015). The use of the > 2802 cpm cut point resulted in a significantly lower estimation of daily time spent in moderate to vigorous intensity in comparison with the Evenson et al. (2008) cut point (mean difference = -3.70 ± .80 minutes; 95% CI -7.22, -.19, $p < .05$), which equates to a 20.90% mean difference. This trend is apparent in the original study results, with Boddy et al. (2015) reporting mean physical activity of 49.80 ± 3.80 (boys) and 45.30 ± 8.0 (girls) minutes per day, and Phillips and Holland (2011) reporting lower estimates of 28.20 ± 14.90 (boys) and 26.90 ± 6.50 (girls) minutes per day.

Although there is no way of attributing the cause of this variance in moderate to vigorous intensity activity reported between these studies solely to the cut points used, these findings show that results will be significantly affected by cut points. Therefore, the cut points used will limit the comparison between studies measuring physical activity using the same device, population, and environment, i.e. free-living.

4.5.3 Strengths and limitations

This study was the first to investigate the effect of cut points in children with intellectual disabilities and empirically highlight the effect that the lack of population-specific cut points for children with intellectual disabilities is having on this field of research. This study was the first to include the more recent cut points developed by Pulsford et al. (2011), Vanhelst et al. (2011), Mackintosh et al. (2012), and Jimmy et al. (2013), which used different analysis techniques, devices, and criterion measures to the studies included in previous reviews on the effect of cut points in typically developing children. Therefore, this allowed for a more in-depth understanding of the effect that cut points has on the estimation of physical activity intensity, and enabled discussion of additional factors, such as device and method of analysis used, which could affect cut points.

Not without limitations, the design of this secondary data analysis prevented the validation of existing cut points in children with intellectual disabilities. This would have enabled a greater understanding of which cut points are most valid, instead of only focussing on between-cut points effects. Furthermore, as the Get Active, Be Healthy data were only available on a group level, it was not possible to investigate the effect of participant characteristics on intensity estimates. This could have provided additional information on whether the participant-related effects discussed did have an effect on cut points and the estimation of physical activity intensity.

4.5.4 Conclusions

This study highlights that cut points can have significant and clinically meaningful effects on physical activity intensity estimations in children with intellectual disabilities, a finding which is consistent with previous research involving typically developing children (Anderson et al., 2005; Clanchy et al., 2011; Guinhouya et al., 2006; McClain et al., 2008; Reilly et al., 2008; Trost et al., 2011). Furthermore, investigating the effect of cut points in the previous studies conducted by Phillips and Holland (2011) and Boddy et al. (2015) highlights the real-world effect of cut points and the impact on comparability and validity between studies. Considering the limited research investigating physical activity in children with intellectual disabilities, compared with typically developing children, this is a major concern in terms of consolidating research.

Therefore, it is important to calibrate cut points specifically for children with intellectual disabilities. This will provide a single, population-specific method of data interpretation that will increase the validity of consolidation and comparison of research in this field. Furthermore, conducting calibration using the latest GT3X+/wGT3X+ device will keep the interpretation of accelerometer output in line with advancing technology. However, as the methods employed for calibration can impact on the cut points developed, the calibration protocol and analysis used need careful consideration. As there are many limitations associated with regression equations and the bias introduced with this analysis, the use of ROC curves is the emerging form of analysis for calibration research. Furthermore, as there are various methods which can be employed in a calibration study, such as the criterion measure and protocol used, it is important to firstly investigate the feasibility of these methods in children with intellectual disabilities.

Chapter 5 – Feasibility of a laboratory-based accelerometer calibration protocol for children with intellectual disabilities

5.1 Overview of this chapter

The results from Chapter 2 highlight the lack of population-specific methods for interpreting accelerometer output for children with intellectual disabilities. Furthermore, in Chapter 4, the effect of using cut points derived in typically developing children was discussed, with the calibration protocol identified as a possible cause for these differences. Therefore, when designing a calibration protocol, it is important to understand the methods which can be used, possible effects of methods on calibration, and the feasibility of these methods for use in children with intellectual disabilities. This chapter will discuss the elements which can be included in a calibration protocol and decide upon the design and methods which will be investigated in the present study. The experimental findings of this study will be discussed, with conclusions and recommendations provided for the design of a full-scale calibration study for children with intellectual disabilities.

5.2 Introduction

This section will discuss the elements of a calibration protocol and provide a rationale for the methods and measures employed within the present study.

5.2.1 Calibration protocol

Accelerometer calibration is an integral aspect of establishing measurement validity for physical activity in children with intellectual disabilities. However, calibration is complex with many factors to be considered when designing a study protocol. The protocol of a calibration study involves the concurrent measurement of a gold standard biological or behavioural measure of physical activity and accelerometry during activity, and can be laboratory- or field-based. Furthermore, where feasible, a protocol should include a measure of cardiorespiratory fitness, which will allow cardiorespiratory differences within and between groups to be understood, and is primarily important when using a physiological criterion measure (Freedson et al., 2012). Therefore, within a protocol, the primary decisions for a researcher are the setting (field- or laboratory-based), criterion measure, the activity protocol, and whether it is feasible to additionally measure cardiorespiratory fitness.

When designing a calibration protocol, it is important to be aware that the methods and measures employed are, to varying extents, dependent on each other. For example, if a stationary measure of indirect calorimetry is used as the criterion measure, it is not feasible to use this method in a free-living environment. On the other hand, if researchers wish to use a field-based protocol, a direct measure of cardiorespiratory fitness may not be feasible. Therefore, in addition to considering validity and feasibility of methods, researchers may have to prioritise one aspect of the design over another.

5.2.1.1 Criterion measure

The most commonly used criterion measures for calibration in children are indirect calorimetry and direct observation (Bassett et al., 2012). Although the strengths and limitations of these methods have been discussed in Section 1.4.3.1, this section will recap the strengths and limitations of these methods specific to calibration.

Direct observation has generally been used in calibration and validation studies involving toddlers, young children, and children with developmental disabilities (Capio, et al., 2010; Hislop, Bulley, Mercer, & Reilly, 2012; Mackintosh et al., 2012). An advantage of this method is that it is non-invasive and can be used effectively during free-play or activity sessions. Therefore, this allows the calibration protocol to include activities which are commonly conducted by children. However, as this is a behavioural measure, which is a proxy for physiological outcomes, it requires validation for the interpretation of intensity-related outcomes. As the metabolic costs of activity can vary between groups, the use of standardised energy expenditure costs or thresholds could introduce systematic error.

On the other hand, calorimetry methods enable accelerometer counts to be calibrated against a directly measured physiological dimension of physical activity. The calorimetry methods used in previous calibration studies are stationary metabolic measures, portable metabolic measures, and whole room calorimetry, which have differing advantages and disadvantages specific to use in a calibration study (Bassett et al., 2012). Portable metabolic measures allow children to participate in free-living activities; however, it is an expensive measure and may not be feasible for concurrent measurements in multiple children. Therefore, its limited feasibility may restrict calibration to a controlled or laboratory environment. Stationary indirect calorimetry provides a valid measure of $\dot{V}O_2$ or energy expenditure; however, the equipment required for this measure is invasive and restrictive. Therefore, calibration will be laboratory-based, with the activity protocol not being fully representative of children's play behaviours. Finally, whole room calorimetry enables free-play activities to be conducted in a confined environment; however, there are feasibilities issues relating to the multiple-hour measurements required in the calorimeter environment, such as participant comfort (Oortwijn et al., 2009).

As there is little known about the metabolic costs of physical activity in children with intellectual disabilities, the use of a physiological criterion measure will increase our understanding of the $\dot{V}O_2$ or energy expenditure requirements of various activities, in comparison with typically developing children. Furthermore, the use of indirect calorimetry will enable intensity MET thresholds to be based on direct measurements. In contrast, if a criterion measure is not a direct physiological measure but a behavioural measure, such as direct observation, then activity intensity is decided upon prior to data collection using the energy expenditure compendium, which requires validation (Bassett et al., 2012; Ridley et al., 2008). However, as the metabolic costs of activity can vary between groups, the use of standardised energy expenditure costs or thresholds could introduce systematic error. Therefore, as accelerometer calibration has not previously been conducted in children with intellectual disabilities, the use of indirect calorimetry will increase our knowledge of the energy costs of activity in this population group and allow calibration to be conducted on direct measurements.

5.2.1.2 Activity protocol

Deciding upon activities for use in a calibration protocol is important, as the validity of the cut points calibrated will be dependent on the extent to which the activity protocol is representative of the types and intensities of activities conducted by the study population (Welk, 2005). Calibration protocols therefore have to find the balance between activities which result in the desired intensity level, but which are also common movements among the population of interest. This is difficult as the patterns of activities vary considerably between populations. Understanding the activities commonly completed by children with intellectual disabilities would benefit the development of a calibration activity protocol. A scientific approach to this would be to design an activity protocol based on previous research which reports activities commonly conducted by children with intellectual disabilities (Bassett et al., 2012).

Children with intellectual disabilities have been reported to be a sedentary population, although the research relating to the types of physical activities completed by this population is limited. Television and computer time have previously been measured using parental logs, with mean afterschool screen time of 83 ± 64 minutes per day reported, although this was in a small sample of only 9 children with intellectual disabilities (Foley & McCubbin, 2009). In relation to types of physical activities conducted, adolescents with intellectual disabilities have a preference for walking, jogging, and sports (Lin, Lin, Lin, Chang, Wu, & Wu, 2010). However, Shine, Perry, and Weiss (2012) report that only 15% of children with severe and profound intellectual disabilities participate in team sport activities, suggesting that level of intellectual disabilities may affect activity preferences - although this could also be a result of limited opportunities to participate in sports.

Conclusions regarding the types of activities that children with intellectual disabilities participate in is difficult due to the lack of population-specific research. Furthermore, this research is limited by small sample sizes and subjective proxy-respondent measures. However, previous research suggests that activity preferences in children with intellectual disabilities is affected by disability type and severity. It is also important to note that these studies focus on activities which children with intellectual disabilities prefer to participate in, and not necessarily the activities in which they commonly do participate in. The dearth of research aiming to understand the types of physical activities that children with intellectual disabilities regularly participate in therefore limits the development of an activity protocol with known population-appropriate activities.

However, as indirect calorimetry is to be used as the criterion measure within the present study, this limits the protocol to a laboratory environment. A laboratory-based protocol generally involves treadmill-based activities. An advantage of a treadmill-based protocol is that activities can be completed at a constant and predetermined intensity. Furthermore, as walking and running are commonplace movements, and are the type of movements most accurately measured by accelerometers, it is important that these activities are included in a calibration protocol (Welk, 2005). To counter-balance the effect of constant treadmill activities, free-living activities can also be included which will more accurately represent the sporadic nature of children's activity behaviours. It was suggested by Kim et al. (2012) that at least six activities should be completed in a laboratory-based protocol, with three of these being of moderate to vigorous intensity, and a combination of treadmill-based and free-living. Activity intensity is generally categorised into four levels: sedentary, light, moderate, and vigorous. Puyau et al. (2002) define sedentary activity as being seated or reclined with minimal movement, and light, moderate, and vigorous as being in a standing position with low, medium, and high levels of exertion, respectively.

Protocols can vary, however, as no standardised guidelines exist for the type of activities to be conducted, or the amount of time these activities should be conducted for. Therefore, to ensure the development of a protocol which contains activities which children with intellectual disabilities can complete, the feasibility of various activities conducted in previous calibration protocols in typically developing needs to be investigated. This will increase our understanding of the types of activities that children with intellectual disabilities enjoy and whether it is feasible to generalise an activity protocol used in typically developing children to children with intellectual disabilities.

5.2.1.3 Measurement of cardiorespiratory fitness

Children with intellectual disabilities have lower levels of cardiorespiratory fitness than their typically developing peers (Pitetti, Yarmer, & Fernhall, 2001). Due to the effect that cardiorespiratory fitness has on energy expenditure and $\dot{V}O_2$ during activity, the generalising of cut points calibrated in a population with higher fitness could lead to an underestimation or misclassification of activity intensity for a population with lower fitness, which could have significant implications on results. Due to validity issues such as these, Freedson et al. (2012) notes the importance of investigating and classifying fitness for calibration studies, which will provide information on relative activity and health.

Maximal oxygen uptake has been widely acknowledged as a valid measure of aerobic fitness in children (Dencker & Anderson, 2011). In a laboratory setting, $\dot{V}O_{2max}$ is directly measured using a maximal exercise test to exhaustion with respiratory gas exchange measurements. The primary criterion for the attainment of $\dot{V}O_{2max}$ during a maximal test is a plateau in $\dot{V}O_2$ with increasing workload. However, if this primary criterion is not achieved, the attainment of $\dot{V}O_{2max}$ can be confirmed if two of the three following criteria are met: high levels of blood lactate post-test, respiratory exchange ratio (RER) > 1.00, and heart rate within 10 beats per minute (bpm) of maximal estimation heart rate (Howley, Bassett, & Welch, 1995). If these criteria are not met, data are reported as the peak oxygen uptake ($\dot{V}O_{2peak}$) score recorded during the test.

Although measuring $\dot{V}O_{2max}$ is the most valid measure of cardiorespiratory fitness, there are limitations with this method which researchers need to be aware of. There is some debate in the literature surrounding the criteria for the attainment of $\dot{V}O_{2max}$, such as the lactate or RER thresholds used, and the subjectivity of viewing the plateau in $\dot{V}O_2$. As a result, the primary and secondary criteria used can vary between studies (Howley et al., 1995). Therefore, the use of different criteria between studies can introduce measurement error into results. Furthermore, another potential cause of error is the use of $\dot{V}O_2$ as the primary outcome, as this is technically not a measurement but a calculation based on expired oxygen and carbon dioxide fractions, ventilation, and is affected by barometric pressure, gas temperature, and water pressure of the gas. To reduce the effect of these limitations, it is important for researchers to fully describe the methods and criteria used.

The primary methods of cardiorespiratory fitness testing in a laboratory setting are treadmill- and cycle ergometer-based graded exercise tests. The protocol of a treadmill-based graded exercise test, which involves incremental increases in gradient, is more effective at producing $\dot{V}O_{2max}$ scores compared to a cycle-based test due to less localised muscle fatigue (American College of Sports Medicine, 2013; McArdle, Katch, & Pechar, 1973). A treadmill based protocol has been validated for children with intellectual disabilities for the achievement of $\dot{V}O_{2peak}$ (Fernhall et al., 2000). However, the effectiveness and feasibility of this test for the attainment of $\dot{V}O_{2max}$ has not been investigated as part of a wider calibration protocol.

The attainment of $\dot{V}O_{2max}$ using a graded exercise test is dependent on participants' willingness to exercise to exhaustion and motivation to complete strenuous exercise. Participant familiarisation to the procedures and equipment prior to testing can improve the effectiveness of the test. Specifically, adherence to the testing protocols can be improved with explanation and demonstration of the test, and sufficient practice allowed. Furthermore, the use of verbal encouragement and rewards can improve participant motivation (Rintala, McCubbin, & Dunn, 1995). As little is known about the effectiveness of achieving $\dot{V}O_{2max}$ scores using a treadmill-based graded exercise in children with intellectual disabilities, this is another important aspect of a calibration protocol for which feasibility needs to be investigated. Furthermore, as a graded exercise test should end with exhaustion, the feasibility of including this test in a calibration protocol, i.e. whether the energy requirements of other aspects of the calibration protocol affect children's ability to effectively complete the graded exercise test, needs to be examined.

5.2.2 Summary

Designing a calibration protocol is complex. There are various methods and measures which can be used, each having an effect on calibration. Therefore, it is important to design a calibration study relative to the population of interest. As little is known about the energy costs of physical activity in children with intellectual disabilities, a direct criterion measure using stationary indirect calorimetry will increase our knowledge and help ensure valid calibration. These measurements, however, have not been extensively conducted in children with intellectual disabilities. Furthermore, many of the activities used in previous studies involve sport-specific skills, co-ordination, and concentration, and therefore may not be suitable for children with intellectual disabilities. Furthermore, testing cardiorespiratory fitness is important for calibration. Although treadmill-based tests have been validated for children with intellectual disabilities, the feasibility of including a maximal exercise test within a calibration protocol in unknown, and requires further investigation. Therefore, prior to conducting a full-scale calibration study, it is important to design an effective protocol for children with intellectual disabilities.

5.2.3 Research questions

The purpose of the present study is to test the feasibility of an accelerometer calibration protocol for children with intellectual disabilities. The findings from this study will inform the protocol of a future calibration study that will aim to calibrate accelerometry in children with intellectual disabilities.

The research questions being investigated in this chapter are:

RQ 6: Is it feasible to recruit children with intellectual disabilities from additional support needs schools to participate in a calibration study?

RQ 7: Are activities conducted in a calibration protocol designed for typically developing children feasible for children with intellectual disabilities?

RQ 8: Is it feasible to measure resting energy expenditure (REE) in children with intellectual disabilities?

RQ 9: Is a treadmill-based graded exercise test to measure $\dot{V}O_2max$ feasible in children with intellectual disabilities?

RQ 10: Is the use of respiratory gas exchange equipment feasible in children with intellectual disabilities?

RQ 11: Does altering Ultima CPX breath-by-breath system threshold settings have an effect on $\dot{V}O2$ in children with intellectual disabilities and typically developing children?

RQ 12: Is there a significant difference in the relationship between $\dot{V}O2$ and accelerometer counts between children with intellectual disabilities and typically developing children?

2.3 Method

5.3.1 Ethical considerations

This study was approved by the Medical, Veterinary, and Life Sciences College Ethics Committee, University of Glasgow (Appendix i). Written informed consent was required from both participants and parents. If any child was unable to read or sign the consent form due to disability severity, parents were asked to read it to them and verbal consent was attained from the child. Each participant received a £30 voucher for participating and all participant and parent travel expenses were reimbursed.

5.3.2 Participants

5.3.2.1 Recruitment

Participants with intellectual disabilities were recruited from additional support needs schools in Glasgow, Scotland in May/June 2013. A researcher (AM) visited two schools, explained the study to children, and handed out parent and child information packs and consent forms (Appendices ii and iii). If children were interested in participating, parents were asked to return a parent and child consent form to the researcher to allow discussion regarding participation. A convenience sample of typically developing children was recruited from the Glasgow area.

5.3.2.2 Inclusion and Exclusion criteria

The inclusion and exclusion criteria for this study were developed to ensure that an appropriate sample of children with intellectual disabilities were recruited. As calibration is affected by age, this type of study requires a relatively homogeneous sample in terms of age; however, a wider age range of children were included in this study to increase our understanding of the feasibility of the methods and measures for children of various ages. Furthermore, to ensure participants could safely complete the protocol, and to prevent movement-related factors confounding the results, children had to be independently ambulatory. Therefore, the following inclusion and exclusion criteria were used as the basis for study recruitment:

Inclusion criteria:

- have intellectual disabilities
- aged between 8 to 14 years
- independently ambulatory

Exclusion criteria:

- having a physical disability
- having a developmental disability, without a specific diagnosis of intellectual disabilities

For the sample of typically developing children, the inclusion and exclusion criteria were:

Inclusion criteria:

- aged between 8 to 14 years
- independently ambulatory

Exclusion criteria:

- have intellectual disabilities
- have a physical disability

5.3.3 Protocol

The study protocol, which is described in Table 5.1, was designed based on previous research relating to laboratory-based calibration studies involving typically developing children. A summary of these previous laboratory-based protocols, which is adapted from McGarty, Penpraze, and Melville (2015), is presented in Table 5.2. This study was conducted in three main phases: 1) familiarisation, 2) preparation, and 3) data collection. All experimental procedures were conducted in an exercise laboratory at the University of Glasgow, with two researchers present for each session. All sessions were conducted between August and November, 2013.

Initially, the sample of typically developing children completed the protocol during one session. However, to account for the sample of children with intellectual disabilities possibly requiring a longer familiarisation and preparation phase, the calibration protocol for this group was conducted over two separate sessions, as presented in Figure 5.1. Therefore, for children with intellectual disabilities, the familiarisation and preparation phases were conducted during session 1. Data collection was conducted in two stages; stage one was conducted during session one, with stage two completed during the second session.

5.3.3.1 Familiarisation phase

The aim of the familiarisation phase was to make the participants feel as comfortable as possible in the laboratory environment. Participants were introduced to the researchers involved and shown around the laboratory environment and surrounding areas; this included being shown the changing rooms and discussing general safety procedures. This also allowed an opportunity for participants and parents to ask any questions.

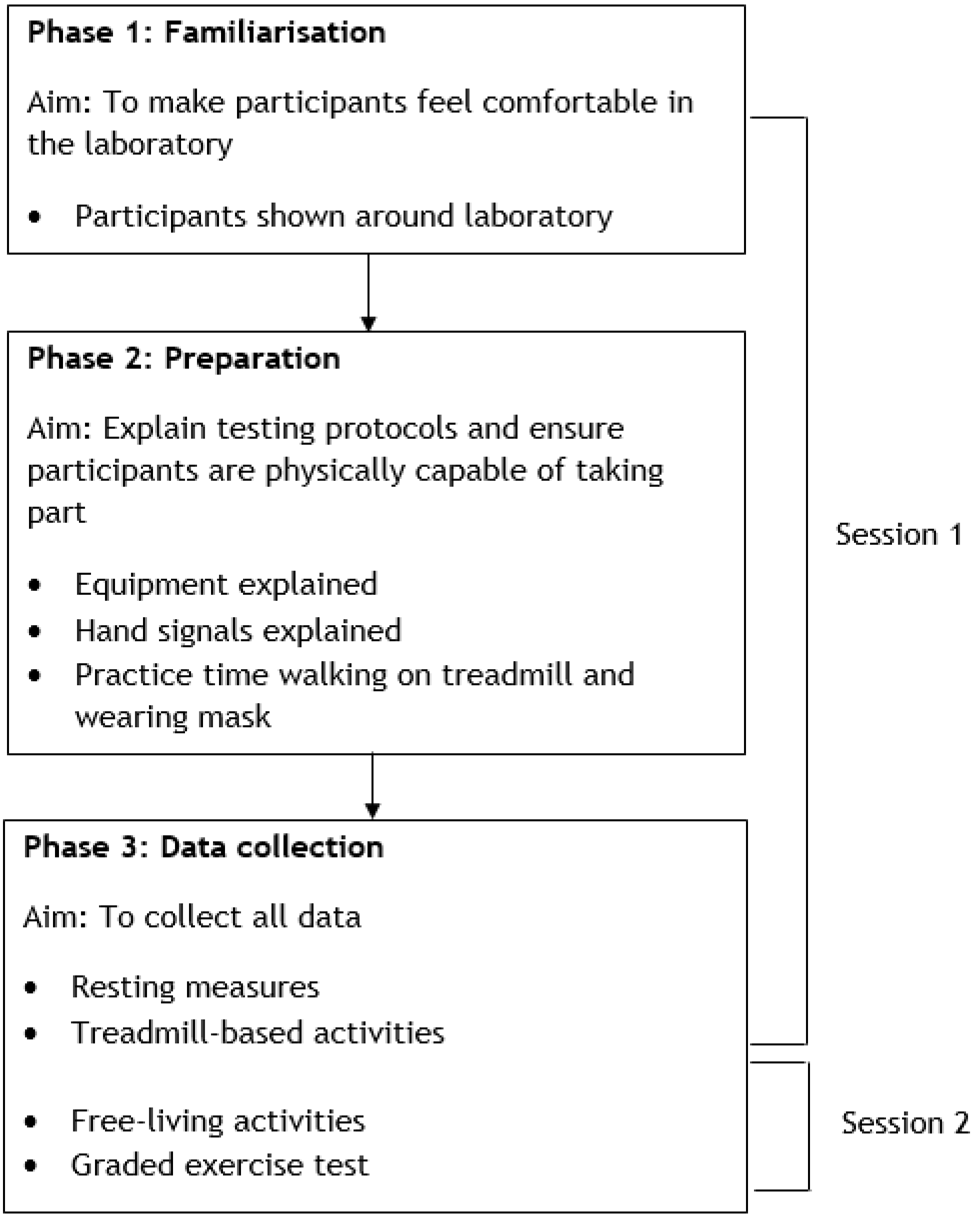


Figure 5.1. Flow chart of study procedures conducted during each session

5.3.3.2 Preparation phase

The aim of the preparation phase was to allow participants to become more familiar with the equipment and procedures, and to ensure they were physically capable of safely completing the protocol. If any participant was uncomfortable or having difficulties, the researcher made a judgement as to whether another preparation session was required, or whether the participant should not take part in the data collection phase.

The purpose of the equipment to be used was simply explained, e.g. “the accelerometer is worn around the waist and measures movement”. Participants practiced breathing through the respiratory collection mask and walking on the treadmill. Due to the nature of the testing and the respiratory gas exchange mask, hand signals were used for communication; thumbs up = okay/able to continue, rocking of hand = tiring/starting to struggle, horizontal movement of hand = stop. The participants were shown these signals and allowed to practice them. It was explained that throughout testing, and before each increase of gradient during the graded exercise test, they will be asked if they feel able to continue; the participant was asked to communicate using these signals.

5.3.3.3 Data collection phase

The first phase of data collection was conducted during the latter part of the first laboratory session. Height and weight measurements were taken, REE was measured, and the treadmill-based activities were completed. The second phase of data collection was conducted during the second session. During this session, participants were asked to complete all free-living activities and the graded exercise test.

5.3.3.3.1 Resting measures

Respiratory gas exchange was measured for 15 minutes to allow resting $\dot{V}O_2$ to be established. Throughout this measurement, participants sat in a reclined position and watched an age-appropriate DVD.

5.3.3.3.2 Activities

The activity protocol was designed based on activities included within previous laboratory-based studies involving typically developing children, which are described in Table 5.2. As there is limited research describing the activities commonly participated by children with intellectual disabilities, a wide-range of activities were included. Furthermore, treadmill and free-living activities were included to increase the ecological validity of the movements conducted. Activities also ranged from sedentary to vigorous intensity, and were skill-specific and non-skill-specific.

Table 5.1. Measurements and activities performed during laboratory sessions

Activity	Description
<u>Session 1</u>	
Rest	Sitting in reclined position watching DVD
Treadmill-based activities	
Light intensity	
3 km/h	Walking at 3 km/h at zero gradient
Moderate intensity	
6 km/h	Jogging at 6 km/h at zero gradient
5 km/h at 5%	Walking briskly at 5 km/h at 5% gradient
Vigorous intensity	
8 km/h	Running at 8 km/h at zero gradient
<u>Session 2</u>	
Free-living activities	
Sedentary	
Sitting playing computer game	Sitting playing handheld Nintendo DS
Watching DVD	Sitting watching DVD
Drawing	Sitting drawing
Light intensity	
Passing football	Passing a football with a researcher
Playing catch	Standing throwing/catching a ball with a researcher
Standing playing computer game	Standing playing handheld Nintendo DS
Moderate intensity	
Step aerobics	Continual stepping on and off aerobic step
Hula hoop	Continual twirling of hula hoop around the waist
Interactive computer game	Playing interactive bowling on an Xbox Kinect
Vigorous intensity	
Jumping jacks	Continual jumping jacks/star jumps
Graded Exercise test	Treadmill-based incremental fitness test

Including a wide range of activities will increase our knowledge about the type of activities children with intellectual disabilities enjoy participating in and also the activities which they are physically able to complete. This will inform the development of a future calibration protocol.

Participants were asked to complete 14 activities for 5 minutes (Table 5.1). These types of activities have been extensively conducted in calibration studies involving typically developing children (Kim et al., 2012). The intensity classifications were based on those defined by Puyau et al. (2002). Prior to conducting the activities, participants completed a 2-minute treadmill-based warm-up. Additionally, rest periods were given between activities, specifically during moderate and vigorous intensity activities, to allow measurements to return to within a resting range.

5.3.3.3.3 Graded exercise test

Participants walked on the treadmill at a constant and self-selected pace. The gradient was increased from zero in increments of 2.5% every 2 minutes. If a participant reached the maximum treadmill gradient (20%), the speed was then increased by 1 km/h every minute. The test ended when the participant reached the point of exhaustion or felt unable to continue, or if the researcher deemed it unsafe to continue. Verbal encouragement was given to participants throughout the test.

This protocol of increasing gradient in increments of 2.5% has been successfully conducted in adults with intellectual disabilities (Fernhall & Tymeson, 1987), with similar gradients used in children without intellectual disabilities. Concurrent with previous protocols conducted in adolescents with intellectual disabilities, walking speed was self-determined (Pitetti, Jongmans, & Fernhall, 1999). It was suggested by the American College of Sports Medicine (2013) that the use of self-determined pace for exercise testing in those with disabilities may prevent the test being discontinued due to participant anxiety.

The primary criterion for the attainment of $\dot{V}O_{2max}$ was a plateau in $\dot{V}O_2$ with increased workload. A secondary criteria of an increased RER > 1.0 and a heart rate within 10 bpm of an age-adjusted estimate of maximal heart rate were also

Table 5.2. Activity protocols of previous laboratory-based calibration studies involving children

Study	Participant age range (years)	Treadmill Activities	Treadmill Activity Time (min)	Free-living Activities	Free-living Activity Time (min)	Criterion measure
Chu et al. (2003)	11-15	Walk 4.5 km/h	5			Respiratory gas exchange (indirect calorimetry) (stationary)
		Run 6.6 km/h	5			
		Run 8.8 km/h	5			
Eston et al. (1998)	8-10	Walk 4 km/h	4	Playing catch	4	Respiratory gas exchange (indirect calorimetry) (stationary)
		Walk 6 km/h	4	Hopscotch	4	
		Run 8 km/h	4	Sitting crayoning	4	
		Run 10 km/h	4			
Evenson et al. (2008)	6-8	Light		Sedentary		Respiratory gas exchange (indirect calorimetry) (portable)
		Walk 2 mph	7	Rest	15	
		Moderate		Sitting watch DVD	7	
		Walk 3 mph	7	Sitting colouring in books	7	
		Vigorous		Moderate		
		Run 4 mph	7	Stair climb	7	
				Dribble basketball	7	
				Vigorous		
				Cycle on stationary bike	7	
				Jumping jacks	7	

Table 5.2. Continued

Study	Participant age range (years)	Treadmill Activities	Treadmill Activity Time (min)	Free-living Activities	Free-living Activity Time (min)	Criterion measure
McMurray et al. (2004)	8-18	Walking 4 km/h	10	Standing arcade game	10	Respiratory gas exchange (indirect calorimetry) (portable)
		Walking 5.6 km/h	10	Stretching	10	
		Running 8km/h	10	Sweeping	10	
				Vacuuming	10	
				Shovelling	10	
				Stair climb (88 steps/min)	10	
				Rope skipping	10	
Pate et al. (2006)	3-5			Rest	10	Respiratory gas exchange (indirect calorimetry) (portable)
				Walk 2 mph	5	
				Walk 3 mph	5	
				Jog 4 mph	5	
Puyau et al. (2002)	6-16	<u>Light</u>		<u>Sedentary</u>		Whole room calorimetry
		Walk 2.5 mph	10	Sitting playing Nintendo	20	
		<u>Moderate</u>		Arts and crafts	20	
		Walk 3.5 mph (6-7yrs)	10	Free play sitting, e.g. cards, lego	20	
			10	<u>Light</u>		
		Walk 4 mph (8-16yrs)		Aerobic warm up	10	

Table 5.2. Continued

Study	Participant age range (years)	Treadmill Activities	Treadmill Activity Time (min)	Free-living Activities	Free-living Activity Time (min)	Criterion measure
Puyau et al. (2002) Continued		<u>Vigorous</u>		<u>Moderate</u>		
		Jog 4.5 mph (6-7yrs)	10	Tae Bo martial arts exercises	10	
		Jog 5 mph (8-10yrs)	10	Free play standing, e.g. hula hoop,	10	
		Jog 6 mph (11-16yrs)	10	throwing ball, jumping jacks		
Puyau et al. (2004)	7-18	Walk 2 mph	7	Rest	30	Whole room calorimetry
		Walk 3.5 - 4 mph	7	Sitting playing handheld Nintendo	20	
		Jog/run 4.5 – 7 mph	7	Sitting playing computer	20	
				Cleaning (dusting)	10	
				Aerobic exercises	12	
				Ball toss	10	
Rowlands et al. (2004)	9	Walk 4 km/h	4	Sitting playing computer game	10	Respiratory gas exchange (indirect calorimetry) (stationary)
		Walk 6 km/h	4	Pass football	4	
		Run 8 km/h	4	Hopscotch	4	
		Run 10 km/h	4			

Table 5.2. Continued

Study	Participant age range (years)	Treadmill Activities	Treadmill Activity Time (min)	Free-living Activities	Free-living Activity Time (min)	Criterion measure
Treuth et al. (2004)	13-14			Rest	15	Respiratory gas exchange (indirect calorimetry) (portable)
				Sitting watch TV	7	
				Sitting playing computer game	7	
				Sweep floor	7	
				Walk 2.5 mph	7	
				Walk 3.5 mph	7	
				Step aerobics	7	
				Ride bike	7	
				Shoot baskets	7	
				Stair walk	7	
				Run 5 mph	7	
Trost et al. (1998)	10-14	Walk 3mph	5			Respiratory gas exchange (indirect calorimetry)
		Walk 4 mph	5			
		Jog 6 mph	5			

Table 5.2. Continued

Study	Participant age range (years)	Treadmill Activities	Treadmill Activity Time (min)	Free-living Activities	Free-living Activity Time (min)	Criterion measure
Vanhelst et al. (2010)	10-16	<u>Light</u> Walk 1.5 km/h, 3% gradient <u>Moderate</u> Walk 3 km/h, 3% gradient Run 4 km/h, 3% gradient <u>Vigorous</u> Run 6 km/h, 3% gradient	15 mins consecutive	<u>Sedentary</u> Lying in bed watching TV Sitting reading Sitting playing computer game <u>Light</u> Standing drawing Passing football	15 mins consecutive	Respiratory gas exchange (indirect calorimetry)

used (Howley et al., 1995; Power & Howley, 2007). Predicted maximal heart rate was calculated using the equation:

$$\text{Heart rate}_{max} = 220 - \text{age (years)}$$

5.3.4 Measures

Throughout the data collection phase, respiratory gas exchange measurements were collected and two accelerometers were worn. Participants also wore a heart rate monitor during the graded exercise test.

5.3.4.1 Respiratory gas exchange

The criterion measure of interest in this study was $\dot{V}O_2$, which was measured through respiratory gas exchange, using the stationary metabolic cart methodology as described in Section 5.2.1.1. Respiratory gas exchange was measured using the Ultima CPX (Medical Graphics, MN, USA) which analyses expired gases on a breath by breath basis. Prior to each test, airflow, ventilatory volume, and gas analysers were calibrated using standard measures in accordance with the manufacture's guidelines. Participants wore a preVent (Medical Graphics, MN, USA) material mask which covers the nose and mouth. This was attached directly to a bidirectional flow meter, a sampling line, and measurement sensor. Data were recorded using standard threshold settings of:

- minimum 50mL $\dot{V}O_2$ and $\dot{V}CO_2$
- minimum 180mL tidal volume
- RER between 0.5 and 2.6.

5.3.4.2 Accelerometry

The ActiGraph wGT3X+ (Actigraph LLC, Pensacola, FL) device was used to record physical activity. Prior to each session, each device was initialised according to manufacturer's specifications. Throughout data collection, participants wore two devices on their waist, with one positioned on the hip (above the iliac crest) and one at the centre of the back. As data varies between device positions, the use of multiple devices in calibration studies enables inter-instrument comparisons and the calibration of placement-specific cut points (Freedson et al., 2012). Therefore, this allowed the feasibility of using multiple devices to be investigated. In line with manufacturer guidelines for use in children, the devices were independently attached using an elastic belt.

5.3.4.3 Heart rate

Heart rate was measured using a chest-worn heart rate monitor (Vantage, Polar Electro). The sensor was attached directly to the skin using an elastic belt and measurements (bpm) were recorded on the device receiver computer which was held by the researcher. Heart rate was recorded every minute during the graded exercise test and at the termination of the test.

5.3.4.4 Anthropometric measures

All anthropometric measurements were conducted in accordance with the International Standards for Anthropometric Assessment (Stewart, Marfell-Jones, Olds, & de Ridder, 2011). Height was measured to the nearest 0.1 cm using a stadiometer (Seca Scales, Hamburg, Germany). Without shoes, participants stood with their heels together and their back against the scale. The head was facing forward with the chin level, and arms were relaxed. Two separate measurements were conducted and the mean value calculated.

Weight was measured to the nearest 0.1 kg using digital scales (Seca Scales, Hamburg, Germany). Two measurements were also conducted with light clothing and no shoes and the mean value calculated.

From the height and weight measurements, BMI was calculated using the following formula:

$$\text{BMI (kg/m}^2\text{)} = \text{weight (kg)}/[\text{height (m)}]^2$$

5.3.5 Management of data

Accelerometer data were recorded in real time whereas respiratory gas exchange measurements were recorded in relation to session duration, i.e.

recoding began at 0.00. To ensure the synchronisation of these measurements, activity start and stop times were recorded in both real time and session duration.

Missing values were intermittently present when the respiratory gas exchange data were initially downloaded in the breath by breath format using the standard threshold settings. It was hypothesised that these data points were being excluded by the BreezeSuit software (Medical Graphics, MN, USA) as they were outwith the standard threshold settings. To allow further investigation, data were additionally downloaded with no threshold settings applied. Therefore, respiratory gas exchange data were downloaded in two formats: 10-second time averaged with standard thresholds and 10-second time averaged with no thresholds.

Respiratory gas exchange data are presented in 10-second time averaged format as time averaging data reduces variability and random error (Robergs, Dwyer, & Astorino, 2010). After data were time averaged, standard and no thresholds were applied using BreezeSuite software (Medical Graphics, MN, USA). Time averaging data removed most of the missing data points that were present in the breath by breath format. However, one participant with intellectual disabilities had one minute of missing data during the treadmill activity protocol for standard threshold data. Data were imputed from the no threshold data, which had no missing measurements, as the measurements for standard and no thresholds were the same during the activity in which the missing data were present.

Prior to being exported into SPSS, data were initially downloaded into an Excel spreadsheet (Version 14.0; Microsoft, 2010) to obtain the measurements required for statistical analyses. Measurements of $\dot{V}O_2$ (mL/Kg/min) were manually extracted for each activity. Additionally, measurements of RER during the graded exercise test, and $\dot{V}CO_2$ (mL/min) during the measurement of REE were extracted for standard threshold data. As steady state measurements are more valid, only minutes 2 to 4 for each activity were included in the analysis (Compher, Frankenfield, Kim, & Roth-Yousey, 2006). The final minute of data were excluded from the analysis as some participants became fatigued and agitated toward the end of the 5-minute measurement period, e.g. not completing the activity as advised or fidgeting with the mask. Therefore, this data was not deemed fully representative of the calibration activities.

Accelerometer data were sampled at a rate of 30 Hz and post-processed using ActiLife 6 software and reduced to 10-second epochs of data. Due to the intermittent movements used within the free-living activities, a shorter epoch will more accurately capture this sporadic activity (Vanhelst et al., 2010). Data were downloaded into an Excel spreadsheet (Version 14.0; Microsoft, 2010) where count data for all activities and measures was extracted for the hip placement vertical axis and vector magnitude. Accelerometer data epochs were then time matched to the corresponding respiratory data epoch. Accelerometer and $\dot{V}O_2$ data were organised for: total activity, activity by activity, and individual participants. Heart rate data collected during the graded exercise test were downloaded from the device receiver in 60-second epochs, with heart rate (bpm) at the termination of the test, as shown on the device receiver, recorded by hand.

5.3.6 Statistical analysis

All statistical data were analysed using SPSS 21 IBM statistical package (SPSS IBM, New York, NY, USA).

Normality was assessed for all variables to ensure the use of an appropriate statistical test. Each variable was plotted using a histogram with normal distribution curves and a boxplot to produce a visual representation of the data distribution. Skewness and kurtosis were tested using z-scores, with < 1.96 representing normal distribution. Normality was additionally assessed using the Shapiro-Wilk test. For data that were not normally distributed, logarithmic and square root transformations were separately applied to the data and normality was retested. If transformations were not effective in producing normally distributed data, non-parametric tests were used.

Descriptive statistics (mean ± SD) were calculated for age, sex, height, weight, and BMI. Additionally, independent t-tests were used to compare any significant differences in these variables between intellectual disabilities and typically developing participants.

The seven research questions within this study were investigated using the following analysis:

- Research question six (recruitment): percentages relating to recruitment, with detailed notes taken on the reasons given by teachers and parents who declined the invitation to be involved in this study.

- Research question seven (feasibility of activity protocol): qualitative observations and activity completion rates; detailed notes were taken on the times participants completed each activity for and any observations relating to participant and parent factors that could have affected completion of the protocol.

- Research question eight (feasibility of measuring REE): attainment of a steady state was measured using a coefficient of variation, with a coefficient of < 10% for $\dot{V}O_2$ and $\dot{V}CO_2$ signifying steady state. Appropriateness of the measurement was assessed using qualitative observations recorded using detailed notes.

- Research question nine (feasibility of graded exercise test): Mean (SD) scores for test duration, self-selected speed, peak $\dot{V}O_2$, RER_{peak}, HR_{peak}, and age-predicted maximum HR were calculated. Differences in these variables between intellectual disabilities and typically developing participants were investigated using independent t-tests.

- Research question ten (feasibility of measuring respiratory gas exchange): qualitative observations and feedback from participants, recorded using detailed notes.

- Research question eleven (effect of ultima CPX thresholds): the effect of thresholds was investigated within intellectual disabilities and typically developing participants for total $\dot{V}O_2$ data (all activity data combined) and activity-by-activity data using the dependant t-test for normally distributed data and the Wilcoxon signed-rank test for data that were not normally distributed. Individual differences were investigated using percentage change.

- Research question twelve (relationship between $\dot{V}O_2$ and accelerometer counts): within group differences were investigated using Spearman's correlation coefficient. These correlation coefficients were converted to z-scores to investigate differences in the relationship between $\dot{V}O_2$ and accelerometer counts between intellectual disabilities and typically developing participants.

5.4 Results

This section will present the results relating to the feasibility of a laboratory-based calibration protocol in children with intellectual disabilities. Results will be presented in relation to the seven specific research questions being investigated within this study.

5.4.1 Participants

Five typically developing children (1 male; 4 female) aged between 11 and 14 years were initially recruited to ensure the suitability of the protocol and allow comparison. Subsequently, five children with mild to moderate intellectual disabilities (4 males; 1 female) aged between 9 and 11 years participated in this study. Therefore, as children with intellectual disabilities were recruited after typically developing children, and due to the low recruitment rate of children with intellectual disabilities, it was not possible to match these groups for age. Descriptive statistics for both groups of participants are presented in Table 5.3. There were significant differences in age ($t = -3.11$, $df = 8$, $p < .05$), height ($t = -2.93$, $df = 8$, $p < .05$), and weight ($t = -2.46$, $df = 8$, $p < .05$) between the two groups, and no significant difference in BMI ($t = -.79$, $df = 8$, $p > .05$). Four typically developing participants and four participants with intellectual disabilities were within the healthy BMI range for children, with one typically developing participant and one participant with intellectual disabilities classified as overweight.

Table 5.3. Descriptive statistics for all participants

Participant	Sex	Age (years)	Height (m)	Weight (kg)	BMI (km/m^2)
ID 1	M	11	1.47	34.6	16.01
ID 2	M	11	1.53	40.5	17.3
ID 3	M	11	1.48	48.6	22.19*
ID 4	F	9	1.43	30.4	14.87
ID 5	M	9	1.33	28.9	16.34
TD 1	F	12	1.61	40.8	15.74
TD 2	F	11	1.51	49.1	21.53*
TD 3	F	14	1.72	56.1	18.96
TD 4	F	12	1.63	44	16.56
TD 5	M	13	1.52	47.6	20.6
ID Mean (SD)		10.20 (1.10)	1.45 (.08)	36.60 (8.08)	17.34 (2.85)
TD Mean (SD)		12.40 (1.12)	1.60 (.09)	47.52 (5.78)	18.68 (2.50)

ID = participants with intellectual disabilities
TD = typically developing participants
* BMI classified as overweight relative to age

5.4.2 Research question 6

Is it feasible to recruit children with intellectual disabilities from additional support needs schools to participate in a calibration study?

Five additional support needs schools in Glasgow were approached to discuss being a point of contact to children eligible for the study. Of these schools, two agreed for a researcher (AM) to visit and speak with pupils. This stage of recruitment was conducted in the three weeks preceding the school summer holidays; of the schools which declined, one gave the busy schedule during this period in the school year as a factor, with the remaining two schools citing insufficient time to attain the required parent consent for a researcher to visit before the school holidays.

A researcher attended each of the participating schools on one occasion. During the visit, the study was explained to children who met the criteria of being aged between 8 and 14 years. In total, 78 information packs were handed out. Ten

(12.82%) initial consent forms and parental contact details were returned. Of the seven parents whom it was possible to make contact with, five participated (6%). One parent declined the invitation for their child to participate due to the need to travel to the laboratory, while the other parent cited insufficient time to arrange the laboratory sessions.

In summary, the recruitment strategy for this study was not feasible, with the time demands of organising two laboratory-based sessions a parental barrier for participation. The academic school year also affected participation as the additional demands on schools associated with the end term limited the number of schools involved in recruitment.

5.4.3 Research question 7

Are activities conducted in a calibration protocol designed for typically developing children feasible for children with intellectual disabilities?

Only one participant with intellectual disabilities was able to complete all activities for the required time. In comparison, all typically developing participants completed the protocol. Two participants with intellectual disabilities completed all treadmill-based activities. Three km/h at zero gradient was the only activity which all participants were able to perform for 5 minutes. The activities which were not performed were deemed to be too physically demanding. Table 5.4 shows the activities completed and not completed by participants with intellectual disabilities.

No participants with intellectual disabilities had previous experience of using a treadmill and, although ample practice time was given, balance difficulties were present at speeds 6 km/h and 8 km/h. However, considering the unfamiliarity of the treadmill, all participants with intellectual disabilities completed these activities enthusiastically, and all commented that the treadmill-based activities were the most enjoyable.

Three participants with intellectual disabilities completed all free-living activities. During the final activities, one participant reported they were too tired due to the demands of the session and did not complete the hula hoop, Xbox Kinect, and jumping jacks activities. All other activities not performed were due to participants opting out of activities they did not perceive as enjoyable. Feedback from participants was that the computer activities, i.e. those involving the Nintendo DS and Xbox Kinect, were the most enjoyable free-living activities.

The design of this study was to test the feasibility of a variety of activities and combine the most appropriate activities into a shorter protocol. Hence, this protocol contained a greater number of activities than a standard calibration protocol. This had an effect on the completion rate and data collection, as participants became fatigued during the latter stages of the session, in particular the second session. Additionally, the intensity of activities increased as the session progressed which resulted in participants not completing some moderate and vigorous activities with the intended vigour and intensity, which would have an effect on a full-scale calibration protocol. Therefore, the conclusions regarding the most feasible activities could be affected by participant fatigue.

An additional finding regarding the feasibility of activities was the views of parents regarding which activities they perceived their child to be capable of completing. For example, one parent suggested that her child did not participate in the 6 km/h or 8 km/h activities, which required running, as she had never seen him run and assumed he was not capable of doing so; however, this participant (ID2) completed the 6 km/h activity for 5 minutes and 8 km/h for 3.5 minutes. From observations, the views of parents did not affect the completion rates within this study, yet it did seem to affect the vigour with which participants completed activities, i.e. participants whose parents encouraged them to continue with or attempt difficult activities completed activities with more enthusiasm.

In summary, it is not feasible to generalise activities from a calibration protocol designed for typically developing children to children with intellectual disabilities. Consideration has to be given to the energy demands and order of activities to prevent this negatively impacting on the data collected, i.e. participants not reaching the desired intensity due to fatigue.

Table 5.4. Activity completion rates of each participant with intellectual disabilities

Activities	Participant				
	ID 1	ID 2	ID 3	ID 4	ID 5
REE				7 min 40 sec	
3 km/h at 0%					
6 km/h at 0%				2 min 13 sec	
5 km/h at 5%					2 min 20 sec
8 km/h at 0%		3 min 30 sec			
Sitting playing DS					
Watching DVD					
Drawing		3 min 50 sec			
Passing football					
Throw/catch					
Standing playing DS					
Step aerobics					
Hula hoop		1 min 50 sec			
Xbox					
Jumping jacks					
Graded exercise test					

Green: Participant completed activity for required 5 min
Orange: Participant attempted activity but did not complete for 5 min (actual time provided)
Red: Participant did not perform activity

5.4.4 Research question 8

Is it feasible to measure REE in children with intellectual disabilities?

Four participants completed this measure for the 15-minute period. One participant became very agitated and could only continue with the measurement for 7 minutes and 40 seconds, although all participants expressed feeling uncomfortable with the duration of the measurement.

Table 5.5. Coefficient of variation (%) for achievement of steady state for REE measurements

	ID 1		ID 2		ID 3		ID 4		ID 5	
	$\dot{V}O_2$	$\dot{V}CO_2$	$\dot{V}O_2$	$\dot{V}CO_2$	$\dot{V}O_2$	$\dot{V}CO_2$	$\dot{V}O_2$	$\dot{V}CO_2$	$\dot{V}O_2$	$\dot{V}CO_2$
REE (final 10 minutes)	35.97	32.40	27.71	32.34	16.92	15.62	-	-	16.90	17.03
REE (final 5 minutes)	26.12	32.95	21.87	23.10	16.97	14.68	21.41	22.51	8.23	8.75

$< 10\%$ $\dot{V}O_2$ and $\dot{V}CO_2$ signifies the attainment of a steady state
$\dot{V}O_2$ and $\dot{V}CO_2$ data are presented as mL/kg/min

To reduce perceived anxiety and focus attention, participants watched a DVD throughout the REE measurement. However, no participant achieved a steady state, which will reduce validity. Table 5.5 shows the coefficient of variation for the achievement of a steady state during the final 10 minutes and 5 minutes of this measurement. The order of the protocol has again to be considered as REE was the first respiratory gas exchange measure taken during session one, which could have increased anxiety and had an effect on the data recorded.

In summary, it was not feasible for children with intellectual disabilities to achieve steady state REE measures within the suggested measurement time of 5 to 10 minutes (Compher et al., 2006). Watching a DVD was not an effective strategy for reducing participant agitation and perceived anxiety. However, conducting REE measurements at the beginning of the protocol could have had an additional effect on anxiety.

5.4.5 Research question 9

Is a treadmill-based graded exercise test to measure $\dot{V}O_{2max}$ feasible in children with intellectual disabilities?

Four participants with intellectual disabilities performed the graded exercise test. Participant ID2 did not perform the test due to fatigue. In comparison, all typically developing participants performed the test, however, due to a system error, no respiratory gas exchange measurements were recorded for TD1. Individual test data is presented in Table 5.6.

Each test was terminated by the participant, by signalling they were too exhausted to continue. No participant met the primary criteria for the attainment of $\dot{V}O_{2max}$, i.e. a plateau in $\dot{V}O_2$. Therefore, results are presented as the peak scores attained during the test. One typically developing participant (TD3) met the secondary criteria for a maximal test, with a RER > 1.0 and a HR_{max} within 10 bpm of an age-adjusted estimate.

Table 5.6. Individual scores attained during graded exercise test

Participant	$\dot{V}O_{2peak}$ (mL/kg/min)	HR_{peak} (bpm)	Age-predicted HR_{max} (bpm)	RER_{peak}	Speed (km/h)	Time (min)
ID1	34.30	195	209	1.06	5.50	21.50
ID3	30.40	172	209	0.92	5.00	9.00
ID4	25.70	152	221	0.97	4.00	16.00
ID5	31.60	135	221	0.89	3.50	7.00
TD 1	-	177	208	-	6.00	20.50
TD 2	29.40	172	209	1.04	5.50	15.50
TD 3	43.80	199	206	1.21	6.00	14.00
TD 4	30.20	179	208	0.96	6.00	12.00
TD 5	38.90	193	207	0.94	5.50	14.00
ID Mean	30.50	163.50	215.00	0.96	4.50	13.38
(SD)	(3.59)	(25.88)	(6.93)	(.07)	(.91)	(6.65)
TD Mean	35.57	184.00	207.60	1.04	5.80*	15.20
(SD)	(6.97)	(11.45)	(1.14)	(.12)	(.27)	(3.21)

* TD participants significantly ($p < .05$) different from ID participants

Test duration was longer for typically developing participants (M = 15.20 minutes, SD = 3.21 minutes) than participants with intellectual disabilities (M = 13.38 minutes, SD = 6.65 minutes), although this was not significant ($t = -.55$, $df = 7$, $p > .05$). Typically developing participants also self-selected faster speeds (M = 5.80 km/h, SD = .27 km/h) than participants with intellectual disabilities (M = 4.50 km/h, SD =.91 km/h), which was significant ($t = -3.01$, $df = 7$, $p < .05$). There were no significant differences in peak scores between typically developing and intellectual disabilities participants for $\dot{V}O_2$ ($t = -1.30$, $df = 6$, $p > .05$), heart rate ($t = -1.61$, $df = 7$, $p > .05$), or RER ($t = -1.08$, $df = 6$, $p > .05$).

In summary, the protocol of a treadmill-based graded exercise test is feasible for children with intellectual disabilities, although it is not feasible for the attainment of $\dot{V}O_{2max}$. However, the unfeasible attainment of $\dot{V}O_{2max}$ scores was not limited to participants with intellectual disabilities, as typically developing participants continued with the test for longer, and completed it at significantly faster speeds, yet were also unable to attain maximal scores.

5.4.6 Research question 10

Is the use of respiratory gas exchange equipment feasible in children with intellectual disabilities?

Difficulties were encountered with participants wearing the respiratory gas exchange mask. The two primary reasons for this were participant anxiety and the weight of the mask when the bidirectional flow meter and sampling line were attached.

All participants expressed concern about wearing the mask, although only during the longer duration measure of REE did this effect one participant's ability to complete the activity. The level of stress and anxiety experienced due to the mask was high; three participants recorded RER scores greater than one, indicating hyperventilation, and one became very upset. Methods employed to reassure participants who were experiencing higher levels of observed anxiety were a researcher talking to them and a researcher also wearing a mask. However, reported anxiety caused by the mask reduced as the session progressed and participants became more familiar with the equipment.

During dynamic movements, the weight attached to the mask caused it to slip down, leaving the nose and mouth partially uncovered. All participants were asked to wear a nose clip to limit the amount of expired gas not measured or ambient air captured, but no participant agreed to the nose clip. To prevent the mask coming off completely or slipping off the nose, a researcher held the sampling line to reduce the weight the mask had to support. The preVent (Medical Graphics, MN, USA) mask used was the smallest size available and no alternative masks were suitable.

The mask was attached to the measurement sensor with a 2.10 metre sampling line. All activities were therefore completed within a marked area to prevent damage or injury caused by moving too far from the measurement sensor whilst wearing the mask and sampling line. This restricted movement had no effect on participants' ability to complete any activity.

In summary, the use of the Ultima CPX system (Medical Graphics, MN, USA) was not feasible due to the usability issues associated with the mask. However, if this equipment issue is addressed and ample practice time is provided to reduce anxiety, respiratory gas exchange equipment could be feasible for use in children with intellectual disabilities.

5.4.7 Research question 11

Does altering Ultima CPX breath by breath system threshold settings have an effect on $\dot{V}O_2$ in children with intellectual disabilities and typically developing children?

At intermittent points in the standard threshold data, when participants were known to be active and wearing the mask, the measurements recorded were inconsistent, e.g. breath by breath measurements being recorded 20 seconds apart. However, when no thresholds were applied to the same data, there were fewer points of missing data. It was hypothesised that data were being excluded due to the use of standard thresholds. The effect of threshold settings was therefore investigated. As this analysis aimed to investigate the effect of thresholds and not the effect of missing data, missing data points were excluded from this analysis.

Distribution of the data was investigated using the Shapiro-Wilk test, z-scores, and histograms with normal distribution curves. Data for both groups of participants were not normally distributed, therefore logarithmic and square root transformations were separately applied to the data. Both transformation methods reduced the kurtosis and skew of the data, however neither was effective in producing a normal distribution. The Shapiro-Wilk normality test showed all transformed data to still be significantly ($p < .001$) different from a normal distribution. Based on this, non-parametric tests were deemed to be most appropriate.

For participants with intellectual disabilities, $\dot{V}O_2$ was significantly higher when standard threshold settings (Mdn = 9.30 mL/kg/min) were applied compared to no thresholds (Mdn = 9.00 mL/kg/min), $z = -12.43$, $p < .001$, $r = -.27$. Similarly, for typically developing participants, $\dot{V}O_2$ was significantly higher with standard threshold (Mdn = 11.10 mL/kg/min) settings compared to no thresholds (Mdn = 11.00 mL/kg/min), $z = -4.29$, $p < .001$, $r = -.09$.

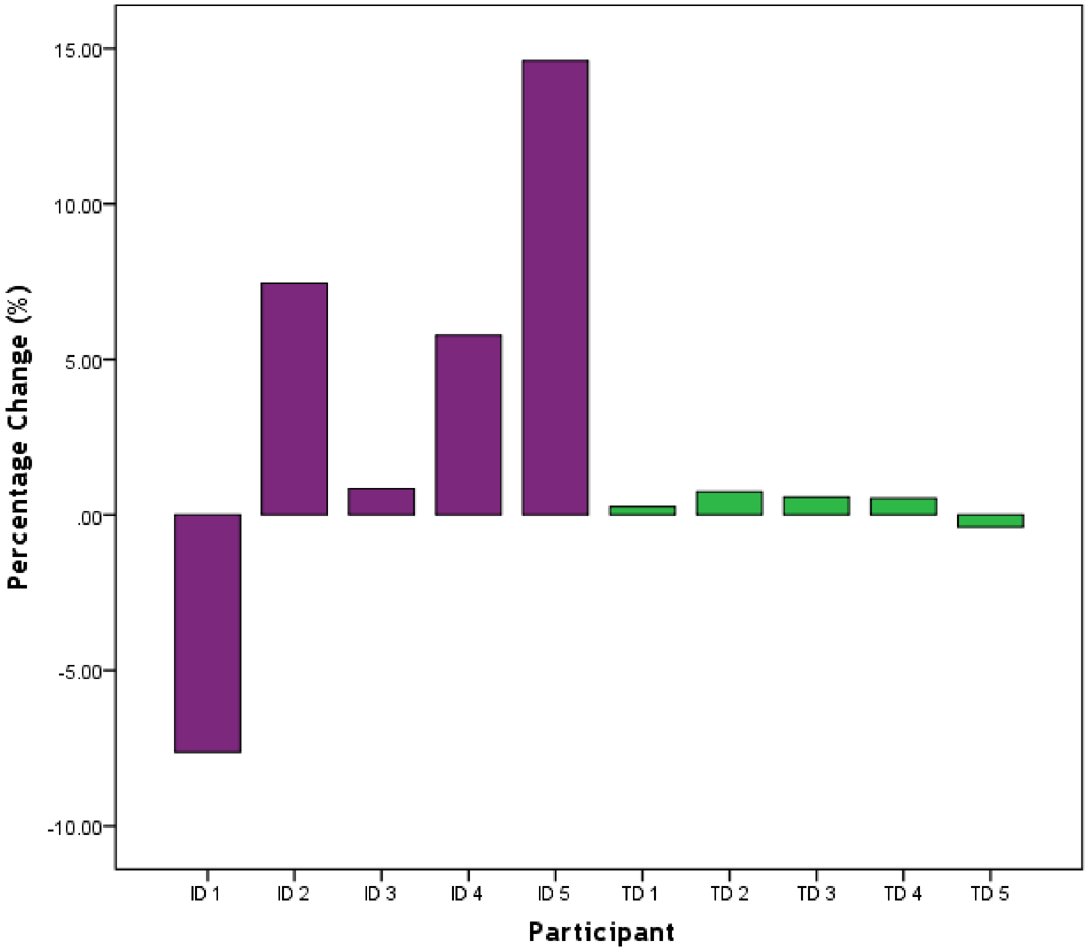


Figure 5.2. Difference between standard and no threshold settings represented as percentage change for each participant

When the effect of threshold settings was examined on an individual level, however, participants with intellectual disabilities had a greater variance than typically developing participants. Figure 5.2 illustrates the percentage change between standard and no thresholds, i.e. the mean difference between standard and no threshold measurements represented as a percentage. This shows individual differences relating to threshold settings, and also group differences that were not identified within the previous group analysis, which could be a result of the large data set. Percentage change for participants with intellectual disabilities (-7.63% to 14.61%) had a larger range compared to typically developing participants (-.39% to .74%). This therefore suggests that altering threshold settings not only has an effect on an individual level but that the effect is different between intellectual disabilities and typically developing participants.

Table 5.7. Test statistics for within group analysis comparing the effect of threshold settings on $\dot{V}O_2$ for each activity

Activity	Parametric Data				Non Parametric Data				
	Mean (SD)		t	df	Median		z	r	p
	ST	NT			ST	NT			
3km/h									
ID	11.47 (3.61)	10.99 (3.85)	3.57	89					.001**
TD	10.33 (2.81)	10.26 (2.78)	1.48	89					.14
6km/h									
ID	16.61 (6.17)	21.67 (8.59)	-4.70	61					.000***
TD					13.80	13.65	-.45	-.03	.66
5km/h 5%									
ID	21.69 (.89)	19.04 (.65)	4.09	77					.000***
TD					9.60	9.60	.00	.00	1.00
8km/h									
ID					27.70	36.40	-3.24	-.32	.001**
TD					19.75	19.75	.00	.00	1.00
Sitting DS									
ID	4.51 (1.75)	3.89 (1.86)	5.19	89					.000***
TD					6.60	6.20	-2.03	-.15	.04*

Table 5.7. Continued

Activity	Parametric Data				Non Parametric Data				
	Mean (SD)		t	df	Median		z	r	p
	ST	NT			ST	NT			
Watching DVD									
ID	4.66 (1.87)	3.95 (2.05)	5.52	89					.000***
TD					5.35	5.25	-1.60	-.27	.11
Drawing									
ID	4.93 (1.20)	3.41 (1.85)	6.56	64					.000***
TD	5.84 (1.64)	5.79 (1.64)	1.79	71					.08
Passing football									
ID	9.42 (4.43)	8.48 (4.49)	4.13	89					.000***
TD					8.70	8.05	-2.21	-.18	.03*
Throw/catch									
ID	10.62 (4.54)	9.32 (4.75)	4.49	71					.000***
TD	9.97 (5.32)	9.85 (5.33)	3.04	89					.003**
Standing DS									
ID	6.01 (2.14)	5.61 (2.25)	5.84	89					.000***
TD					10.15	10.15	-1.41	-.11	.16

Table 5.7. Continued

Activity	Parametric Data				Non Parametric Data				
	Mean (SD)		t	df	Median		z	r	p
	ST	NT			ST	NT			
Step aerobics									
ID	17.27 (6.19)	16.87 (6.49)	1.97	89					.05
TD	18.04 (8.87)	17.90 (9.03)	1.56	89					.12
Hula hoop									
ID	12.44 (5.31)	11.72 (5.49)	4.05	71					.000***
TD					16.35	16.35	-1.29	-.10	.20
X-box Kinect									
ID	9.23 (1.81)	8.91 (1.65)	3.64	71					.001**
TD	18.09 (5.96)	17.96 (6.07)	2.41	71					.02*
Jumping jacks									
ID	14.47 (7.14)	13.25 (8.06)	5.22	71					.000***
TD					8.30	13.35	-1.01	-.12	.31

The effect of thresholds on $\dot{V}O_2$ outcomes was additionally investigated in relation to each activity for both groups of participants. Normality of the data was tested, as described above, with non-parametric tests used for data that was not normally distributed. For participants with intellectual disabilities, threshold settings had a significant effect on results for all activities, with the exception of step aerobics (Table 5.7). For typically developing participants, however, there were only significant differences between threshold settings for four activities: sitting playing DS ($p < .05$), passing football ($p < .05$), throw/catch ($p < .01$), and XBox Kinect ($p < .05$). Figures 5.3 and 5.4 illustrate the mean $\dot{V}O_2$ scores for standard and no threshold settings for participants with intellectual disabilities and typically developing participants for each activity, respectively. These results also show that thresholds have a significant effect on $\dot{V}O_2$ for children with intellectual disabilities, with this effect greater than that found in typically developing participants.

In summary, altering threshold settings had a significant ($p < .001$) effect on $\dot{V}O_2$ measures for both groups of participants, with standard thresholds producing significantly higher measurements. This effect was not constant across all participants, with a greater variance seen for participants with intellectual disabilities. Additionally, the effect of thresholds was greater for participants with intellectual disabilities when data were investigated for each activity, which will have implications for the classification of activity and intensity for calibration. However, it was not possible within the design of this study to determine whether the results using the standard or no threshold settings were most valid.

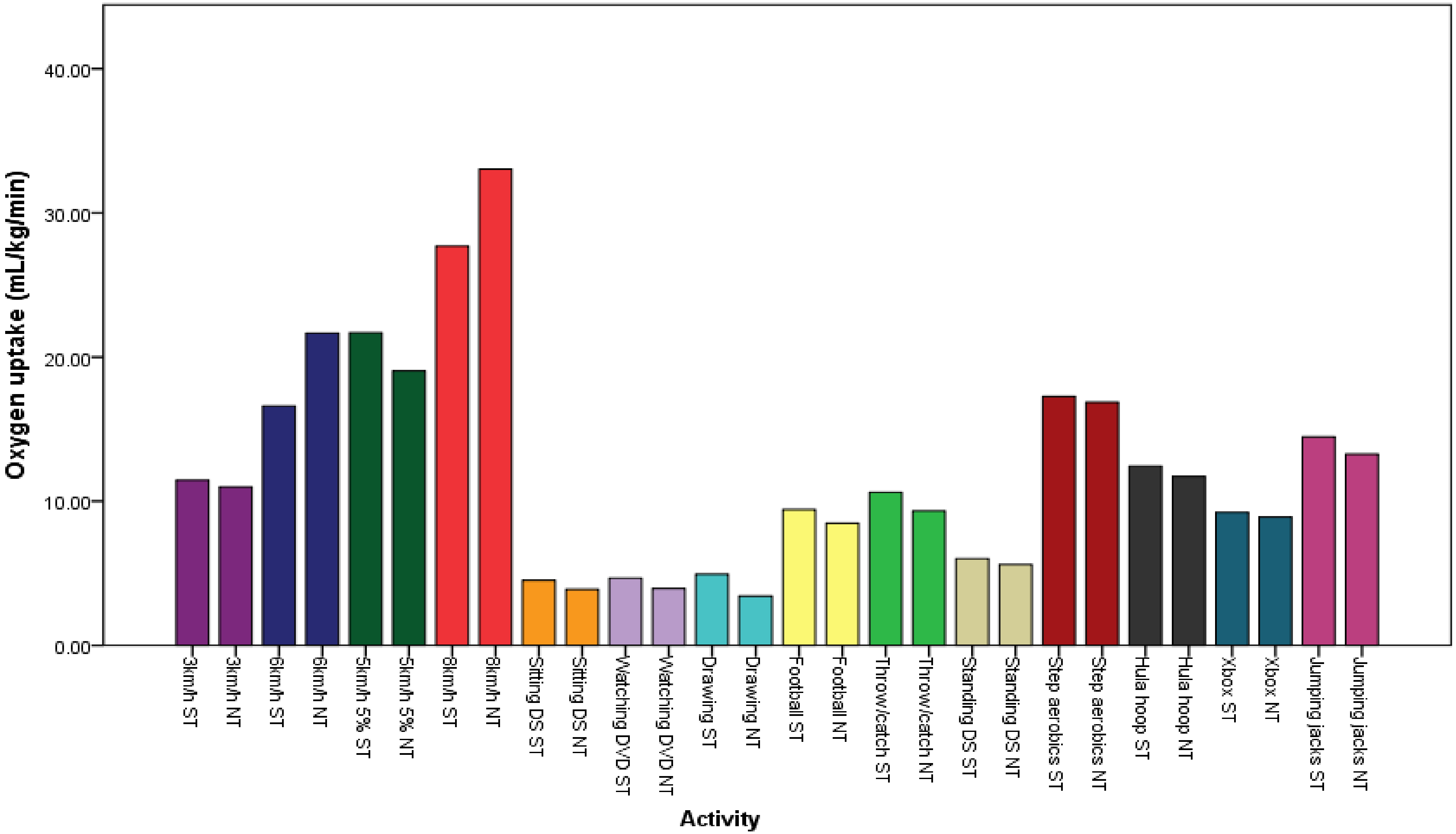


Figure 5.3. Mean $\dot{V}O_2$ of participants with intellectual disabilities for each activity
ST; standard thresholds
NT; no thresholds

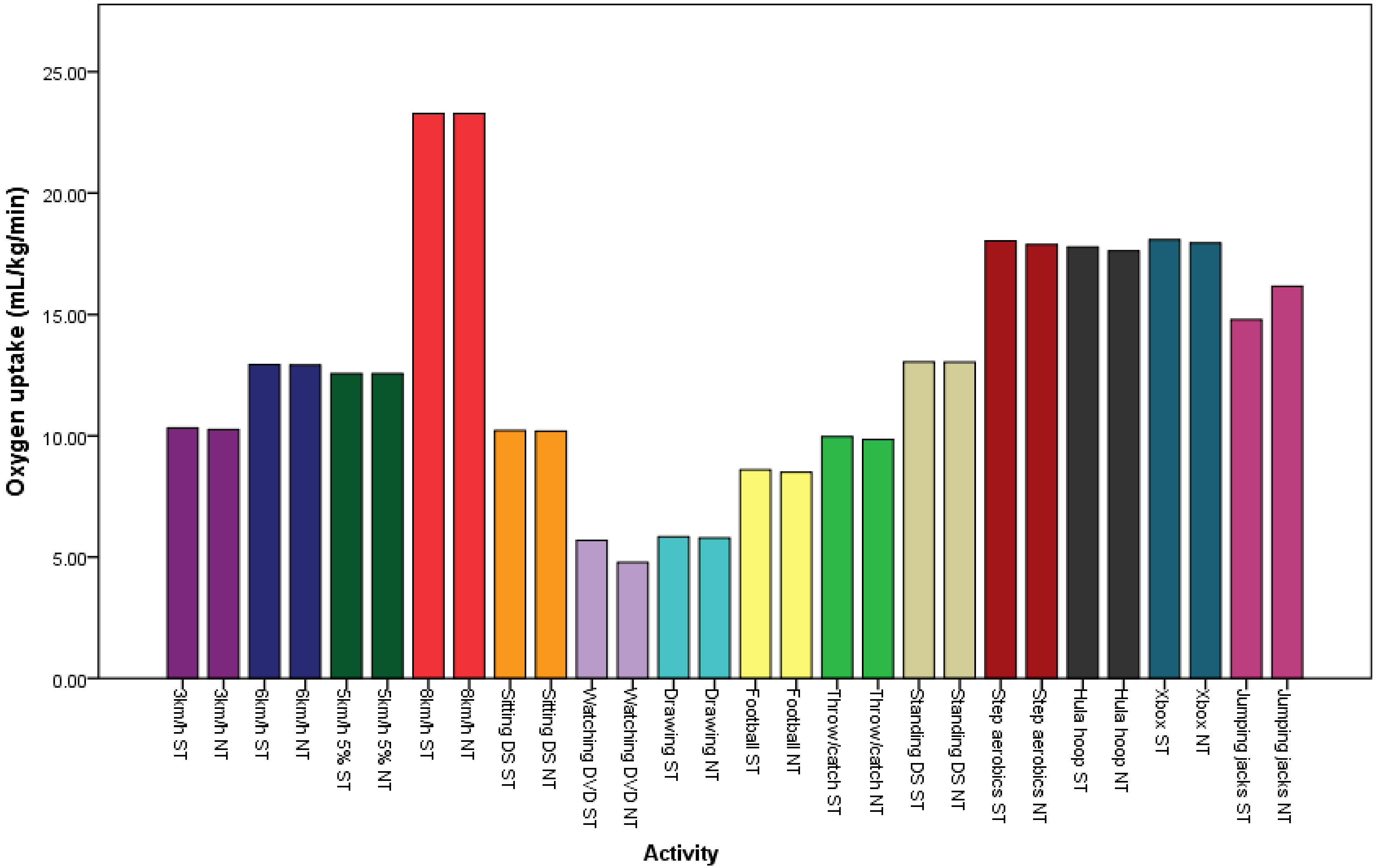


Figure 5.4. Mean $\dot{V}O_2$ of typically developing participants for each activity
ST; standard thresholds
NT; no thresholds

5.4.8 Research question 12

Is there a significant difference in the relationship between $\dot{V}O_2$ and accelerometer counts between children with intellectual disabilities and typically developing children?

Only counts for the vertical axis (hip placement) and counts for vector magnitude (hip placement) were used in this analysis as pre-existing cut points are most commonly calibrated using the hip placement for the vertical axis and vector magnitude. Also, standard threshold $\dot{V}O_2$ was used as no previous studies have investigated the use of no threshold settings. Normality tests showed all data not to be normally distributed, including after log and square root transformations. Therefore, the non-parametric Spearman's correlation coefficient was used. Additionally, all tests were one-tailed as it was hypothesised that there would be a positive relationship between counts and $\dot{V}O_2$.

Table 5.8. $\dot{V}O_2$ and count scores recorded for each activity
Data are presented as mean (SD)

Activity	$\dot{V}O_2$ (mL/kg/min)		Vertical axis counts (counts/10-sec)		VM counts (counts/10-sec)	
	ID	TD	ID	TD	ID	TD
Treadmill						
3 km/h	11.47 (3.61)	10.33 (2.81)	120.04 (58.78)	138.31 (50.46)	263.03 (79.94)	267.41 (100.69)
6 km/h	16.61 (6.17)	12.94 (6.03)	469.98 (183.16)	696.90 (191.98)	703.88 (259.00)	845.80 (219.27)
5 km/h at 5%	21.69 (7.82)	12.57 (7.21)	318.90 (138.93)	528.36 (156.62)	531.87 (174.98)	663.40 (182.30)
8 km/h	27.69 (9.68)	23.28 (8.68)	768.02 (170.53)	1126.93 (181.48)	978.44 (161.07)	1278.26 (190.13)
Free-living						
Sitting DS	4.51 (1.75)	10.22 (7.33)	1.66 (8.89)	.16 (1.48)	4.36 (21.82)	.22 (1.60)
DVD	4.66 (1.87)	5.69 (2.20)	2.27 (7.69)	1.89 (4.87)	8.74 (20.77)	1.89 (4.87)
Drawing	4.93 (1.20)	5.84 (1.64)	7.59 (32.60)	.07 (.59)	25.24 (61.14)	.61 (3.23)

Table 5.8. Continued

Activity	$\dot{V}O_2$ (mL/kg/min)		Vertical axis counts (counts/10-sec)		VM counts (counts/10-sec)	
	ID	TD	ID	TD	ID	TD
Football	9.42 (4.43)	8.60 (3.14)	22.36 (58.90)	62.88 (66.07)	208.66 (138.82)	337.24 (163.68)
Throw/catch	10.62 (4.54)	9.97 (5.32)	32.71 (49.66)	11.79 (27.17)	151.70 (124.90)	103.42 (101.04)
Standing DS	6.01 (2.14)	13.05 (8.05)	1.92 (14.74)	.12 (.85)	18.61 (81.03)	3.22 (16.34)
Step aerobics	17.27 (6.19)	18.04 (8.87)	387.63 (323.58)	432.12 (96.50)	591.32 (333.91)	696.43 (154.07)
Hula hoop	12.44 (5.31)	17.79 (11.40)	135.65 (166.00)	17.63 (11.38)	386.45 (306.55)	819.54 (339.83)
Xbox	9.23 (1.81)	18.09 (5.96)	27.40 (43.51)	18.58 (26.75)	147.11 (125.08)	132.56 (106.73)
Jumping jacks	14.47 (7.14)	14.79 (10.19)	870.03 (885.65)	2189.89 (1066.63)	992.69 (937.67)	2231.66 (1035.49)

VM = vector magnitude
ID = participants with intellectual disabilities
TD = typically developing participants

Table 5.8 shows the mean $\dot{V}O_2$ and counts recorded for each activity. There was a significant difference in the relationship between $\dot{V}O_2$ and the vertical axis counts ($z = 13.21$, $p < .0001$) and $\dot{V}O_2$ and vector magnitude counts ($z = 14.23$, $p < .0001$) between participants with intellectual disabilities and typically developing participants. For participants with intellectual disabilities, $\dot{V}O_2$ was significantly correlated with the vertical axis counts ($r_s = .70$, $p < .001$) and vector magnitude counts ($r_s = .73$, $p < .001$), with no significant differences between these correlation coefficients ($z = -1.79$, $p > .05$). Similarly, $\dot{V}O_2$ was significantly correlated with both vertical axis counts ($r_s = .29$, $p < .001$) and vector magnitude counts ($r_s = .31$, $p < .001$) for typically developing participants, with no significant differences between correlation coefficients ($z = .59$, $p > .05$). Figures 5.5–5.8 illustrate the relationships between $\dot{V}O_2$ and counts for sedentary and moderate to vigorous intensity activities.

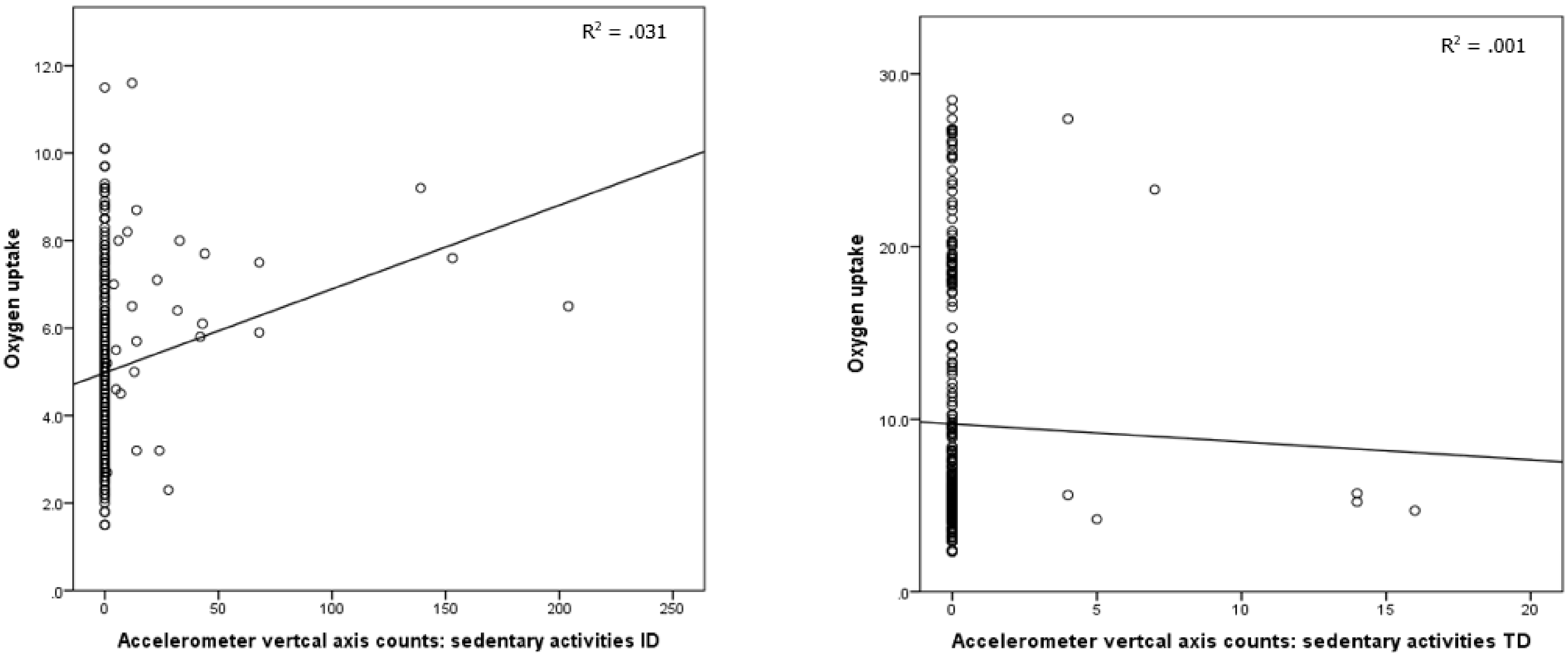


Figure 5.5. Relationship between $\dot{V}O_2$ and vertical axis accelerometer counts for sedentary activities
ID; participants with intellectual disabilities: TD; typically developing participants

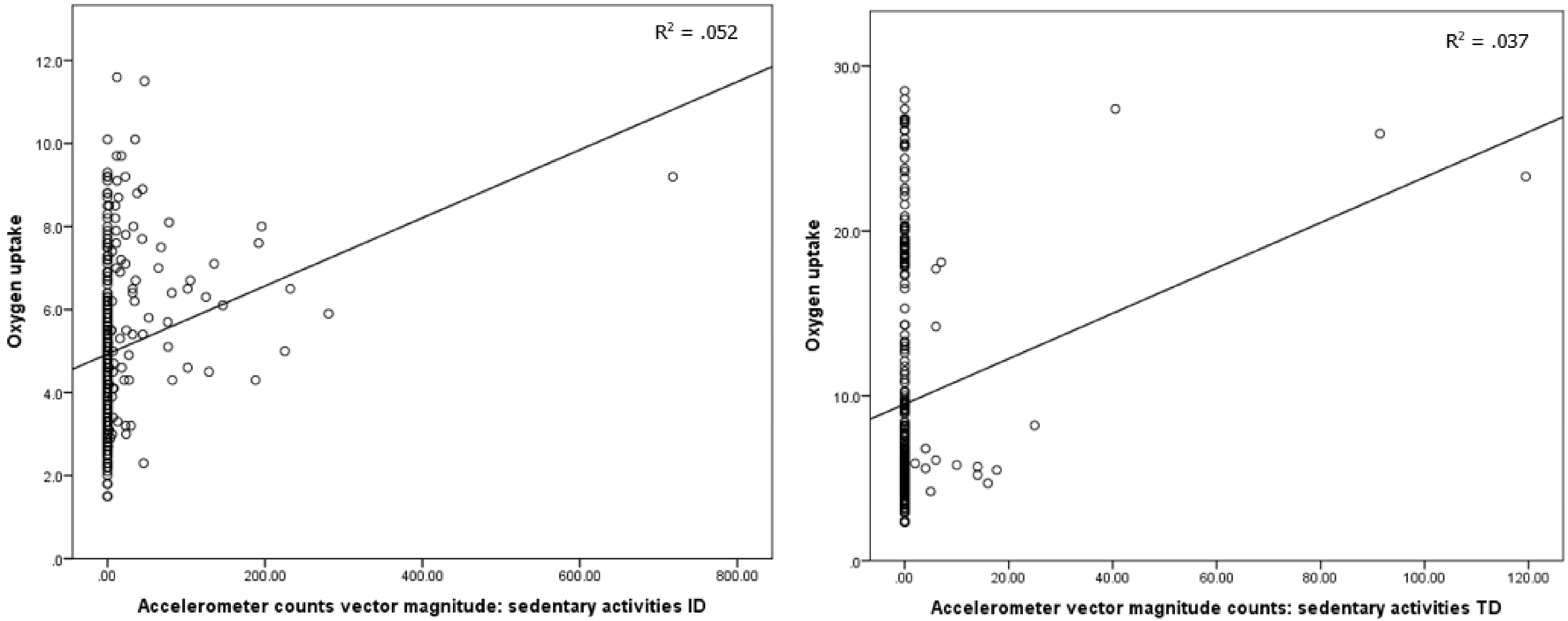


Figure 5.6. Relationship between $\dot{V}O_2$ and vector magnitude accelerometer counts for sedentary activities
ID; participants with intellectual disabilities: TD; typically developing participants

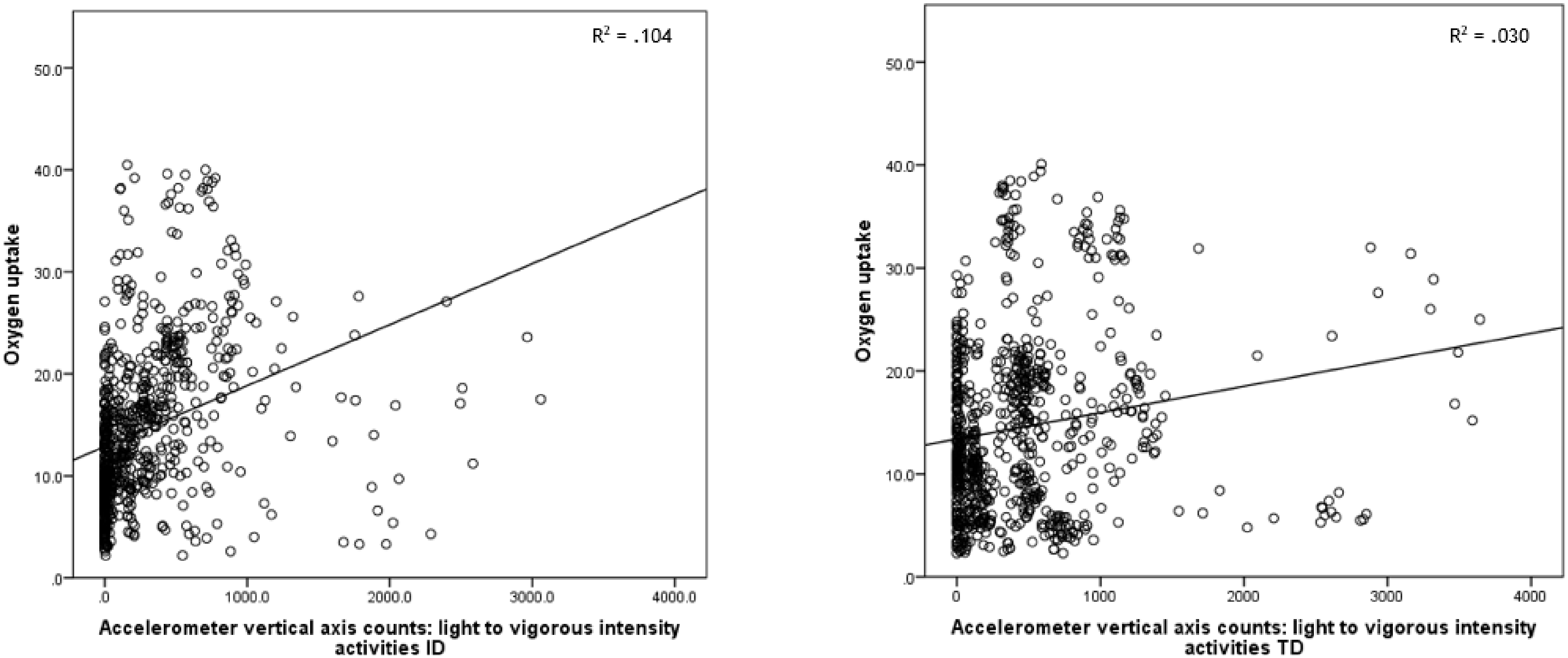


Figure 5.7. Relationship between $\dot{V}O_2$ and vertical axis accelerometer counts for light, moderate, and vigorous intensity activities
ID; participants with intellectual disabilities: TD; typically developing participants

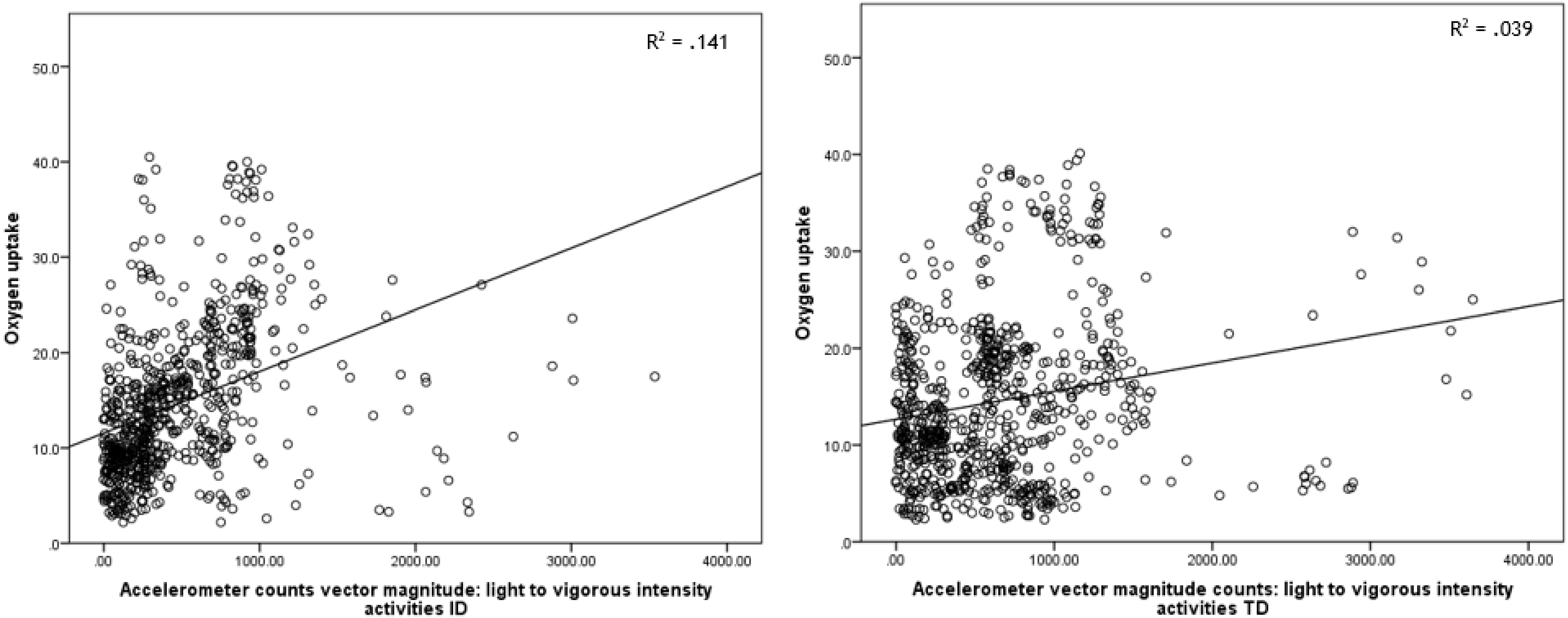


Figure 5.8. Relationship between $\dot{V}O_2$ and vector magnitude accelerometer counts for light, moderate, and vigorous intensity activities
ID; participants with intellectual disabilities: TD; typically developing participants

This additionally highlights high $\dot{V}O_2$ scores which were recorded whilst the participants were not moving, i.e. zero counts. Furthermore, the relationship between counts and $\dot{V}O_2$ is more linear for physical activity in comparison to sedentary behaviours for children with intellectual disabilities for the vertical axis and vector magnitude. In typically developing children, a more linear relationship for physical activity was only found for the vertical axis.

In summary, there was a significant difference in the relationship between $\dot{V}O_2$ and accelerometer counts between typically developing participants and participants with intellectual disabilities for vertical axis and vector magnitude counts. For both the vertical axis and vector magnitude counts, the correlation with $\dot{V}O_2$ was strongest in participants with intellectual disabilities. Vector magnitude also had a stronger correlation with $\dot{V}O_2$, compared to the vertical axis, for both groups. High $\dot{V}O_2$ scores were also present for both intellectual disabilities participants and typically developing participants when zero counts were recorded, which could be partially caused by the anxiety described previously.

2.4 Discussion

5.4.1 Section overview

This section will discuss the findings from this study in the context of previous research. Findings will be discussed in relation to the following areas: recruitment (research question six), activities (research questions seven), REE (research question eight), graded exercise test (research question nine) breath by breath respiratory gas exchange (research questions ten and eleven), and accelerometry and $\dot{V}O_2$ (research question twelve). Additionally, strengths and limitations of this study will be discussed, with suggestions for future research.

5.4.2 Recruitment (research question 6)

The low participation rate within this study suggests that recruitment from only additional support needs schools is not feasible. The initial aim was to recruit 10 participants to both the intellectual disabilities and typically developing group, as 10 is the minimum suggested sample size for a calibration study (Freedson et al., 2005). Although five participants is not a suitable number for a calibration study, it is comparable to participation rates in previous research investigating the feasibility of calibration protocols. Oortwijn et al. (2009) recruited five typically developing children for a study which aimed to test the feasibility of an accelerometer calibration protocol within a whole room calorimeter. However, Oortwijn et al. (2009) report a response rate of 83%, which is notably higher than the 12.82% initial response rate within the present study.

Small sample sizes and an over-representation of boys are common limitations in health-related research involving children with intellectual disabilities (Maïano et al., 2014). The over-representation of boys within this study could be partially attributed to the higher prevalence of intellectual disabilities in boys compared to girls (Croen et al., 2001). Furthermore, as boys generally participate in more physical activity than girls, the activity-focussed protocol may have been of less interest to girls, which could have further limited their recruitment. Although recruiting large sample sizes of children with intellectual disabilities is difficult, it is important that protocols and measures are pilot tested on as many participants that can be recruited (Rikli, 1997).

This suggests that for a laboratory-based calibration study, a greater number of children with intellectual disabilities need to be approached to generate the required sample size, compared to a study involving typically developing children. To achieve this, other points of contact, such as sports clubs for children with intellectual disabilities, could be used for recruitment.

Feedback from parents who opted out of participation suggests that the two session laboratory-based protocol was a barrier due to time and travel requirements. However, this study included a longer protocol with a greater number of activities than a standard calibration study, with the aim that the most feasible activities would be combined into a single session for a full-scale study. Therefore, it is possible that this barrier would lessen for a full-scale calibration study that was conducted during a single session.

School-based recruitment strategies have previously been described as effective for recruiting typically developing children to exercise-related studies (Rowland, 1994). The use of incentives, such as the vouchers and reimbursement of travel expenses used within this study, have also been effective in the recruitment of typically developing children (Rowland, 1994). However, there is limited research regarding the recruitment of children with intellectual disabilities. The lack of detail presented within previous studies on the specifics of the recruitment strategy employed and discussion on its effectiveness is limiting comparison between studies. Furthermore, this is also hindering the development of a comprehensive method of recruiting children with intellectual disabilities to health-related research. For adults with intellectual disabilities, recruitment strategies involving direct contact with participants has been shown to be most effective in comparison with telephone recruitment or the use of a third-party contact within a service organisation (Lennox et al., 2005). Furthermore, recruitment of adults with intellectual disabilities is lower for studies involving more invasive measures and physical tests (Cleaver, Ouellette-Kuntz, & Sakar, 2010). Therefore, the nature of the testing procedures within this study may have contributed to the low recruitment rate.

In summary, the recruitment strategy employed was not effective within the timeframe of this study. For future studies, the low response rate from parents needs to be accounted for, although the inclusion of a greater number of schools and service organisations could provide the required number of participants for a full-scale calibration study. However, as invasive measures and physical tests are barriers for recruitment in adults with intellectual disabilities, further investigation is needed to determine whether this low recruitment rate is a direct result of an ineffective recruitment strategy, or whether this type of study was one that children with intellectual disabilities do not wish to participate in.

5.4.3 Activities (research question 7)

This study suggests that the activities used within a calibration protocol in typically developing children cannot be fully replicated in children with intellectual disabilities. Although the treadmill activities were based on previous use in typically developing children of a similar age, the speeds were not appropriate for this sample, as the physical demands were too high to allow participants to attempt or complete all activities. Some previous studies, however, have aimed to ensure the suitability of activities by proposing speeds per age group or within a range of speeds. Puyau et al. (2002) included vigorous activities that were age-specific; furthermore, Puyau, Adolph, Vohra, Zakeri, and Butte (2004) proposed moderate and vigorous activities within speeds of 3.5 to 4 mph and 4.5 to 7 mph, respectively.

When compared to these previous studies, the moderate (5 km/h at 5% and 6 km/h) and vigorous (8 km/h) speeds in the present study are within the lower ranges used within Puyau et al. (2002) and nearest to the 8 to 10 years vigorous speed within Puyau et al. (2004). With previous studies identifying age as having an effect on the suitability of activities, the high completion rate for typically developing participants could be due to age, as they were significantly older than participants with intellectual disabilities. However, three participants with intellectual disabilities within this study were still not able to complete the treadmill speeds which were deemed age-appropriate in these previous studies, suggesting slower speeds should be used within this population. Employing a range of speeds or age-specific speeds for children with intellectual disabilities could therefore increase the rate of completion.

The completion rate was high for free-living activities with three participants with intellectual disabilities completing the protocol, suggesting that generalising free-living activities from a typically developing protocol is feasible. The participants who did not complete all activities opted out of drawing and throw/catch due to a dislike for the activities. No previous studies, however, have discussed participants opting out of activities. There were no activities which participants were not capable of completing. This could be partially due to the intensity of activities not being fixed, as in the treadmill-based activities. Therefore, participants could complete the activities at in intensity that was comfortable for them and could intermittently stop when fatigued. Although this could have a positive effect on completion rates, it is important to ensure that activities are structured so participants reach the desired intensity for the purposes of calibration. The use of unstructured activities in children where there is no fixed intensity, such as during free-play, negatively affects the validation of accelerometry due to the limited time that children spend in moderate to vigorous intensity activity (Kahan, Nicaise, & Reuban, 2014).

It is important to note, however, that the aim of this protocol was to test the feasibility of a variety of activities and combine the most appropriate activities into a shorter protocol. Therefore, this protocol contained a greater number of activities than a standard calibration protocol. This was deemed the most effective way to identify feasible and enjoyable activities, which would be important for calibration and recruitment in a future study. However, this had an effect on completion rates and data collection, as participants became fatigued during the latter stages of the session, in particular the second session. As previously discussed in Section 5.2.1.2, an alternative method to design a protocol would be to choose activities based on previous research which has investigated activity preferences in the population of interest; however, this research is very limited in children with intellectual disabilities. Therefore, increasing our understanding of the types of activities that children with intellectual disabilities enjoy participating in is an important area for future research.

An interesting finding in the present study was the influence of parents and their views in relation to their child's ability to complete activities. Although this did not appear to have an effect of the completion of activities, i.e. no participant opted out of an activity due to their parent's views, it raises important questions regarding the effect that parental perceived competence could have on the overall physical activity levels of children with intellectual disabilities. Furthermore, although not within the design of this study, the influence of parental perceived competence could have had an additional effect on recruitment, as some parents may have considered their child not to be capable of completing the protocol.

No previous research seems to have addressed the effect of parental influence or perceived competence on physical activity in children with intellectual disabilities. However, in typically developing children parental encouragement is a significant ($p < .01$) predictor of physical activity and perceived competence in children (Welk, Wood, & Morss, 2003). More specifically, mothers' perceived competence is a predictor of their child's physical activity, whereas a father's perception of perceived competence is not, which is relevant to the present study as all children attended the session with their mother (Bois, Sarrazin, Brustad, Trouilloud, & Cury, 2005). Although, in practice, it is not feasible for parents not to be present during the session as any negative competence-related observations did not outweigh the observed comfort-related benefits that parents being present had on participants. However, further investigation into parental influence and perceived competence, based on theoretical frameworks, could increase our understanding of the engagement of children with intellectual disabilities in physical activity and recruitment for research studies.

In summary, the design of this protocol had an effect on data collection as participants became fatigued during the latter stages of the sessions. Additionally, the intensity of activities increased as the sessions progressed which resulted in participants not completing some moderate and vigorous activities with the intended vigour and intensity, which would have an effect on a full-scale calibration protocol. Therefore, the conclusions regarding the appropriateness of activities could be affected by participant fatigue. That said, the findings suggest that treadmill activities cannot be generalised for children with intellectual disabilities, and speeds per age group or within a range of speeds should be considered. In contrast, the high completion rates indicate that free-living activities can be generalised from a typically developing protocol to a calibration study involving children with intellectual disabilities.

5.4.4 Resting energy expenditure (research question 8)

Observed participant anxiety caused by the respiratory gas exchange equipment had an effect on REE measurements. Measuring REE is therefore not deemed feasible in children with intellectual disabilities within this calibration protocol. Anxiety has been previously noted as a limitation of using respiratory gas exchange measurements in children (Corder et al., 2008). Therefore, the attainment of a steady state is important to optimise results and is particularly important for resting metabolic measures (McClave et al., 2003). If a steady state is not achieved, repeated measurements are needed to ensure validity (Compher et al., 2006). Furthermore, due to the high within-participant variability in populations with disabilities, the attainment of reliable baseline scores may not be feasible, with results having to instead be averaged over multiple measurements (Rikli, 1997). Based on this, the resting measurements from this study may not be a valid representation of REE.

As REE was the first physiological measurement conducted within this protocol using the respiratory gas exchange equipment, this may have further increased anxiety. Conducting this measurement at a later stage in the protocol may limit this effect. If a steady state is achieved, however, only 5 to 10 minutes of measurement is required, with the initial 5 minutes discarded (Compher et al., 2006). As the protocol for this measure was similar to that used for sedentary activities (watching DVD), it could therefore be feasible to estimate REE from the measurements obtained during continuous sedentary activities.

Resting energy expenditure is important for the calibration process, as METs should be presented for activities, which requires a measurement of energy expenditure at rest (Freedson et al., 2005). Resting rate is constant among adults, with 3.5 mL/kg/min the standard measurement used for the calculation of activity METs in this population. However, REE in children aged 5 years is approximately 6 mL/kg/min which declines to 3.5 mL/kg/min at 18 years (Schofield, 1985). This decline with age therefore requires REE to be individually calculated for children as the use of adult MET thresholds for calibration would introduce systematic error (Freedson et al., 2005). However, REE can be approximated through age-specific estimates, therefore, a direct measurement is not essential for calculating MET thresholds.

In summary, the high coefficients of variation for $\dot{V}O_2$ and $\dot{V}CO_2$ show that participants were hyperventilating during the REE measurement due to anxiety. Therefore, based on these findings, measuring REE is not feasible. However, when considering the design of the protocol, conducting REE as the first measurement using the respiratory gas exchange equipment could have contributed to the high levels of anxiety observed. Although additional preparation time could reduce anxiety, the use of age-specific estimates is deemed most appropriate for measuring REE in children with intellectual disabilities.

5.4.5 Graded exercise test (research question 9)

No participants with intellectual disabilities reached $\dot{V}O_{2max}$, which could be due to a number of factors. Firstly, test duration ranged from 7 to 21.5 minutes, with one participant reaching the maximum treadmill gradient. It is suggested that an exercise test should be within a duration of 8 to 12 minutes to prevent premature termination of the test due to localised muscle fatigue, rather than the attainment $\dot{V}O_{2max}$ (Balady et al., 2010). Specific to a treadmill-based graded exercise test, the protocol of increasing gradient can cause calf muscle fatigue and lower back discomfort, which limits the participant's ability to continue with the test (McArdle, 1973). Therefore, due to the longer test durations in this study, it is reasonable to assume that participants experienced fatigue and discomfort, which could have attribute to premature termination of the test prior to achieving $\dot{V}O_{2max}$.

From observation, as the gradient increased, participants with intellectual disabilities became unstable, which may have been a contributing factor to the termination of the test before $\dot{V}O_{2max}$. Although encouraged to walk as normally as possible, all participants with intellectual disabilities used the treadmill handrail for support as the test progressed. However, the additional stability provided by the handrail may have increased the test duration as this would reduce the work load and affect the relationship between $\dot{V}O_2$ and work rate (Balady et al., 2010). For a graded exercise test in children with intellectual disabilities, Fernhall et al. (2000) used a maximum gradient of 12%; if participants reached this gradient, the speed was then increased by 0.5 km/h every minute until termination of the test.

Although this protocol could lessen the effect of localised muscle fatigue caused by the gradient, it could be limited by the increasing speed, as some participants within this study were not physically able to complete moderate to vigorous treadmill activities. From the range of chosen speeds (3.5 to 5.5 km/h), all participants chose a slow walking speed, which results in a lower work rate and could therefore additionally limit the likelihood of $\dot{V}O_{2max}$ attainment before muscle fatigue. Alternatively, participants could have been encouraged to select a slightly faster speed to conduct the test. This would increase the work load and theoretically result in a shorter test, therefore lower treadmill gradients would be used. However, as feasibility issues were identified with the use of faster treadmill speeds, this may not be a feasible solution for all participants.

The mean $\dot{V}O_{2peak}$ score for children with intellectual disabilities attained within this study was 30.50 mL/kg/min. This mean score is lower in comparison to $\dot{V}O_{2peak}$ scores recorded in previous research involving children with intellectual disabilities during treadmill-based tests. Fernhall et al. (2000) and Baynard et al. (2008) note the attainment of mean $\dot{V}O_{2peak}$ scores of 39.80 mL/kg/min and 39.40 mL/kg/min for children aged 9 to 15 years and 13.7 years, respectively. However, the attainment of high $\dot{V}O_{2peak}$ or $\dot{V}O_{2max}$ scores can be difficult in children who have no prior experience of strenuous exercise and the physical effects and discomfort associated with the protocol of an exercise test (Katch et al. 2011). Therefore, as none of the sample in this study had prior experience of a treadmill, an alternative test could limit this effect.

An alternative method of measuring cardiorespiratory fitness is through a submaximal test from which $\dot{V}O_{2max}$ can be estimated. Multiple field-based tests have been developed which are less dependent on expensive measures and complex protocols and therefore provide more feasible methods for data collection. These tests are primarily based on the relationship between heart rate and work rate; however, generalising these estimates to children with intellectual disabilities requires caution as lower maximal heart rates have been reported in this population. This is primarily apparent in children with Down syndrome, therefore the use of unadjusted equations and estimates developed in typically developing children will underestimate cardiorespiratory fitness (Fernhall et al., 2001). However, submaximal tests, including the one mile walk test (Teo-Koh & McCubbin, 1999), 600 yard walk/run test, 20 metre and modified 16 metre shuttle run test (Fernhall et al., 2000), have shown reliability and concurrent validity for the prediction of $\dot{V}O_{2max}$ in children with intellectual disabilities. Therefore, these tests could be considered within a calibration protocol as an alternative to a maximal test.

It is also important to consider the effect of the overall calibration protocol on the effectiveness of the graded exercise test. Mean time for session two was 101 ± 19.85 minutes, which concluded with the graded exercise test. The preceding activities could have reduced the physiological or mental capabilities of participants to complete the test to their functional limit ($\dot{V}O_{2max}$), e.g. due to muscle fatigue or concentration/tiredness. Therefore, it may be more appropriate to conduct a submaximal test to estimate $\dot{V}O_{2max}$, or conduct the test at another stage in the protocol. Alternatively, as done by Evenson et al. (2008), $\dot{V}O_{2max}$ could be estimated from measurements obtained during treadmill-based activities.

In summary, the treadmill-based graded exercise test protocol employed within this study was not feasible for the attainment of $\dot{V}O_{2max}$, which could be a result of participant fatigue and instability on the treadmill during the latter part of the test. With the physical demands of a calibration protocol already high for participants, researchers in future studies should give consideration to submaximal tests or, where feasible, estimate $\dot{V}O_{2max}$ from treadmill-based activities.

5.5.6 Breath by breath respiratory gas exchange (research questions 10 and 11)

This study highlighted usability and measurement issues regarding the Medical Graphics breath by breath equipment. As previously discussed, wearing the mask caused anxiety for participants. From observations, this caused participant breathing rates to increase and become shallow, with RER data showing participants were hyperventilating. The intermittent increases in measurements caused by anxiety will introduce random error and negatively affect validity and reliability. These higher measurements may also affect the attainment of a steady state. It is important that the periods of measurement used for calibration are steady state, as this confirms that all that energy demands for that activity are being provided by the aerobic energy system, i.e. the $\dot{V}O_2$ measurements used for calibration represent the full energy demands of the activity.

From a practical perspective, there were difficulties identified with the suitability of the preVent mask (Medical Graphics, MN, USA) and equipment. The size of the mask, which was the smallest available size, was too big for all participants, which resulted in the mask moving during activities and occasionally uncovering the nose and mouth. As no participant agreed to wear a nose clip, the method employed to prevent this was a researcher holding the sampling line to take the weight off the mask. This was effective in keeping the mask in an appropriate position; however, this increased researcher burden may not be feasible for a full-scale calibration study. As no smaller mask size is available, the alternative equipment is a mouthpiece; however, as a nose clip has to be worn with the mouthpiece, this may also not be feasible. No previous studies were identified that discussed similar equipment issues, however, wearing the mask has previously been discussed as a cause of anxiety in children (Corder et al., 2008).

Additional measurement issues were also identified with the use of the Ultima CPX breath by breath system (Medical Graphics, MN, USA), specifically in relation to the use of thresholds. The application of no thresholds significantly lowered mean $\dot{V}O_2$, indicating that measurements are being recorded below the standard threshold settings. However, as illustrated in Figure 5.2, this effect was not consistent between participants, with a greater variance seen for participants with intellectual disabilities. Furthermore, the decrease in mean $\dot{V}O_2$ with no thresholds varied on an individual level, with two participants (ID1 and TD5) having higher mean $\dot{V}O_2$ with no thresholds. Although not a parameter of this study, individual disability type could have had an effect as participant ID5, who showed the greatest difference between thresholds (14.61%), had Down syndrome, which can affect aerobic capacity (Baynard et al., 2008; Mendonca, Pereira, & Fernhall, 2010).

The significant differences in $\dot{V}O_2$ between standard and no threshold settings will have significant implications for calibration. If the threshold setting used is not a valid representation of $\dot{V}O_2$, this will cause systematic error which will affect the validity of calibration. As $\dot{V}O_2$ is the criterion measure, it is essential that this measurement is accurate. However, within the design of the present study, it was not possible to distinguish which threshold setting is most valid. Therefore, until this is investigated further, the use of this breath by breath system is not feasible as a criterion measure.

Threshold settings were investigated within this study due to missing data points that were identified during data processing. A literature search identified no previous studies that discussed the effect of thresholds in relation to breath by breath respiratory gas exchange. However, the lack of data processing guidelines for this type of measurement has been discussed. Robergs et al. (2010) discuss that there is currently no universally accepted method for processing breath by breath $\dot{V}O_2$ data and that the resulting lack of consistency is impacting on the validity of processing and interpreting data.

To overcome these issues and improve the accuracy of measurements, researchers have reported electing to instead use the traditional Douglas bag method, as it allows a greater level of control over data collection and processing, and is less prone to error (Bassett et al., 2012; Macfarlane, 2001). However, as the use of the Douglas bag method still requires a mask or mouthpiece and nose clip, it will have the same usability limitations previously discussed. Furthermore, an advantage of the use of a breath by breath system over the time-averaged Douglas bag method is that is more accurately captures the intermittent and sporadic movements conducted by children, which will enable more precise calibration. Considering the limitations with respiratory gas exchange measurements, the use of another criterion measure should therefore be considered. Freedson et al. (2005) discuss the complexity and difficulties associated with using and interpreting a biological criterion measure in children, and suggest that a behavioural criterion measure, specifically direct observation, to be an effective alternative method.

In summary, many feasibility and validity issues were identified for the use of breath by breath respiratory gas exchange. Considering this is a criterion measure, any validity issues require further investigation. Furthermore, as between group differences were identified for the effect of alerting thresholds, use of the Ultima CPX breath by breath system in children with intellectual disabilities needs to be better understood. Based on the limited feasibility and validity for children with intellectual disabilities, breath by breath respiratory gas exchange is not a feasible method for use in this population and alternative, non-invasive methods, such as direct observation, should be considered.

5.5.7 Accelerometry and $\dot{V}O_2$ (research question 12)

As the relationship between $\dot{V}O_2$ and counts is the basis for calibration, a difference in this relationship between typically developing children and children with intellectual disabilities would raise questions on the validity of previous studies which generalised cut points between these groups. Specifically, the validity of using cut points developed for typically developing children in children with intellectual disabilities is partially based on the assumption that the relationship between $\dot{V}O_2$ and counts is the same for both these groups. The findings from this study, however, show a significant ($p < .0001$) difference in the relationship of counts and $\dot{V}O_2$ between typically developing participants and participants with intellectual disabilities. Therefore, the prediction of intensity classification for children with intellectual disabilities based on typically developing cut points will introduce systematic error and validity issues.

The relationship between counts and $\dot{V}O_2$ is complex, as it is not linear across all activities. It varies across activity intensities and patterns, with a non-linear relationship present during sedentary behaviours and a more linear relationship present during physical activity (Freedson et al., 2005; Treuth et al., 2004). This is confirmed in Figures 5.5-5.8, which show the linear model to be a better fit for physical activity data than sedentary data in children with intellectual disabilities, although a more linear relationship for physical activity was only found for the vertical axis in typically developing children; yet, the coefficient of determination values are still low. However, no previous studies were identified which specifically investigated differences in this relationship between population groups. Within this study, as illustrated in Figures 5.5–5.8, the linear models for both physical activity and sedentary data were a better fit for participants with intellectual disabilities compared to typically developing participants for both the vertical axis and vector magnitude.

These low coefficient of determination values suggest that a linear model is not appropriate for this data, even though there were significant correlations for participants with intellectual disabilities and typically developing participants for $\dot{V}O_2$ and counts. The poor fit of this linear model suggests the need for further investigation into a more appropriate regression model to compare the relationship between participants. However, there is no regression model that can be generalised to any data set, with the large number of regression equations developed limiting comparison between studies (Bassett et al., 2012).

Further investigation into the development of an appropriate model for the comparison between participants within this study, however, would be speculative due to the threshold issues previously discussed, thus preventing the use of a known valid measure of $\dot{V}O_2$. However, as age and maturation affect calibration and the relationship between $\dot{V}O_2$ and counts, the difference in this relationship could instead be a factor associated with age, as typically developing participants were significantly older (Freedson et al., 2005). However, the low participant numbers in this study prevented matching or direct comparison.

Another finding within this study that could affect the comparison between groups and the calibration of future cut points is the number of data points that show relatively high $\dot{V}O_2$ scores at zero counts. Low activity counts which have high $\dot{V}O_2$ measurements are type I errors, specifically false positives. This results in sedentary or low intensity activity being misclassified as moderate or vigorous intensity, which will affect the accurate prediction of activity energy expenditure and intensity. It is therefore important to set thresholds for the inclusion of data to limit the effect of false positive and false negative data points (Treuth et al., 2004). With the design of a future calibration study in mind, consideration should be given as to whether these errors are exclusively statistical errors or if they could, at least in part, be measurement errors that were an effect of the protocol.

One possible cause of these spurious points could be random error due to stress and anxiety of participants, which has an effect on respiratory measurements. Also, systematic errors for $\dot{V}O_2$ could be present due to the threshold settings used, which may result in measurements that are consistently higher than the true mean. Alternatively, it could be related to the protocol, specifically excess post-exercise oxygen consumption (EPOC). After a bout of exercise, $\dot{V}O_2$ does not immediately return to resting levels but instead decreases gradually. Within session one, the $\dot{V}O_2$ demands for each activity did not steadily increase throughout the session. As can be seen in Figure 5.3, $\dot{V}O_2$ was higher for 6 km/h than 5 km/h at 5%, therefore EPOC could have an effect on the validity of the measurements recorded during 5 km/h at 5%.

Similarly, in session two, stationary activities followed dynamic activities, e.g. standing playing the DS was preceded by passing a football and throw/catch. Therefore, if the rest between these activities was not sufficient for $\dot{V}O_2$ to return to resting levels, EPOC could have produced higher $\dot{V}O_2$ levels which were not representative of the activity and counts recorded. Furthermore, as previously discussed, during the vigorous free-living activities, such as step aerobics, movement and intensity was not constant as participants intermittently stopped due to fatigue, which could result in high $\dot{V}O_2$ scores being recorded when the participant was stationary. Step aerobics was specifically discussed by Treuth et al. (2004) as an activity which had a low correlation between counts and $\dot{V}O_2$, and records a high number of error scores; however, this could be due to the accelerometer being less accurate at measuring the stepping movement.

In summary, this study was the first to highlight the significantly different relationship between accelerometer counts and $\dot{V}O_2$ between children with intellectual disabilities and typically developing children. Although these results need to be interpreted with caution, due to the small sample size and validity issues with $\dot{V}O_2$ identified within this study, this could have important implications for the generalisation of cut points. Therefore, this further highlights the need for population-specific methods of data interpretation to be developed for children with intellectual disabilities.

5.5.8 Recommendations for future research

Although additional research is required before definitive conclusions can be made regarding feasibility, initial methodological recommendations for the design of a calibration study involving children with intellectual disabilities are:

1. Treadmill-based activities should not be generalised from protocols designed for typically developing children; instead, speeds should be self-selected or age-appropriate speeds developed.

2. Free-living activities, which can be successfully generalised from typically developing protocols, should be incorporated due to the high completion rates

3. REE and $\dot{V}O_{2max}$ should be estimated using validated non-invasive methods. In terms of future research, it is recommend that the suitability and validity of breath by breath respiratory gas exchange measurements is further investigated.

4. An effective recruitment strategy has to be developed and reasons for the low recruitment rate of girls needs to be better understood.

5. Considering the urgent need to calibrate accelerometry for children with intellectual disabilities, a field-based calibration protocol utilising a non-invasive criterion measure is recommended.

5.5.9 Strengths and limitations

This was the first study which begins to address the lack of population specific cut points for children with intellectual disabilities. Rather than assuming a calibration protocol could be successfully generalised from a calibration study involving typically developing children, this study aimed to ensure the development of an effective and feasible protocol for children with intellectual disabilities. The feasibility of all the primary aspects of a calibration study, from recruitment to data analysis and outcomes, were investigated. This provides a wide range of information which is not only relevant to the design of future calibration studies, but also physical activity research in children with intellectual disabilities in general. Therefore, the results of this study can be used to ensure that activities included in a full-scale calibration study are appropriate for this group and effective for calibration. In addition, the observational aspect of this study also resulted in interesting findings that require further investigation, such as parental influence and perceived competence.

On the other hand, the low recruitment rate of children with intellectual disabilities is a factor which significantly limits this study. Although the low participant numbers highlights the difficulties with recruitment, it prevented direct comparison with a group of matched typically developing participants. This would have enabled a direct comparison of physiological differences between these groups that could additionally affect calibration. Furthermore, although the effect of thresholds was an interesting finding that requires further investigation, the lack of a criterion measure for $\dot{V}O_2$ limits the validity of the results related to $\dot{V}O_2$ within this study. However, results including $\dot{V}O_2$ measurements for activities were presented with $\dot{V}O_2$ scores for both threshold settings to limit this effect.

5.5.10 Conclusions

The purpose of this study was to investigate the feasibility of a laboratory-based accelerometer calibration protocol in children with intellectual disabilities. Findings from this study suggest that the methods used within a calibration protocol for typically developing children cannot be generalised to children with intellectual disabilities. The physical demands of the treadmill-based activities were too high for participants with intellectual disabilities to enable protocol completion, therefore a range of speeds for each intensity is suggested. The direct measurement of aerobic fitness, using a treadmill-based graded exercise test, and REE was not feasible within this study; however, as these are important aspects of the calibration process, consideration should be given to the use of methods which estimate these measurements. Consideration should also be given to the order of the protocol to limit the error caused by anxiety and EPOC.

As a significant difference in the relationship of counts and $\dot{V}O_2$ between groups was identified, equations calibrated in typically developing children may not be appropriate for children with intellectual disabilities. This further highlights the need for cut points to be specifically calibrated for children with intellectual disabilities. It is therefore crucial that the findings from this study are used to inform the design of a calibration study to ensure the validity of physical activity measurement in children with intellectual disabilities. However, due to the measurement and equipment difficulties relating to respiratory gas exchange measures and $\dot{V}O_2$, this laboratory-based methodology is not feasible. The use of a behavioural criterion should therefore be considered. This criterion measure would also allow the study to be conducted outwith a laboratory setting, which could additionally overcome the problems with recruitment.

Chapter 6 – Calibration of the ActiGraph wGT3X+ accelerometer in children with intellectual disabilities

6.1 Overview of this chapter

The previous chapters in this thesis have highlighted the need for accelerometer calibration to be conducted in children with intellectual disabilities. However, as the feasibility of conducting accelerometer calibration using a laboratory-based protocol was deemed not to be feasible for children with intellectual disabilities, an alternative protocol needs to be developed. This chapter will discuss the design of a field-based protocol and the development of the first population-specific accelerometer cut points for children with intellectual disabilities.

6.2 Introduction

Similar to the design of a laboratory-based study, as discussed in Section 5.2, the methods, protocol, and criterion measure used in a field-based study will affect calibration. The following sections will discuss the methods employed in previous field-based studies which calibrated accelerometry in typically developing children and provide a rationale for why a field-based study is a feasible and valid alternative design for a calibration study in children with intellectual disabilities.

6.2.1 Calibration

Calibration is the process of developing new cut points by calibrating activity counts against a known biological or behavioural measure. The translation of raw acceleration into a biological value is a form of validity-based research specifically referred to as “value calibration” (Welk, 2005). Calibration is a complex process, however, and there are many challenges in deriving a biological meaning from raw biomechanical measures of acceleration (Freedson et al., 2005).

There are also additional issues associated with calibration involving children, due to the relationship between energy expenditure and body mass, and the influence of maturation (Freedson et al., 2005). Ambulatory movements vary between children, with the biomechanics of walking still developing and changing up to the age of 12 years (Cavagna, Franzetti, & Fuchimoto, 1983). During maturation, leg length relative to trunk length increases, muscle fibres increase, and control of fine and gross motor function improves, which all impact on the biomechanics of walking and the forces generated during lower limb movement (DeJaeger, Willems, & Heglund, 2001). Furthermore, as the acceleration signal recorded by the ActiGraph is affected by stride length and cadence, the raw output will potentially vary even in children walking at the same speed (Brage, Wedderkopp, Andersen, & Froberg, 2003).

The energy costs of movement relative to body mass decreases with age, which could introduce error between samples of different ages, as the net cost of walking can be up to 70% higher in younger children compared to adolescents and adults (DeJaeger et al., 2001). Furthermore, weight status impacts on calibration as the energy costs of activity are higher in children who are heavier (Brage et al., 2003; Davies, 1980). This has important implications for children with intellectual disabilities, as this group have higher rates of obesity than typically developing children (Borremans, Rintala, & McCubbin, 2010; Rimmer, Rowland, & Yamaki, 2007; Whitt-Glover et al., 2006).

This variance between children may be higher in children with intellectual disabilities, as gait abnormalities and walking difficulties are associated with this population. Specific disabilities are associated with abnormal gait patterns, with the energy costs of walking being as much as three times higher in children with cerebral palsy in comparison with their typically developing peers (Unnithan, Clifford, & Bar-Or, 1998). Down syndrome is also associated with hypotonia, which can decrease the force generated with movement and increase the energy costs, due to lower walking efficiency (Ulrich, Haehl, Buzzi, Kubo, & Holt, 2004). Furthermore, children with intellectual disabilities have low reported levels of cardiorespiratory fitness in comparison with typically developing children (Frey et al., 2008). Therefore, the energy costs of activity will be higher in individuals with a lower level of fitness, which will introduce bias if cut points are used in a population with different fitness levels from the calibration sample (Freedson et al., 2012).

In summary, considering possible cardiorespiratory and biomechanical differences between children with intellectual disabilities and typically developing children, it is important that cut points are calibrated specifically for children with intellectual disbilities. However, developing a calibration protocol is complex and must be designed with the population of interest, study outcomes, and available resources in mind. As little research has been conducted in this field, our knowledge is limited regarding the feasibility of many of the methods employed for calibration in children with intellectual disabilities, as discussed in Chapter 5. Therefore, the following sections will discuss the different methods which can be used for calibration using a field-based protocol, and the validity and feasibility issues related to these methods, specific to children with intellectual disabilities.

6.2.2 Field-based protocol

A field-based study refers to a protocol which is conducted in the participant's own environment, such as a school. The primary advantage of this design is that it is possible to develop a protocol which more accurately captures the idiosyncratic activities conducted by children, in comparison to a structured laboratory-based protocol, which has important implications for calibration (Bassett et al., 2012). As the cut points developed during calibration are based on the activities included within the protocol, if calibration is conducted on movements which are not representative of children's activity behaviours, this will introduce systematic error and reduce the ecological validity of the developed cut points. The protocols used in previous field-based studies involving typically developing children have included unrestricted free-play, semi-structured sessions, and structured activity protocols. However, accelerometer calibration using field-based protocols has not been widely conducted; therefore, as the protocol for accelerometer validation requires the same elements as for calibration, the design of these studies will also be discussed. Table 6.1 describes previous field-based calibration and validation studies conducted in children which used a non-invasive criterion measure.

Table 6.1. Previous field-based calibration and validation studies conducted in children

Study	Participants	Protocol	Direct observation method
De Decker et al. (2013)	45 preschool children 4-6 years	1 hour unrestricted free play during preschool class time.	Observation tool developed to distinguish between sedentary and non-sedentary behaviour, based on previous tools Second by second measurements recorded using Vitessa software. Data then converted into 15-sec epochs depending on whether > or < 10-sec of epoch was recorded sedentary
Kahan et al. (2013)	69 preschool children 4-5 years	Unstructured outdoor play	OSRAC-P tool Data recorded in 30-sec epochs (5-sec observe/25-sec record)
Hislop et al. (2012)	31 preschool children 3-5 years	1 hr outdoor nursery free play	Adapted CARS tool Data recorded in 15-sec epochs

Table 6.1. Continued

Study	Participants	Protocol	Direct observation method
Mackintosh et al. (2012)	28 children 10-11 years	6 free-living activities: - drawing (10-min) - DVD (10-min) - self-paced walking (5-min) - self-paced jog (5-min) - playground games (hopscotch, Frisbee, reaction ball; 3.3-min each) - free choice games (10-min)	SOFIT Activity coded in 10-sec epochs
Kelly (2005)	78 preschool children 3-4 years	Structured play class (not structured by researchers). Duration ranged between 39 to 45-min	CPAF 1-min epochs recording activity of duration > 15-sec
De Bock et al. (2010)	33 preschool children 3-6 years	Observed during preschool day for 150-min	CARS All activity lasting > 3-sec was recorded in 15-sec epochs

Table 6.1. Continued

Study	Participants	Protocol	Direct observation method
Capio et al. (2010)	31 children with cerebral palsy 6-14 years	Structured session (12-min, each activity conducted for 2-min): - sitting - standing - standing with intermittent ball dribbling - walking with intermittent standing ball dribbling - walking - jogging Unstructured session: - 10-min free play	SOFIT Activity coded in 15-sec epochs
Welk et al. (2007)	30 children 8-12 years	Structured session (calibration; each activity conducted for 2-min): - sit - stand & dribble ball	CARS used as basis for analysis; adapted from 5 to 4 codes based on pilot testing: category 4 & 5 combined into 1 category

Table 6.1. Continued

Study	Participants	Protocol	Direct observation method
Welk et al. (2007) continued		- walk & dribble ball - continuous walking - jogging and dribble ball - walking/jogging - jogging Unstructured session (cross-validation): - 10 to 12-min free play	Activity was recorded using Behavioral Evaluation System & Taxonomy (BEST) to allow real time coding.
Coe & Pivarnik (2001)	10 boys aged 12.8 ± .40 years	Basketball team session (55mins), including: warm-up, ball handling, shooting, running, and scrimmages drills	CARS Data recorded at each change in intensity and averaged into 1-min epochs

Unstructured free-play protocols are generally conducted during school, specifically during recess and classroom time (De Decker et al., 2013; Hislop et al., 2012; Kahan et al., 2013). This protocol has the highest ecological validity due to calibration being conducted on activities which are fully representative of children's free-play. However, considering children spend little time active at a moderate to vigorous intensity, this could impact on calibration. In typically developing children, Kahan et al. (2013) conducted an unstructured free-play protocol for an accelerometer validation study and found the study sample did not participate in sufficient moderate to vigorous intensity activity during free-play, which limited the scope for investigation into validity at this activity intensity.

Semi-structured protocols vary in comparison with free-play protocols as the sessions are designed to get children active. These protocols can include physical education, sport training sessions, activity sessions, or a combination of semi-structured and free-play sessions (Capio et al., 2010; Coe & Pivarnik, 2001; De Bock et al., 2010; Kelly, 2005). Furthermore, field-based studies have also included constant and structured activities more commonly used in laboratory-based studies, such as sitting, standing, or constant running, thus limiting the ecological validity of these studies (Mackintosh et al., 2012; Welk, Eisenmann, Schaben, Trost, & Dale, 2007).

It is important to consider the feasibility and suitability of previous field-based protocols specific to children with intellectual disabilities to ensure the most appropriate protocol for calibration is used. However, the literature relating to how active children are in different environments is limited and conflicting. Previous studies report that children with intellectual disabilities spend between 38.10% and 78.30% of recess at a moderate to vigorous intensity (Faison-Hodge & Porretta, 2004; Pitetti et al., 2009). Similarly, the percentage of physical education classes spent in moderate to vigorous intensity ranges from 24.00% to 52.80%; however, higher percentages have been reported for adapted physical education (78.20%; Faison-Hodge and Porretta, 2004; Pitetti et al., 2009; Sit et al., 2008). With this conflicting evidence regarding how active children with intellectual disabilities are during free-play and physical education, a semi-structured activity protocol may be more appropriate, as activities could be included to ensure children are active at the required intensities.

In summary, the type of free-living activity protocol used will potentially impact on the effectiveness of calibration and therefore requires consideration. As there is no consistent findings on how active children with intellectual disabilities are during unstructured free-play and physical education, designing a study-specific semi-structured protocol will help ensure children are sufficiently active at each intensity without reducing the ecological validity associated with structured activities.

6.2.3 Criterion measure

Of the criterion methods which can measure dimensions of physical activity (calorimetry methods, doubly labeled water, and direct observation), for a field-based protocol direct observation is the only feasible method which will capture physical activity intensity. Furthermore, as direct observation is non-invasive, use of this measure will resolve the feasibility and validity issues associated with breath by breath respiratory gas exchange, as discussed in Chapter 5.

Unlike calorimetry methods and doubly labeled water, which measure physiological parameters, direct observation is a behavioural measure. Due to the complex movements of children and the changing movements with age, there are various direct observation tools available which categorise different types of movements and postures. The second element of direct observation is the sampling method used to code activity, which also varies between observation tools. Therefore, the following sections will discuss four observation tools which have previously been used in calibration and validation studies involving children, and the data coding methods used for each of these tools.

6.2.3.1 Children's Activity Rating Scale (CARS)

The CARS is one of the most commonly used direct observation tools to measure physical activity, and has been validated in typically developing children (Puhl, Greaves, Hoyt, & Baranowski, 1990). This tool codes activity on a 5-point scale where: 1 = stationary/no movement; 2 = arm/trunk movement whilst stationary; 3 = slow, easy-paced movement; 4 = medium/moderate paced movement; 5 =

fast/strenuous activity. Activities are continually coded every time the participant transitions from one category to another for > 3 seconds, with these codes averaged for a 1-minute epoch.

In practice, however, use of this tool varies between studies. This tool was used by Coe and Pivarnik (2001) to validate accelerometry in boys during a basketball training session, and used as per its original design. However, more recent studies have modified the data sampling method from that originally devised by Puhl et al. (1990). In an accelerometer calibration study involving preschool children, De Bock et al. (2010) coded free-living school activity using the original sampling procedures, however, these activity levels were not time averaged to one minute epochs. Hislop et al. (2012) also used the original coding categories but with a modified 15-second time sampling method, which was originally developed by Sirard, Trost, Pfeiffer, Dowda, and Pate (2005). For this method, rather than continually recording changes in activity and averaging these measurements, activity was observed for 15 seconds and recorded for 15 seconds, resulting in two measurements per minute. As children have sporadic movement patterns, similar to accelerometry, the use of a real time or shorter time sampling epoch will more accurately capture children's activity patterns. However, as a limitation of direct observation is that it is a time intensive measure for researcher to use, this will be amplified with increasing numbers of measurement epochs.

6.2.3.2 Children's Physical Activity Form (CPAF)

The CPAF categorises physical activity into four categories: 1 = stationary with no movement; 2 = stationary with limb movement; 3 = slow trunk movement; and 4 = rapid trunk movement. The tool has shown moderate validity against heart rate in children ($r = .61$, $p < .05$) and used as a criterion measure for accelerometer validation during structured activity (Kelly, et al., 2004; O'hara, Baranowski, Wilson, Parcel, & Simons-Morton, 1989). The coding of activity using CPAF, however, is more complex than other tools as only "clean" epochs are included in the analysis. Activity is coded across 1-minute epochs, but is only recorded if the activity is conducted for > 15 seconds. Furthermore, only one code of activity can be included within each 1-minute epoch. That is, only epochs where the intensity was completed for the entire epoch are included in the analysis.

An advantage of this tool is that direct observation codes are matched with accelerometer epochs that are not skewed by the changes in activity during the epoch. However, as children do participate in sporadic activity, there are generally a large number of epochs excluded from the analysis. For example, Pulakka et al. (2013) used the CPAF to calibrate accelerometry in young children during free-living activity, yet of the 9,081 epochs recorded during this study, 6,904 were excluded for not being clean. Therefore, as a result of using the CPAF, calibration is conducted on data which does not fully represent the activity patterns conducted by the study sample.

6.2.3.3 Observation System for Recording Physical Activity in Children – Preschool Version (OSRAC-P)

The OSRAC-P is based on the CARS tool in relation to categorising physical activity intensity (Brown et al., 2006). However, unlike the CARS, the OSRAC-P measures behavioural, social, and contextual decisions. The OSRAC-P accounts for the previously discussed limitation with CARS, as it utilises shorter 5-second time sampling observation intervals, which will more accurately capture physical activity behaviours. This tool was used as a criterion measure by Kahan et al. (2013) to validate pre-existing cut points in preschool children aged 4 to 5 years during unstructured free-play. The elements of this tool have additionally been used to adapt pre-existing child tools for use in pre-school children (Sharma, Chaung, Skala, & Atteberry, 2011).

6.2.3.4 System for Observing Fitness Instruction Time (SOFIT)

The SOFIT tool was originally developed for the measurement of physical activity, lesson context, and teacher behaviour during physical education classes (McKenzie et al., 1991). However, the physical activity element of this tool, which codes activity into 5 categories of body posture and movement, has since been used independent of the other elements for the measurement of physical activity in children in various environments. Within SOFIT, codes 1 to 3 represent body posture (lying down, sitting, and standing, respectively), code 4 represents walking, with code 5 representing any activity conducted at a higher intensity than ordinary walking (McKenzie, 2009).

SOFIT was originally designed for data to be recorded in 20-second epochs, in 10-second observe/record intervals, i.e. three measurements per minute. However, similar to other tools, SOFIT has been adapted in more recent studies. Lafleur et al. (2013) combined codes 1 and 2 to develop a four-point scale and recorded data every 10 seconds. To overcome the limitations with averaging observe/record coding intervals, Spruijt-Metz et al. (2009) used a modified continuous observation system (SOFITCO) which also continuously recorded fidgeting to account for non-exercise energy expenditure. Furthermore, Keating, Kulinna, and Silverman (1999) developed a computerised version of SOFIT to replace the traditional pencil and paper method. However, these modifications have not been widely validated and the use of these modified versions has been limited outwith these studies.

This tool has been widely used for the measurement of physical activity in both typically developing children and children with intellectual disabilities (Hinckson & Curtis, 2013; McKenzie, 2002; Sallis & Saelens, 2000). Furthermore, unlike the other direct observation tools discussed, an advantage of SOFIT is that it has been validated specifically in children with intellectual disabilities, both for the psychometric properties of the test and against criterion measures of physical activity (Faison-Hodge & Porretta, 2004; Taylor & Yun, 2006).

6.2.4 Summary

Based on the previous findings discussed in Chapter 5, a calibration study including a field-based protocol and non-invasive measure needs to be conducted for children with intellectual disabilities. This section has highlighted the various types of field-based protocols which can be used and discussed the strengths and weaknesses associated with each of these designs. Considering the importance of collecting data across a range of intensities, the use of a semi-structured protocol will help ensure that children complete sufficient activity at each intensity to ensure calibration is conducted on a large, representative data set.

For a field-based study, direct observation is the only feasible non-invasive criterion measure. There are, however, various tools which enable the collection of direct observation data, each using various activity categories and coding procedures. However, most of the research discussed has included a sample of preschool children, thus limiting the generalising of previous validity and feasibility findings to older children. As it is vital that the tool used for calibration is valid for use in children with intellectual disabilities, SOFIT is the most appropriate tool. Furthermore, the relative simplicity of the SOFIT activity categories and coding procedures make it the most feasible direct observation tool for use in the present study.

6.2.5 Research questions

The purpose of this study is to calibrate the ActiGraph wGT3X+ accelerometer for the estimation of physical activity intensity in children with intellectual disabilities. Furthermore, this study also aims to add to the existing literature regarding the validity of accelerometer counts and heart rate in children with intellectual disabilities. These aims will be achieved using the following research questions:

RQ 13: Does heart rate provide acceptable criterion validity for the measurement of total physical activity in children with intellectual disabilities?

RQ 14: Does the ActiGraph wGT3X+ provide acceptable criterion validity for the measurement of total physical activity in children with intellectual disabilities?

RQ 15: What are the optimal ActiGraph wGT3X+ vector magnitude cut points for the classification of sedentary, moderate, and vigorous intensity activity for children with intellectual disabilities?

RQ 16: What are the optimal ActiGraph wGT3X+ vertical axis cut points for the classification of sedentary, moderate, and vigorous intensity activity for children with intellectual disabilities?

6.3 Method

6.3.1 Ethical considerations

This study was approved by the Medical, Veterinary, and Life Sciences College Ethics Committee, University of Glasgow (Appendix iv). Written informed consent was required from both participants and parents prior to participation. Verbal consent was additionally sought from participants prior to each activity session.

6.3.2 Participants

6.3.2.1 Recruitment

Five additional support needs primary schools in the West of Scotland, which were specifically for children with mild to moderate intellectual disabilities, were used for recruitment and data collection. One researcher (AM) visited the participating schools, explained the study to children in primaries four to seven, and handed out information packs containing parent and child information sheets and consent forms (Appendices v & vi). In total, 86 information packs were handed out to eligible children (60 boys, 26 girls). If children were willing to participate, parents were asked to return a signed parent and child consent form to the school. Dates and times for the sessions were decided approximately two weeks after the information packs had been handed out through discussion with teachers. This allowed time for consent forms to be returned so that an appropriate number of sessions could be arranged, depending on the number of participants, and at a time that all participants were available.

The process of recruitment was slightly altered after discussion with a teacher at one of the participating schools. Specifically, the wording on the information sheets was changed from “intellectual disabilities” to “learning disabilities”, as the teacher discussed that parents had previously noted concerns with the term “intellectual disabilities”, therefore the term “learning disabilities” had instead been adopted by the school.

6.3.2.2 Inclusion and exclusion criteria

The following inclusion and exclusion criteria were developed to ensure a sample of children with intellectual disabilities within a relatively small age-range were recruited. Furthermore, due to the nature of the physical activity sessions, it was important that participants were independently ambulatory. Therefore, the following inclusion and exclusion criteria were developed:

Inclusion criteria:

- having intellectual disabilities
- aged between 8 to 11 years
- independently ambulatory

Exclusion criteria:

- having a physical disability
- having a developmental disability, without a specific diagnosis of intellectual disabilities

6.3.3 Protocol

6.3.3.1 Development of the session

The physical activity session was designed specifically for this calibration study. As previously discussed in Section 6.2.2, the content of a session for the purpose of calibration can have important implications on results. In addition to developing the session based on previous research, discussions were had with an Active Schools coordinator and the deputy head teacher at one of the participating schools; Active Schools is a programme aimed at developing and supporting the delivery of quality sporting opportunities for children in Scotland.

The Active Schools coordinator assisted with ensuring the appropriateness of the games and activities to be included within the session. They provided information regarding activities which were regularly conducted in the participating schools, which allowed the inclusion of activities that were familiar to the children. It was assumed that the inclusion of familiar activities would help the session flow well by reducing the in-depth instructions required for unfamiliar games, and thus increase compliance. After initial development of the session, the appropriateness of the content was discussed with the teacher. These discussions were primarily focussed on ensuring that the activities were explained in an appropriate way, specifically that activities were explained with a visual demonstration, where possible, rather than verbal instruction.

Therefore, based on previous research and the discussions with the Active Schools coordinator and teacher, the session was designed with the following factors in mind:

- Activities should include a variety of movements, with a focus on activities which correspond with the SOFIT categories

- The intensity of activities increase from sedentary through to vigorous as the session progresses, which is of specific importance to the measurement of heart rate

- Activities, specifically the vigorous games, are familiar to the study sample to reduce the instruction time

- Activities to be included which do not require complex instruction and could be visually demonstrated

- Activities should not involve complex skills or movements.

Furthermore, the appropriateness of the session was assessed during an initial pilot session. This session was conducted by two researchers, included eight participants (boys = 7) and lasted 25 minutes (session 1; Table 6.6). Session organisation was investigated relating to the time taken to do anthropometric measurements, layout of the hall, and feedback from participants relating to enjoyment, which were recorded using timing sheets and notes. In addition,

Table 6.2. Description of school-based session protocol and predominant movements conducted

Content	Content Description	Predominant Movement	Intensity	Time (min)
Instruction	• Explain lay out of hall, e.g. stay within coned area, keep back from/ignore cameras	Sitting	Sedentary	2
	• Explain session: will be fun & include different games, etc.			
Warm up	• Lying down stretches			
	- Find space on floor: make star shape, make arrow shape	Lying	Sedentary	2
	• Standing Stretches	Standing	Light	3
	- Stretch up/touch toes x 5			
	- Stretch arm across chest x 3 (per arm)			
	- Arm rotation forward and back (x 10 each)			

Table 6.2. Continued

Content	Content Description	Predominant Movement	Intensity	Time (min)
Warm up continued	• Active Stretches - Sitting: 1) pass ball overhead to person behind, then back down line (x 2) 2) repeat and pass ball at right side going behind, left side coming down line (x 2)	Sitting	Sedentary	5
	- Standing: 1) pass ball through legs going back, overhead going forward down line (x 2)	Standing	Light	5
Instruction game	• Walk about area following instructions - touch floor with right/left hand - high five next person they pass - walk with hands on head	Walking	Moderate	5

Table 6.2. Continued

Content	Content Description	Predominant Movement	Intensity	Time (min)
Instruction game continued	• Walk about area following instructions - touch “X” coloured cone - get into groups of 2/3 - turn all cones upside down - turn all cones right way up	Walking	Moderate	10
Obstacle game	• Complete obstacle game (x 2) and walk back to join line. Repeat x 5. 1) Obstacle course 1: - walk between and touch zig zag cones - 5 step-ups on aerobic step - kick ball against bench - walk to back of line	Standing Walking Running	Light Moderate/ vigorous Vigorous	10

Table 6.2. Continued

Content	Content Description	Predominant Movement	Intensity	Time (min)
Obstacle game continued	• Complete obstacle game (x 2) and walk back to join line. Repeat x 5. 2) Obstacle course 2: - run through agility ladder - pull hula hoop over head - head ball against bench - run to back of line	Running	Moderate/ vigorous Vigorous	
Dodge ball game	• For each game, two participants were selected to be “catchers” and given a ball. The other participants had to avoid/“dodge” being touched with ball.	Running	Vigorous	15
Active cool down	• Walk and collect in all equipment	Walking	Moderate	5

session descriptive statistics (Table 6.6) were calculated to investigate if the session was effective at getting participants active at each intensity.

6.3.3.2 Session observation

Physical activity sessions were recorded using two wide-lensed video cameras (GoPro Hero3, CA, USA). The use of video recording has previously been used successfully in conjunction with direct observation in children with intellectual disabilities (Faison-Hodge & Porretta, 2004). One camera was placed at the front of the hall (camera one) and the other was adjacent at the side of the hall (camera two). Both cameras were positioned on tripods to allow the greatest visibility of the hall. Camera one was used for data analysis with camera two footage only used if a child was obscured from the view of camera one. Recordings started prior to the session commencing and were stopped at the completion of the session. The video cameras automatically recorded the time when each recording commenced. This time was manually synchronized to match the internal clock of the computer that was used to initialise the accelerometers, which ensured that the video, accelerometer, and heart rate data could be accurately time-matched to the second.

6.3.4 Measures

Throughout the activity session, participants wore an accelerometer and heart rate monitor. In addition, the session was recorded to allow direct observation analysis. To ensure the accurate identification of participants during the analysis of the session recordings, participants wore a coloured bib which corresponded with their participant identification number. Similar methods of participant identification, specifically coloured wristbands, have been used in previous research utilising direct observation in children (Pope, Coleman, Gonzalez, Barron, & Heath, 2002).

6.3.4.1 SOFIT

The SOFIT is a momentary time sampling direct observation tool which enables physical activity behaviours to be recorded (McKenzie et al., 1991). This observation tool consists of three phases (student activity, lesson context, and teacher behaviour) and was initially developed for the assessment of physical education classes. The ‘student activity’ element of this tool, which is the only SOFIT element used in the present study, is designed to categorise physical activity behaviours; code 1 = lying down, code 2 = sitting, code 3 = standing, code 4 = walking, code 5 = very active. Table 6.3 describes the SOFIT coding categories in full.

Table 6.3. SOFIT activity codes and movements associated with each code

Code	Typical category movements
1. Lying down	– Lying on front or back – Body parallel to floor – If moving, energy expenditure of movement should not exceed that of ordinary walking
2. Sitting	– In seated posture – If moving, energy expenditure should not exceed ordinary walking, e.g. sit-ups are very active
3. Standing	– Body posture adjacent to the floor – If moving, for example standing stretching or moving on the spot, energy expenditure should not exceed that of ordinary walking
4. Walking	– Walking from one points to another – Walking speed equal to or slower than ordinary walking; fast-paced walking is coded as very active
5. Very active	– When expending more energy than during ordinary walking, e.g. running, jogging, skipping – Includes movements associated with other activity codes that require higher energy expenditure, e.g. sit ups, or standing with vigorous upper body movements

Activity is coded every 20 seconds using 10-second observe/record intervals, yielding 3 observations per minute. This coding process is paced using pre-recorded audio MP4 files, developed by McKenzie (2009), which prompts the rater when to observe and record activity. The behaviour being conducted at the record prompt, i.e. at the end of the 10-second observe interval, is coded. If the participant is transitioning from one activity to another, the activity is recorded as the higher code; for example, if the participant is transitioning from sitting to standing at the end of the observe interval, the activity would be coded as standing.

The validity of SOFIT as a measure of physical activity has been established in typically developing children and children with intellectual disabilities. In typically developing children, McKenzie et al. (1991) originally validated SOFIT against heart rate and found incremental increases in heart rate with each SOFIT activity category. From these heart rate measurements, energy expenditure (kcal/kg/min) was estimated for each activity category: lying down = .029, sitting = .047, standing = .051, walking = .096, very active = .144. McKenzie, Sallis, and Armstrong (1994) subsequently showed that these estimated energy expenditure costs significantly correlated ($r = .74$, $p < .001$) with the CALTRAC accelerometer in a sample of 69 typically developing children. SOFIT has also been found to be valid against criterion measures of heart rate and energy expenditure for deciphering between moderate to vigorous and non-moderate to vigorous intensity activity (Rowe, Schuldheisz, & van der Mars, 1997; Rowe, van der Mars, Schuldheisz, & Fox, 2004). Specifically, SOFIT codes 1 and 2 are a valid representation of sedentary behaviours, with codes 4 and 5 a valid measure of moderate to vigorous intensity activity.

For children with intellectual disabilities, Faison-Hodge and Porretta (2004) validated SOFIT for the estimation of moderate to vigorous intensity activity in 8 children and found a strong association with heart rate ($r = .81$, $p = .01$). Furthermore, Capio et al. (2010) used SOFIT as a criterion measure to validate accelerometry ($r = .75$, $R^2 = .56$, $p < .001$) and heart rate ($r = .65$, $R^2 = .56$, $p < .001$) in children with cerebral palsy. Similar to the procedures of the present study, SOFIT has previously been used as a criterion measure of physical activity for the calibration of accelerometer cut points (Mackintosh et al., 2012) and for the development of prediction equations for activity energy expenditure in typically developing children (Honas et al., 2008).

6.3.4.1.1 SOFIT observer training

In-depth guidelines for observer training were developed by the SOFIT developer (McKenzie, 2009). These guidelines suggest that observers complete standardised classroom training which consists of understanding coding procedures, memorising coding definitions, and practice video analysis. Further to this, validity should be assessed using gold standard video segments and reliability measured in the field setting. To allow standardised training for all observers, McKenzie (2009) developed seven SOFIT training videos which cover an introduction to SOFIT, coding practice, and assessment.

Three observers (AM, CM, & VP) were trained for the coding of data within this study. Eight hours of observer training was conducted over two sessions, in accordance with the McKenzie (2009) guidelines. Session one consisted of understanding coding procedures and definitions, and initial video practice. Session two consisted of additional video analysis practice and validity assessment. Observers achieved a combined score of 86% accuracy with the gold standard assessment video, which exceeded the minimum recommended requirement of 80%.

6.3.4.1.2 SOFIT reliability measures

Field-based reliability was established in accordance with McKenzie (2009) recommendations which suggest that prior to data collection inter-observer agreement (IOA) of ≥ 80% should be achieved. To ensure consistency in the results, IOA was additionally tested at the approximate midpoint of data collection. Furthermore, intra-observer reliability was tested for the primary observer (AM). Inter- and intra-observer agreement were calculated as a percentage using the following formula:

IOA (%) = (number interval agreements / number total intervals) x 100

Initial reliability measures were conducted on two randomly selected participants from the pilot session. This represented 25% of the class, which exceeds the recommendation that reliability should be established in at least 12% of participants (McKenzie, 2009). This represented a total of 150 20-second observation intervals. Initial inter-observer reliability between the three observers was 79%, with intra-observer reliability 89%. This initial result was lower than the recommended 80% agreement, however, McKenzie (2009) suggested that IOA < 80% does not prevent measures being conducted, and instead discrepancies should be discussed. Therefore, the three observers further discussed the coding procedures and the epochs where there were discrepancies between observers.

Mid-point reliability measures were conducted using two randomly selected participants from session three. This represented 22% of the session participants and included a total of 226 20-second observation intervals. Midpoint inter- and intra-observer reliability was 85% and 91%, respectively, confirming that the lead rater (AM) was achieving the recommended standard for data collection.

6.3.4.2 Accelerometry

Physical activity was measured using ActiGraph wGT3X+ accelerometers. This device is wireless ANT+ enabled, which was utilised during this study. ANT+ allows interoperability between other wireless devices with are ANT+ enabled. Participants wore one device on their right hip at the iliac crest for the duration of the activity session. The internal specifications and use procedures of the wGT3X+ have been fully described in Sections 4.2.1.2.4 and 5.3.4.2.

6.3.4.3 Heart rate

Heart rate was measured using an ANT+ enabled wireless monitor (CooSpo, ANT+, China). As these devices are also ANT+ enabled, heart rate data was recorded wirelessly by the wGT3X+ accelerometers. An advantage of this was that participants did not need to wear a heart rate device receiver, which is usually worn on the wrist. Heart rate monitors have previously been used in children with intellectual disabilities, however, the use of a wrist worn receiver has been noted as a distraction for participants (Faison-Hodge & Porretta, 2004). Therefore, the use of wireless devices limits the amount of measurement devices worn by participants.

6.3.4.4 Anthropometric measures

Height was measured to the nearest 0.1 cm using a stadiometer (Seca Scales, Hamburg, Germany). Without shoes, participants stood with their heels together and their back against the scale. Two separate measurements were conducted and the mean value calculated. Weight was measured to the nearest 0.1 kg using digital scales (Seca Scales, Hamburg, Germany). Measurements were conducted twice, with light clothing and no shoes, with the mean value calculated. From height and weight measurements, body mass index was calculated. Full anthropometric measurements procedures have been described in Section 5.3.4.4.

6.3.5 Management of data

ActiGraph data were downloaded using ActiLife version 6.11.5 software (ActiGraph LLC, Pensacola) in 10-second epochs, measured at a sampling rate of 30 Hz. Accelerometer counts for the vertical axis and vector magnitude, and heart rate scores were extracted for analysis. Video data was time matched to the accelerometer data to ensure the SOFIT analysis started at the beginning of a 10-second accelerometer epoch. Two 10-second epochs of vertical axis counts, vector magnitude counts, and heart rate therefore corresponded with one 20-second SOFIT epoch.

For vertical axis and vector magnitude counts, the two 10-second epochs which corresponded with a SOFIT score were summed using an Excel macro to provide a combined count score for each 20-second epoch. Heart rate scores were recorded as beats per minute, therefore two 10-second epoch scores were averaged to provide a mean 20-second epoch value. This resulted in the following data formats: vertical axis counts (counts/20-sec), vector magnitude counts (counts/20-sec), heart rate (bpm), and SOFIT classification (score/20-sec). Data were then screened for spurious scores. SOFIT scores where the participant left the gym hall were excluded from the analysis. Data in this format were used for all analyses.

Prior to conducting the calibration analyses, data for 14 participants were removed to enable cross-validation analyses, with two participants randomly selected from each of the seven sessions; full details on the cross-validation procedures are presented in Chapter 7. This resulted in data from 36 participants being used for the calibration analysis presented in this Chapter.

6.3.6 Statistical analysis

All statistical analysis was conducted using SPSS 22 IBM statistical package (SPSS IBM, New York, NY, USA). Normality was assessed for all variables to ensure appropriate statistical tests were used. Full details of normality testing procedures are presented in Section 5.3.6. Descriptive statistics (mean ± SD) were calculated for all participant (age, sex, height, weight, and BMI) and session variables (session duration, percentage of session spent in each SOFIT category, and number of girls/boys in each session).

6.3.6.1 Evaluation of the pilot session

The effectiveness of the pilot session was investigated using session descriptive statistics, session observations, and participant feedback.

6.3.6.2 Validation of heart rate and accelerometry

As previously discussed, SOFIT is a criterion measure of physical activity against which other methods of measurement can be validated. Therefore, correlational analysis was used to test the relationship between SOFIT and heart rate and SOFIT and accelerometry (total activity). As heart rate data and accelerometer counts for the vertical axis and vector magnitude were not normally distributed, with log and square root transformations ineffective, relationships with SOFIT were investigated using the non-parametric Spearman's rank correlation coefficient (r_s).

6.3.6.3 Calibration of accelerometer cut points

ROC curve analyses were conducted to determine optimal cut points for the classification of sedentary, moderate, and vigorous intensity activity. ROC curve analysis quantifies the relationship between positive and negative scores for continuous data and allows a cut point to be identified which best discriminates between two conditions (Krzanowski & Hand, 2009). A score is referred to as "positive" if it represents the condition of interest (actual condition), whereas a “negative” score is not the condition of interest. The derived cut point should maximise the probability of correctly classifying positive and negative scores, i.e. true positive and true negative scores, respectively, and limit the probability of misclassifying positive and negative scores, i.e. false positive and false negative scores, respectively.

ROC curves are interpreted using sensitivity, specificity, and the AUC of the ROC curve results. Sensitivity refers to the accuracy of a cut point to correctly classify activity intensity (true positive), e.g. correctly classify vigorous activity as vigorous activity. Similarly, specificity refers to the accuracy of a cut point to exclude data which is not of the specified intensity (false positive), e.g. not misclassify moderate activity as vigorous activity. In addition, the AUC gives a statistical representation of the accuracy of the optimal cut point. The AUC is the average true positive classification rate, independent of false positive classifications. Therefore, a cut point which perfectly classifies all scores will have an AUC of 1.0, with a cut point equivalent to chance having an AUC of 0.5. The AUC scores will be interpreted using the following scale: ≥ .90 is excellent, .80-.89 is good, .70-.79 is fair, and < .70 is poor (Metz, 1978; Zweig & Campbell, 1993).

In line with previous accelerometer calibration studies, the aim of this ROC curve analysis was to identify the cut points which maximise both sensitivity and specificity (Evenson et al., 2008; Jimmy et al., 2013; Mackintosh et al., 2012; Pulsford et al., 2011; Vanhelst et al., 2011). However, it is possible to weigh sensitivity as more important than specificity, or vice versa, which will affect the chosen cut point. As ROC curves examine classification between two conditions, six separate ROC curve analyses were conducted to identify cut points for sedentary, moderate, and vigorous activity for vertical axis counts and vector magnitude counts. Subsequently, the sedentary and moderate cut points were used as lower and upper boundaries for the classification of light activity, with the vigorous cut point used as the upper boundary for moderate intensity.

Accelerometer counts (counts/20-sec) represent the independent variable. The dependant variable was a binary classification of intensity based on the SOFIT scores, with binary code 1 representing a positive score (intensity of interest) and binary code 0 a negative score (not intensity of interest). For the sedentary cut point, SOFIT codes of 1 and 2 formed a binary code of 1 (sedentary), with SOFIT codes 3, 4, and 5 forming a binary code of 0 (not sedentary). For moderate activity, SOFIT codes 1, 2, and 3 created a binary code of 0 (not moderate), with codes 4 and 5 creating a binary code of 1 (moderate). Finally, for vigorous activity, SOFIT codes 1 to 4 created a binary code of 0 (not vigorous), with code 5 creating a binary code of 1 (vigorous). The conversion of SOFIT categories into binary codes for analysis is summarised in Table 6.4. As previously discussed in Section 6.3.4.1, these SOFIT categories are a valid representation of sedentary, moderate, and vigorous intensity activity.

Sensitivity and specificity scores were then used to identify the optimal cut point. However, as the weighting of sensitivity and specificity is at the researcher's discretion based on the study aims/outcomes, the optimal cut point was identified manually. Firstly, ROC curve graphs, which plot sensitivity against specificity, were produced using SPSS and then viewed to identify the approximate optimal cut point, i.e. the point of the ROC curve that is closest to the top left corner of the graph axes. Secondly, once identified, the approximate sensitivity and specificity scores of this point were viewed on the SPSS output table which gives all possible cut points, with approximately 3000 possible cut points produced for each intensity. This provided a narrow range to view within the SPSS table, within which the optimal cut point could be identified.

Table 6.4. Summary of the conversion of SOFIT categories into binary codes for ROC curve analysis

Intensities	SOFIT categories	
	Binary code 1 (intensity of interest)	Binary code 0 (not intensity of interest)
Sedentary	1, 2	3, 4, 5
Moderate	4, 5	1, 2, 3
Vigorous	5	1, 2, 3, 4

6.4 Results

6.4.1 Pilot session

6.4.1.1 Participants

Eight children (7 boys, 1 girl, 8-10 years) with intellectual disabilities participated in the pilot session. Descriptive statistics are presented in Table 6.5.

Table 6.5. Descriptive statistics (mean ± SD) of all participants, pilot session participants, and participants in the calibration group

Characteristic		Participants	
Pilot Session	**Boys (n = 7)**	**Girls (n = 1)**	**Total (n = 8)**
Age (yrs)	8.71 ± .76	10.00 ± .00	8.88 ± .84
Height (m)	1.35 ± .07	1.37 ± .00	1.35 ±.06
Weight (kg)	31.59 ± 7.30	38.80 ± .00	32.49 ± 7.22
BMI (kg/m^2)	17.23 ± 2.26	20.67 ± .00	17.67 ± 2.42
All Participants	**Boys (n = 37)**	**Girls (n = 13)**	**Total (n = 50)**
Age (yrs)	9.35 ± 1.03	10.08 ± 1.12	9.54 ± 1.09
Height (m)	1.43 ± .09	1.42 ± .07	1.43 ± .09
Weight (kg)	39.77 ± 11.54	38.08 ± 5.50	39.33 ± 10.28
BMI (kg/m^2)	19.13 ± 4.15	18.99 ± 2.69	19.09 ± 3.80
Calibration participants	**Boys (n = 28)**	**Girls (n = 8)**	**Total (n = 36)**
Age (yrs)	9.32 ± 1.02	10.25 ± 1.04	9.53 ± 1.08
Height (m)	1.43 ± 1.02	1.43 ± .07	1.43 ± .09
Weight (kg)	39.75 ± 12.78	37.63 ± 4.83	39.28 ± 11.46
BMI (kg/m^2)	19.18 ± 4.56	18.31 ± 2.28	18.99 ± 4.15

6.4.1.2 Effectiveness of the session

Descriptive statistics for the pilot session (session 1) are presented in Table 6.6. The design of the session was effective in getting participants active in all intensity categories, in particular vigorous intensity.

Table 6.6. Descriptive statistics on session duration, participants, and percentage of the session spent in each SOFIT category

Session	School	Session duration (min)	Lying down (%)	Sitting (%)	Standing (%)	Walking (%)	Very active (%)	Boys (n)	Girls (n)
1 (pilot)	A	25	3.96	11.08	27.70	36.15	21.12	7	1
2	B	34	6.25	26.72	17.03	29.66	20.34	5	3
3	B	38	4.50	28.23	27.93	22.02	17.32	4	5
4	C	40	4.06	12.19	26.40	24.37	32.97	7	0
5	C	29	1.64	21.51	21.67	41.87	13.30	7	0
6	D	27	3.27	22.67	17.38	29.97	26.70	3	2
7	E	16	3.47	29.86	11.11	31.94	23.61	5	1

All the activities and games were of a suitable skill and cognitive level, and participants reported the session content to be enjoyable. Based on this, the session content was not changed for future sessions and, as a result, data from this session was included in the validation and calibration analysis.

Minor issues were identified during the pilot session in relation to the organisation of the session, specifically in the transition period between activities, as participants became distracted. To limit this effect, participants were asked to be involved in the set-up of games, e.g. setting out cones, which was effective. Another issues noted within the pilot session was the organisation of the anthropometric measurements and fitting of accelerometers and heart rate monitors. In the pilot session, which was conducted during a 1-hour timeslot, a large proportion of the session (approximately 30 minutes) was spent taking anthropometric measurements and fitting devices. This was due to the unfamiliarity of participants with the devices, in particular the heart rate monitor, with participants requiring demonstrations of device wear prior to agreeing to wear it.

Based on these findings, the following organisational changes were made to future sessions:

- Three researchers to be present at session
- Limit sessions to 10 participants

6.4.2 Calibration sessions

6.4.2.1 Participants

Fifty-three children with intellectual disabilities were initially recruited for this study, which resulted in a final participation rate of 50 (37 boys; 13 girls). The reasons for the three children who were initially recruited not participating were absence on the day of the session (n = 2) and being removed at the start of the session, prior to any data collection, due to disruptive behaviour (n = 1). Subsequently, participants were randomly assigned into two groups for calibration (n = 36) and cross-validation (n = 14) analyses. Descriptive statistics for all participants and calibration participants are presented in Table 6.5; descriptive statistics for participants in the cross-validation group and group assignment procedures are presented in Chapter 7.

6.4.2.2 Activity sessions

Seven activity sessions were conducted in five schools; Table 6.6 includes descriptive data of each session. The activity session was designed to be approximately 45 minutes in duration, with an additional 15 minutes for anthropometric measurements and fitting devices. This time frame was effective for most sessions, although the duration was predominately determined by the time the anthropometric measurements took, the engagement of participants, and their abilities to complete various aspects of the session. Session 7 was the only session in which the activities had to be amended due to the level of participants intellectual disabilities; the instruction and drill elements of the session were not conducted as participants had difficulties in understanding the commands and the various activities included in the drills.

6.4.3 Validation

6.4.3.1 Research question 13

Does heart rate provide acceptable criterion validity for the measurement of total physical activity in children with intellectual disabilities?

The mean heart rate scores recorded for each SOFIT category are presented in Table 6.7. Heart rate, $r_s = .42$, p (one-tailed) $< .001$, was significantly associated with SOFIT. As SOFIT is a criterion measure of physical activity, these results indicate that heart rate provides weak criterion validity for the measurement of physical activity in children with intellectual disabilities.

6.4.3.2 Research question 14

Does the ActiGraph wGT3X+ provide acceptable criterion validity for the measurement of total physical activity in children with intellectual disabilities?

The mean accelerometer counts recorded for each SOFIT category are presented in Table 6.7. Vertical axis counts, $r_s = .82$, p (one-tailed) < .001, and vector magnitude, $r_s = .80$, p (one-tailed) < .001, counts were significantly associated with SOFIT. These results indicate that wGT3X+ accelerometer counts provide excellent criterion validity for the measurement of physical activity in children with intellectual disabilities, with the vertical axis having a higher level of validity.

Table 6.7. Mean (± SD) accelerometer counts and heart rate for each SOFIT category

SOFIT category	Vertical axis counts (counts/20sec)	Vector magnitude counts (counts/20sec)	Heart rate (bpm)
1	102.74 (165.82)	331.93 (427.58)	95.53 (27.47)
2	87.39 (150.54)	398.55 (363.56)	107.06 (31.43)
3	222.16 (249.81)	650.34 (432.90)	122.76 (29.73)
4	623.58 (346.97)	1279.18 (476.01)	127.44 (32.49)
5	1402.19 (582.28)	2155.65 (672.91)	143.06 (38.73)

6.4.4 Calibration

6.4.4.1 Research question 15

What are the optimal ActiGraph wGT3X+ vector magnitude cut points for the classification of sedentary, moderate, and vigorous intensity activity for children with intellectual disabilities?

Accelerometer cut points, sensitivity, specificity, and AUC results are presented in Table 6.8. In addition, Figure 6.1 illustrates the ROC curves for vertical axis and vector magnitude counts for the classification of sedentary, moderate, and vigorous activity, with the optimal cut points highlighted.

The vector magnitude cut points which represent the optimal balance between sensitivity and specificity were ≤ 1863 cpm (sedentary), ≥ 2610 cpm (moderate) and ≥ 4215 cpm (vigorous). For the classification of sedentary, moderate, and vigorous activity, the AUC was significant (p < .001), with excellent discrimination for moderate (.92) and vigorous (.92) intensity activity, and good discrimination for sedentary (.86) behaviours. The high sensitivity (80–86%) and specificity (77–82%) scores indicate that these cut points will be effective in not misclassifying activity intensities and correctly classifying activity intensities.

Table 6.8. Sensitivity, specificity, AUC, and optimal cut points for each intensity category for the vertical axis and vector magnitude

	Sensitivity (%)	Specificity (%)	AUC (95% CI)	Cut point (counts/20-sec)	Cut point (cpm)
Vertical axis					
Sedentary	81	81	.87 (.86–.88)	≤ 169	≤ 507
Light	n/a	n/a	n/a	170–335	508–1007
Moderate	86	83	.92 (.91–.93)	336–766	1008–2300
Vigorous	88	85	.94 (.93–.95)	≥ 767	≥ 2301
Vector magnitude					
Sedentary	80	77	.86 (.84–.87)	≤ 621	≤ 1863
Light	n/a	n/a	n/a	622–869	1864–2609
Moderate	86	82	.92 (.91–.93)	870–1404	2610–4214
Vigorous	85	82	.92 (.91–.93)	≥ 1405	≥ 4215

6.4.4.2 Research question 16

What are the optimal ActiGraph wGT3X+ vertical axis cut points for the classification of sedentary, moderate, and vigorous intensity activity for children with intellectual disabilities?

The optimal cut points for vertical axis counts were ≤ 507 cpm (sedentary), ≥ 1008 cpm (moderate), and ≥ 2301 cpm (vigorous). Similar to vector magnitude cut points, the AUC was significant ($p < .001$) for all intensities, with excellent discrimination for moderate (.92) and vigorous (.94) activity, and good discrimination for sedentary (.87) behaviours. High sensitivity (81–88%) and specificity (81–85%) scores indicate that these cut points will be effective in reducing type I and type II errors.

In comparison with the results for vector magnitude discussed in Section 6.4.4.1, the vertical axis demonstrates either equal or higher sensitivity, specificity, and AUC scores for each intensity (Table 6.8). Therefore, these results suggest that the use of vertical axis cut points will provide greater classification accuracy than vector magnitude counts for moderate and vigorous intensity activity and sedentary behaviour in children with intellectual disabilities.

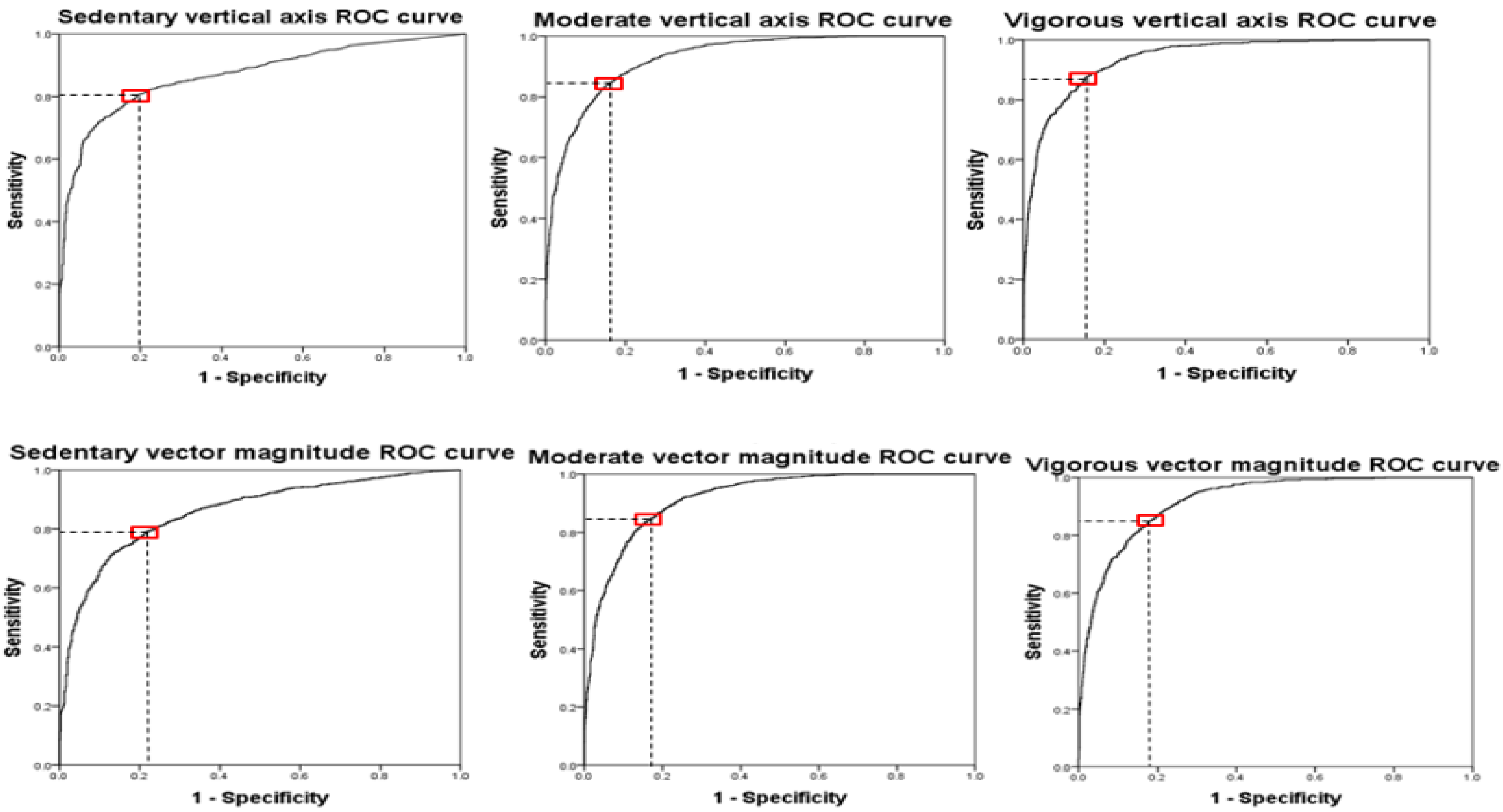


Figure 6.1. ROC curves and approximate optimal cut points for sedentary, moderate, and vigorous intensity for vertical axis and vector magnitude counts

6.5 Discussion

The primary purpose of this study was to calibrate the ActiGraph wGT3X+ accelerometer for the estimation of physical activity intensity in children with intellectual disabilities. Furthermore, this study also investigated the criterion validity of raw accelerometer counts and heart rate. Validation and calibration were conducted against a criterion measure of direct observation during a semi-structured activity session.

6.5.1 Validation

Accelerometer counts for the vertical axis and vector magnitude exhibited excellent criterion validity (r_s = .82 and r_s = .80, respectively), although heart rate only indicated weak criterion validity (r_s = .42).

6.5.1.1 Heart rate

As shown in Table 6.7, heart rate increased with each SOFIT category. This finding is expected as heart rate has a linear relationship with increased workload and the energy demands of activity, which is most apparent ≥ moderate intensity activity (Corder et al., 2008; Trost, 2007b). However, the high standard deviation scores indicate that there is high variability in heart rate between participants. This high variability could be partially attributed to intellectual disabilities. For example, heart defects are common in children with Down syndrome, who also have lower reported peak heart rate and a higher resting heart (Baynard et al., 2008). Furthermore, heart rate can be influenced by various other factors, such as stress and room temperature, which could have additionally contributed to the high variations found.

In terms of measurement, changes in heart rate are not instantaneous in relation to changing workload, which could have reduced validity, as the 20-second SOFIT epochs used may not have captured the lag in heart rate response. In addition, as the SOFIT code given for each 20-second epoch is based on the activity being conducted at the end of the observe interval, the high standard deviations could be a result of the activity code given not being fully representative of the activities conducted during the epoch.

There is limited previous research investigating the validity of heart rate in children with intellectual disabilities. For children with Down syndrome, Esposito et al. (2012) reports a weak positive relationship between heart rate and the Actical accelerometer ($r = .22$, $p < .01$). An interesting aspect of this study was that heart rate was used as a metabolic criterion measure to validate accelerometry. However, considering the variability and limitations associated with heart rate, it is generally not regarded as a criterion measure. Capio et al. (2010) validated heart rate against a criterion measure of SOFIT in children with cerebral palsy. To account for the delayed response of heart rate to changing workload, the authors only included the final 30 seconds of 2-minute structured activity data in the analysis to allow heart rate to reach a steady state. Results from this study show good criterion validity for heart rate ($r = .65$, $R^2 = .43$, $p < .001$). However, only using steady-state measurements limits the generalisability of these findings to the use of heart rate during free-living, sporadic physical activity. Furthermore, these results were calculated using linear regression, which is not an appropriate method of statistical analysis as SOFIT is a categorical measure, thus further illustrating limitations with the standard of previous measurement research conducted in children with intellectual disabilities.

In the present study, heart rate data was only collected for 42 children, as the remaining 8 declined to wear the heart rate monitor. Children who declined to wear the monitor did so for various reasons, including not feeling comfortable with the skin contact of the monitor or with the researcher or teacher putting the device on. Therefore, the heart rate data collected in this study may not be fully representative of the study sample as a whole. The use of heart rate in children with intellectual disabilities has been limited. Faison-Hodge and Porretta (2004) used Polar devices to measure heart rate during physical education and recess in children with mild intellectual disabilities. The authors discuss that the use of Polar heart rate monitors was feasible although the wrist-worn device receiver was a distraction. In the present study, however, participants did not wear a receiver as the ActiGraph wirelessly recorded heart rate data.

Therefore, heart rate is not a feasible measure for all children with intellectual disabilities, and may be more feasible in children with milder intellectual disabilities, with the use of multi-device monitoring potentially further increasing usability. However, considering the weak criterion validity and usability issues associated with the device, heart rate does not provide a valid or consistently feasible method of measuring physical activity in children with intellectual disabilities.

6.5.1.2 Accelerometry

The vertical axis and vector magnitude counts both provide excellent criterion validity for the measurement of physical activity in children with intellectual disabilities. For vector magnitude, accelerometer counts increase with each SOFIT category Table 6.7. For the vertical axis, counts do not consistently increase with the SOFIT categories, as a lower mean counts score is recorded for standing in comparison with the preceding lying down category. This illustrates that triaxial vector magnitude counts can more accurately detect changes in posture, regardless of locomotion. However, with the inclusion of the three axes, it is generally assumed that vector magnitude will provide a more valid representation of children's activity patterns. However, this was not the case in the present study with the vertical axis counts in fact having a higher correlation with SOFIT compared to vector magnitude.

As this is the first study to validate vector magnitude counts in children with intellectual disabilities, there is limited scope for comparison with previous research. For vertical axis counts, Capio et al. (2010) validated the ActiGraph counts against a criterion measure of SOFIT in children with cerebral palsy, showing excellent criterion validity ($r = .75$, $R^2 = .56$, $p < .001$). However, as discussed in Section 6.5.1.1, this validity was also established using linear regression, which is not appropriate for the categorical SOFIT criterion measure.

In the measurement literature relating to typically developing children, the validity between the vertical axis and vector magnitude has been investigated. However, this has generally been in the form of comparing the uniaxial GT1M device to the newer GT3X/GT3X+ triaxial devices, or comparing different uniaxial and triaxial brands of accelerometer. Within these studies, high correlations have generally been recorded for triaxial accelerometers against a criterion measure, including: Tracmor ($r = .79$), Tritrac ($r = .44 -.79$), and

ActivTracer (r = .88 – .92; Eston, Rowlands, & Ingledew, 1998; Plasqui, Joosen, Kester, Goris, & Westerterp, 2005; Tanaka, Tanaka, Kawahara, & Midorikawa, 2007; Welk & Corbin, 1995). Correlations for uniaxial accelerometers, however, have generally been slightly lower: AM1764 (r = .57– .60; Janz, 1994). Furthermore, Hänggi, Phillips, and Rowlands (2013) note that the correlation between the GT1M and a criterion of $\dot{V}O_2$ varies depending the type of activity conducted. Similar to the present study, Hänggi et al. (2013) also compared the relationship between GT3X vertical axis (r = .88) and vector magnitude counts (r = .89), although this study concluded that vector magnitude provides a marginally more accurate measure of physical activity.

The findings in the present study show that the level of criterion validity for the ActiGraph wGT3X+ in children with intellectual disabilities is comparable to that established in previous studies involving typically developing children. The finding in the present study that uniaxial counts were more valid than triaxial counts was unexpected, although the difference is small. Theoretically, triaxial accelerometry should be more valid at capturing the dynamic physical activity behaviours of children in comparison to uniaxial accelerometry; therefore, the validity between numbers of axes used needs further empirical investigation (Bassett et al., 2012). Not only is this important from a measurement perspective, but also in terms of feasibility, as newer triaxial accelerometers are more expensive. However, there is currently no consensus on whether triaxial accelerometry is superior to uniaxial accelerometry, with studies reporting similar validity between these types of accelerometers (Adolph et al., 2012; Hänggi et al., 2013; Vanhelst et al., 2012). Therefore, this is an important area for future research.

In summary, the excellent criterion validity for the wGT3X+accelerometer demonstrates that both the vertical axis and vector magnitude counts can accurately detect changes in physical activity intensity in children with intellectual disabilities, which provides a strong foundation for accelerometer calibration. However, the higher validity identified for the vertical axis requires further investigation.

6.5.2 Calibration

The sedentary, moderate, and vigorous count boundaries developed (cpm) were ≤ 507, 1008–2300, and ≥ 2301 for the vertical axis and ≤ 1863, 2610–4214, and 4215 for vector magnitude, respectively. These cut points exhibit high sensitivity (80–88%) and specificity (77–85%) scores, with the accuracy of the cut points increasing with intensity (AUC = .86–.94). Similar to the validation findings, vertical axis counts produce marginally more accurate cut points compared to vector magnitude. However, these cut points are notably different from previous cut points derived in typically developing children; Table 6.9 presents the ActiGraph cut points discussed in Chapter 4 with the additional inclusion of the cut points developed in the present study.

Table 6.9. Comparison of existing ActiGraph accelerometer cut points for typically developing children with the calibrated intellectual disabilities-specific cut points

Cut points	Sedentary (cpm)	Light (cpm)	Moderate (cpm)	Vigorous (cpm)
Vertical axis				
Current study	≤ 507	508–1007	1008–2300	≥ 2301
Puyau (2002)	≤ 799	800–3199	3200–8199	≥ 8200
Treuth (2004)	≤ 100	101–2999	3000–5200	≥ 5201
Freedson (2005)	≤ 500	NA	501–4000	4001–7600
Mattocks (2007)	≤ 100	101–3580	3580–6129	≥ 6130
Evenson (2008)	≤ 100	101–2295	2296–4011	≥ 4012
Pulsford (2011)	≤ 99	100–2240	2241–3840	≥ 3841
Vanhelst (2011)	≤ 400	401–1900	1901–3918	≥ 3919
Mackintosh (2012)	≤ 372	373–2160	2161–4806	≥ 4807
Jimmy (2013)	n/a	n/a	1596–2315	≥ 2316

The following sections will discuss the calibrated cut points in relation to previous research, with discussion on possible reasons for the differences identified with previously developed cut points. Only the sedentary, moderate, and vigorous cut points will be discussed, as research is predominately focussed on the measurement of sedentary and ≥ moderate intensity activity, due to the health implications of these intensities. Furthermore, the light intensity cut points were not specifically calibrated, therefore the validity of the light intensity cut points will be discussed in Chapter 7.

6.5.2.1 Vertical axis

6.5.2.1.1 Sedentary

The sedentary cut point of ≤ 507 cpm derived within this study is towards the higher range of those previously developed in typically developing children (99–799 cpm). This cut point produces good classification accuracy (AUC = .87) and equal sensitivity and specificity (81%).

There is limited scope for the comparison of classification accuracy with previous cut points developed using regression equations and MET thresholds; however, there is scope for directly comparing the sensitivity, specificity, and AUC with previous calibration studies which used ROC curve analysis. The test statistics from previous ROC curve analysis calibration studies, as discussed in Chapter 4, are presented in Table 6.10. The existing typically developing sedentary cut points provide almost perfect accuracy for classifying sedentary behaviours, and are notably higher than the ROC curve analysis scores for the sedentary cut points calibrated in the present study.

One possible reason for the higher scores in these previous studies is the protocol used. Most of the sedentary activities used in previous studies were structured and did not occur during free-play. Participants in Evenson et al. (2008) completed three structured activities (15-minute rest, watching a DVD, and colouring books for 7 minutes each); Pulford et al. (2011) used a structured protocol where children spent 30 minutes lying down watching a DVD and 5 minutes sitting playing a computer game; and Mackintosh et al. (2012) included sedentary activities of drawing/colouring for 10 minutes. Therefore, as sedentary activity is constant, the accelerometer will record minimal counts and the criterion method will more accurately measure the activity as sedentary, i.e. direct observation will not be effected by transitions or epochs containing more than one intensity of activity. Subsequently, these types of sedentary behaviours will be easier to discriminate from physical activity in the analysis.

A limitation with these previous protocols is that the structured activities may not be fully representative of the activity behaviours conducted by children. In the present study, however, the free-living design of the protocol will better account for the sporadic nature of children's activity behaviours and the transitions from one activity intensity to another. On the other hand, as children with intellectual disabilities have been reported to spend a high proportion of their day sedentary, the use of more prolonged periods of sedentary activity may in fact be more representative of actual sedentary behaviours.

The development of a cut point for sedentary behaviours is important as there is an emerging research area which is specifically focussed on understanding and measuring sedentary behaviour as a construct independent of physical activity (Biddle et al., 2015). However, an important consideration when measuring sedentary behaviour with the ActiGraph is that this device is primarily designed for the measurement of movement, i.e. physical activity. With the increasing focus on sedentary behaviour, devices have been developed which are designed to measure posture, such as the ActivPAL (PAL Technologies, Glasgow, Scotland). That said, however, third generation ActiGraph devices also include an inclinometer which is designed to distinguish between lying down, sitting, and standing; although, this additionally requires calibration (Clemes et al., 2012). Therefore, if the focus of a study is on measuring sedentary behaviours rather than physical activity, the use of a posture-specific device or measure should be considered rather than the use of activity intensity cut points.

In summary, the sedentary cut points developed in this study exhibit good classification accuracy, although this is lower than the classification accuracy of previous calibration studies. This is likely an effect of the protocol used, although the present cut point should be more ecologically valid for capturing the free-living sporadic behaviours of children. However, for studies which aim to only measure sedentary behaviour, consideration should be given to using posture-specific devices or the ActiGraph inclinometer.

6.5.2.1.2 Moderate

The moderate cut point of 1008−2300 cpm developed within this study produces excellent classification accuracy (AUC = .92), with sensitivity and specificity of 86% and 83%, respectively. As can be seen in Table 6.10, this cut point exhibits a higher level of accuracy than the cut points developed by Evenson et al. (2008) and Pulsford et al. (2011).

The developed vertical axis cut point is lower than existing cut points. With the exception of the cut point developed by Freedson et al. (2005), the lower count boundary is between 588 and 2573 cpm lower than previous cut points. Similarly, the upper boundary is between 15 and 5899 cpm lower. Another interesting finding is the small count range between the lower and upper boundary, which is 1292 cpm. With the exception of Jimmy et al. (2013), which had a count range of 719 cpm, the cut point ranges between the upper and lower boundaries in previous studies were higher, ranging from 1599−4999 cpm. Interestingly, the boundary ranges are greater for cut points developed using regression equations (2200−4999 cpm) compared to ROC curve analysis (719−2645 cpm), suggesting that analysis method affects the derived cut points.

The lower boundary of the moderate cut point is additionally important as this also provides the cut point for moderate to vigorous intensity, which represents health-enhancing activity. Some previous research which has validated cut points suggest the use of a lower boundary cut point in the range of 3000 and 3600 cpm for moderate intensity activity (Guinhouya, Apete, & Hubert, 2009; Guinhouya, Hubert, & Zitouni, 2011). The authors of these studies criticise the use of cut points in the range of 2000 cpm, which they suggest are biased due to being calibrated using inappropriate structured activities, such as walking speeds which are too slow to be defined as moderate. However, the suggested use of 3600 cpm suggested by Guinhouya et al. (2009) was based on a classification accuracy of AUC = .64−.66, suggesting that this cut point will in fact provide only fair accuracy.

In contrast, a study by Trost et al. (2011) which compared the validity of different ActiGraph cut points suggests that the lower boundary cut point of 2296 cpm developed by Evenson et al. (2008) is most valid (moderate intensity:

sensitivity = 60%, specificity = 88.%, AUC = .74; moderate to vigorous intensity: sensitivity = 88%, specificity = 92%, AUC = .90). Furthermore, Clanchy et al. (2011) also recommend use of the Evenson et al. (2008) cut point for children with cerebral palsy. However, an interesting aspect of this study is that Clanchy et al. (2011) also calibrated a moderate to vigorous intensity cut point for children with cerebral palsy to allow comparison. The lower developed cut point of ≥ 2012 cpm supports the findings of the present study that cut points developed in typically developing children are too high for children with intellectual disabilities. Considering the moderate cut point of 1008–2300 cpm established in the present study, the use of the recommended Evenson et al. (2008) cut point would underestimate moderate and moderate to vigorous intensity activity in children with intellectual disabilities. Furthermore, the test statistics in the present study provide higher accuracy than that for the Evenson et al. (2008) cut point, suggesting that these lower cut point boundaries are more representative of the activity behaviours of children with intellectual disabilities

In summary, the moderate cut point developed in this study exhibits excellent classification accuracy. In comparison with existing typically developing cut points, this cut point is substantially lower, which has important implications as the lower boundary is also the cut point used for moderate to vigorous intensity activity. Therefore, the use of typically developing cut points in children with intellectual disabilities will generally underestimate physical activity intensity and introduce systematic error into results.

6.5.2.1.3 Vigorous

The vigorous cut point of ≥ 2301 cpm derived within this study is lower than previously developed vigorous cut points in typically developing children. However, this cut point produces excellent classification accuracy (AUC = .94), with sensitivity and specificity of 88% and 85%, respectively, indicting it is accurate in children with intellectual disabilities.

The vigorous cut point developed in the present study exhibits a higher level of accuracy than the Evenson et al. (2008) cut point, although less accuracy than the Pulsford et al. (2011), Mackintosh et al. (2012), and Jimmy et al. (2013) cut points. This cut point is also lower than those developed in previous studies and is actually lower than some existing moderate intensity cut points, i.e. the lower boundary for the moderate cut points developed by Puyau et al. (2002), Treuth et al. (2004), and Mattocks et al. (2007) are > 2301 cpm. Therefore, as this vigorous cut point is more similar to the moderate cut points established in typically developing children, the use of existing vigorous intensity cut points will underestimate the time children with intellectual disabilities spend active at this intensity, introducing a high level of systematic error into results.

The vigorous cut points developed in previous studies (Table 6.10) have generally exhibited lower accuracy than sedentary and moderate intensity cut points. This has been at least partially attributed to the wider range of activity behaviours children exhibit at a vigorous intensity, such as skipping and dodging, which are theoretically not as accurately captured by vertical axis accelerometry (Mackintosh et al., 2012). Furthermore, bouts of vigorous intensity activity conducted by children are often short, therefore the measurement epochs used to calibrate vigorous intensity are not based only on vigorous activity (Shields et al., 2009; Whitt-Glover et al., 2006). In the present study, however, the vigorous cut point has the highest classification accuracy, in comparison with the sedentary and moderate cut points. Furthermore, unlike Evenson et al. (2008), the high and similar sensitivity and specificity scores indicate that this cut point will equally limit the likelihood of type I and type II errors. Considering the field-based protocol, this suggests that free-living physical activity conducted at a vigorous intensity will be accurately classified using this cut point.

In summary, the calibrated vigorous intensity cut point provides excellent classification accuracy. However, a large difference in comparison with existing cut points was identified, as this cut point is in fact lower than some moderate intensity cut points calibrated in previous studies. Therefore, this further highlights the need for up-to-date - in terms of analysis, protocol, and device - and population-specific cut points for children with intellectual disabilities.

Table 6.10. ROC curve statistics established in previous calibration studies involving typically developing children

ROC curve statistics	Cut points				
	Evenson (2008)	Pulsford (2011)	Mackintosh (2012)	Jimmy (2013)*	Present study
Sedentary					
Sensitivity (%)	99	95	99	-	81
Specificity (%)	97	93	97	-	81
AUC	.995	.98	.995	-	.87
Moderate					
Sensitivity (%)	77	60	97	-	86
Specificity (%)	81	76	97	-	83
AUC	.85	.60	.99	-	.92
Vigorous					
Sensitivity (%)	68	95	89	90	88
Specificity (%)	89	91	96	86	85
AUC	.83	.98	.98	.94	.94

* Jimmy et al. (2013) did not establish a sedentary cut points and did not present the ROC curve statistics for the moderate cut point

6.5.2.2 Vector magnitude

The previous sections in this chapter have focussed on the vertical axis cut points. This is primarily because the calibration of ActiGraph vector magnitude counts for children is in its infancy, therefore there is little scope for comparison with previous research.

An interesting aspect of this study, however, is the increased accuracy of vertical axis cut points over vector magnitude. As discussed in Section 6.5.1, this does not seem to fit with the theory that the inclusion of three axes will better capture dynamic activity, yet there is currently no consensus within the literature on which is most valid. Similar to the present study, Jimmy et al. (2013) compared the accuracy between cut points derived using GT3X vertical axis and vector magnitude counts, against a criterion measure of energy expenditure. This study reports that vector magnitude counts provide higher classification accuracy for sedentary behaviours and vigorous activity, although vertical axis cut points were more accurate for moderate intensity activity. However, a limitation with this study is that cut points are based on MET thresholds which, as previously discussed, has multiple limitations in children. Therefore, this is an area which requires further investigation.

6.5.3 Factors affecting calibration

The cut points in this study are markedly different from those established in typically developing children. Within the previous sections, the effect of the activity protocol has been discussed as a possible factor impacting on the cut points derived between studies. However, there are additional factors that need to be considered. Therefore, the following sections will discuss the design of this study and other potential causes of error.

6.5.3.1 SOFIT

As the criterion measure, the validity of the developed cut points is dependent on the accurate use of SOFIT. In the present study, every effort was made to ensure that activity coding was valid and reliable. An advantage of the SOFIT tool is that there is numerous resources and training materials available from the authors (McKenzie, 2009). Prior to data collection, the lead rater and two reliability raters completed the recommended training, achieving 86% validity against the gold standard assessment video, 79% inter-rater reliability prior to data collection, and 85% at the mid-point of data collection. Although 80% inter-rater reliability is recommended prior to commencing data collection, this was marginally not achieved. In line with the SOFIT developer's recommendations for not achieving 80% reliability, the raters in the present study discussed the discrepancies.

From these discussions, a cause of error with recording was identified in relation to when to record activity, i.e. is activity coded at the start of the record prompt or after the record prompt. Although this seems a minor issue, in practice activity often changed in the 1-second from the start to the end of the record prompt. It was subsequently agreed that data would be coded at the start of the record prompt. The raters also discussed whether the attainment of 80% reliability was achievable in children with intellectual disabilities, as participants exhibited atypical behaviours which were difficult to classify within the SOFIT categories. However, when Faison-Hodge and Porretta (2004) used SOFIT to code the physical activity of children with intellectual disabilities, they increased the reliability standard to 90%, which was achieved.

Although SOFIT is considered an objective measure, there is an element of subjectivity within the coding of activity. This is most apparent in the coding of walking, i.e. "ordinary walking", as what is deemed "ordinary" will vary between children. The coding of ordinary walking could have additional implications in children with intellectual disabilities. As discussed within Chapter 5, children with intellectual disabilities walk at slower speeds than typically developing children; therefore, a walking speed deemed ordinary for typically developing children could in fact be fast walking for children with intellectual disabilities. In contrast, if children are walking at a slow or light intensity pace, which is physiologically not of a moderate intensity, according to the SOFIT guidelines, this is still coded as walking, i.e. moderate intensity. This could therefore be one possible reason as to why the boundaries of the moderate intensity cut point are lower than that of previous studies.

Another possible cause of systematic error with the SOFIT coding system is the intensity classification of standing. There is a lack of consensus in the literature on whether standing should be coded as sedentary or light intensity activity. Specific to the ActiGraph, De Decker et al. (2013) compared the accuracy of a < 100 cpm cut point when standing was classed as sedentary, compared to when standing was classified as not sedentary. The classification of standing as sedentary, using a cut point of < 100 cpm for the GT1M device, produced a higher level of accuracy (sensitivity = 46.30%, specificity = 75.80%, AUC = .61) compared to classifying standing as not sedentary (sensitivity = 58.50%, specificity = 61.16%, AUC = .59). In this study by De Decker et al. (2013), standing specifically referred to standing still. In the present study, however, the standing category includes movement, e.g. upper body movement; therefore, the classification of standing as light intensity may be more appropriate.

6.5.3.2 Behavioural characteristics of children with intellectual disabilities

As SOFIT is a behavioural measure, calibration is based on observed movements rather than physiological outcomes. Therefore, if these cut points are shown to provide a valid method of interpreting accelerometer output in children with intellectual disabilities, then at least part of the discrepancy between these cut points and existing cut points can be attributed to behavioural and movement differences between children with intellectual disabilities and typically developing children.

Previous research which has suggested that typically developing cut points are too high for children with intellectual disabilities have generally hypothesised that physiological differences, such as levels of cardiorespiratory fitness, will limit the generalisability of cut points to children with intellectual disabilities (Frey et al., 2008). However, as this study was not based on a physiological criterion measure, the present findings suggest that there are additional behavioural and biomechanical differences between these population groups, which affect the calibration of cut points. Similarly, children with abnormal gait patterns have been reported to have lower movement economy than children with normal gait patterns, but again these findings are based on physiological measures (Johnston et al., 2004; Thomas et al., 2011).

The present findings suggest that during the same biomechanical movements, specifically those included in SOFIT – such as walking-, children with intellectual disabilities produce a smaller acceleration, as measured by the ActiGraph wGT3X+. Additionally, as vertical axis counts provide more accurate cut points, it could be hypothesised that children with intellectual disabilities produce extraneous movement on the other two axes which effects the accuracy of vector magnitude counts. As this is the first study to hypothesis these differences, further research needs to be conducted.

6.5.3.3 ROC curve analysis

As discussed in Chapter 4, ROC curve analysis is the emerging statistical method for accelerometer calibration and could provide increased classification accuracy over previous regression-based methods (Jago, Zakeri, Baranowski, & Watson, 2007; Rothney et al., 2008). The strengths and limitations of this method were also discussed in Chapter 4, however, some of the decisions involved with the use of ROC curves could impact on the developed cut points. ROC curve analysis is traditionally used in medical research where a cut point is developed to determine whether a patient does or does not have the condition of interest. Therefore, depending on the severity of the condition, sensitivity will generally be weighed as more important than specificity to reduce the likelihood of false-negative results.

As ROC curve analysis is in its relative infancy for calibration, there is limited consensus on whether sensitivity or specificity is more important. Therefore, the cut point in which both sensitivity and specificity are optimized is generally used. However, weighing one over the other, even slightly, will alter the cut point boundaries. Mackintosh et al. (2012) investigated the classification accuracy of the moderate to vigorous intensity cut point (> 2160 cpm) developed in their study when sensitivity and specificity were altered. Table 6.11 presents the findings reported in Mackintosh et al. (2012) and shows that small changes in specificity and, in particular, sensitivity can result in markedly different cut point boundaries. However, altering the lower count boundary by ± 200 cpm will have little or no effect on the overall classification accuracy. However, from the data presented in Table 6.11, it would appear that 2070 cpm is in fact the optimal moderate to vigorous intensity cut point.

Table 6.11. Effect of altering sensitivity and specificity on derived cut points

Sensitivity (%)	Specificity (%)	Kappa	Total agreement (%)	Counts/min
94	75	.71	89.00	2160*
95	72	.70	88.30	2250
94	78	.72	89.70	2070
96	70	.70	88.10	2360
93	79	.71	89.60	1960

Note: adapted from Mackintosh et al. (2012)
* Derived cut point when sensitivity and specificity are optimised

6.5.3.4 ActiGraph device

The cut points calibrated in the present study were notably different from existing cut points, with the exception of the cut points derived by Jimmy et al. (2013). The sedentary cut points derived in the present study were higher and the moderate and vigorous cut points lower than those previously established in typically developing children. As discussed in Section 4.2.1.2, there are many differences between versions of the ActiGraph, which can limit comparability. Specifically, third generation devices record lower count values for the same acceleration in comparison with older devices (Ried-Larsen et al., 2012). Furthermore, the GT3X records higher counts for the same acceleration than the GT3X+ device (Robusto & Trost, 2012). This is concurrent with the findings in the present study which suggests that cut points derived using third generation devices will be lower than older versions, therefore the device used will affect the cut points calibrated.

As the cut points derived by Jimmy et al. (2013) were the only cut points calibrated using a third generation device (GT3X), there is limited scope for discussion on the effect of device within this study. That said, however, the cut points calibrated by Jimmy et al. (2013) were most similar to the cut points derived in the present study. Furthermore, considering the GT3X records higher count values than the GT3X+, the higher cut points calibrated by Jimmy et al. (2013) is concurrent with previous research. Therefore, this suggests that the ActiGraph device used will affect calibration, which could at least partially account for the large discrepancies between existing cut points and those established in this study.

The effect of device not only has important implications for measurement research specific to children with intellectual disabilities, but also highlights the need for additional calibration research to be conducted in typically developing children to limit measurement error from generalising cut points between devices. As the cut points calibrated by Jimmy et al. (2013) were the only identified cut points established using a third generation device, is it important that future calibration research is conducted in both children with intellectual disabilities and typically developing children to investigate the effect of device.

6.5.4 Strengths and limitations

This was the first study to calibrate accelerometer cut points for physical activity intensity in children with intellectual disabilities. Based on the findings reported in previous chapters of this thesis, a protocol was developed which included feasible methods and activity protocol. The free-living design of this protocol enabled calibration to be conducted on activities which were representative of children's activity behaviours, increasing ecological validity. With studies calibrating ActiGraph vector magnitude only emerging in the last couple of years, this is the first study which has placed measurement research in children with intellectual disabilities in line with the emerging measurement research seen in typically developing children. Furthermore, even though participation rates are generally low in health-related research in children with intellectual disabilities, the high recruitment rate within this study is similar to that reported in previous research involving typically developing children.

However, this study is not without limitations. Although the free-living protocol was a strength of this study, the limited structure of the protocol resulted in the characteristics of each session, e.g. session duration and time spent in each intensity, varying between sessions (Table 6.6). Specifically, sessions which included participants with more complex needs were generally shorter and included less time in higher intensity activity. Therefore, the data collected may not be fully representative of the study sample as a whole, but instead children with milder intellectual disabilities.

6.5.5 Conclusions

This study was the first to calibrate population-specific accelerometer cut points for the estimation of physical activity intensity in children with intellectual disabilities, thus addressing a substantial gap in measurement research relating to children with intellectual disabilities. The cut points developed in this study show high sensitivity and specificity for the estimation of physical activity intensity in children with intellectual disabilities. Furthermore, accelerometry provides a more valid and feasible method to measure physical activity in comparison to heart rate. However, it is important that the developed cut points are cross-validated in a different sample of children with intellectual disabilities. With the emerging trend in this study of lower cut points being derived in comparison with typically developing children, the need to cross-validate these cut points is of vital importance. Possible causes for these differences relating to study design have been discussed. However, moving forward, it is important to consider whether future research may benefit from taking an additional step back to basics. That is, increase the knowledge base relating to the biomechanics of how children with intellectual disabilities move. This will help our understanding of population-specific factors which may have influenced calibration and will help inform the next phases of improving the validity of objectively measured physical activity in children with intellectual disabilities.

Chapter 7 – Cross-validation of the cut points calibrated using the wGT3X+ in children with intellectual disabilities

7.1 Overview of this chapter

Chapter 6 reports the calibration of the first population-specific accelerometer cut points for children with intellectual disabilities. These cut points exhibit good to excellent classification accuracy for the estimation of sedentary behaviour and physical activity intensity. However, it is important to cross-validation cut points in a different sample of the same population group to allow further examination of validity. Therefore, this chapter will examine the validity of the developed cut points in a sub-sample of children with intellectual disabilities. Furthermore, this chapter will also test the validity of existing cut points, as discussed in Chapter 4, against a criterion measure of SOFIT. The validity of the developed cut points and existing cut points will be discussed, and recommendations made for the most valid cut points for use in children with intellectual disabilities.

7.2 Introduction

The calibration of accelerometer cut points for children with intellectual disabilities is the first stage in improving the validity of interpreting accelerometer output and increasing our understanding of accelerometer measurement in this population group. However, calibration findings are specific to the original study sample and protocol and, therefore, require further validation (Heil, Brage, & Rothney, 2012). Furthermore, validity cannot be fully established and understood in one stand-alone study, therefore multiple studies are required to establish validity (Bassett et al., 2012). The next step in the validation process is cross-validation.

7.2.1 Cross-validation

Cross-validation is when the prediction accuracy of an instrument or method is assessed against a criterion measure, and is generally conducted in a sample which was not included in the initial calibration. Specifically, cross-validation investigates the probability that the score a participant receives on the instrument being validated will be the same as that measured by the criterion measure, i.e. the probability that the intensity classification based on the developed cut points is the same as the SOFIT intensity classification.

Ideally, cross-validation should be conducted in a free-living environment, using activities which are similar to, but different, from the activities used in the original calibration study (Welk, 2005). This will increase our understanding of classification accuracy of cut points over a wide range of representative activities and movements. There are two primary methods in which cross-validation can be investigated: the “split sample” or “leave-one-out” approach (Staudenmayer, Zhu, & Catellier, 2012). The split sample method involves splitting the sample into separate groups; one for calibration and one for cross-validation. An advantage of this method is that cross-validation is conducted on a sample not included in the original analysis. Furthermore, this will more accurately replicate the generalisation of cut points to the wider population of children with intellectual disabilities. However, this method will reduce the statistical power of the calibration analysis. The leave-one-out approach is more complicated and is generally used for the cross-validation of regression equations. In this approach, a regression equation is developed on all but one of the study sample and its validity examined on the “held-out” participant. This process is repeated until all participants have been the “held-out” participant, with the mean of this evaluation analysis reported.

No studies were identified which used the leave one-out approach for the cross-validation of intensity cut points in children. Instead cross-validation has mostly been conducted in a sub-sample of the total recruited study sample (Jimmy et al., 2013; Vanhelst et al., 2011). Another design which has been used for cross-validation is to use a split-protocol approach, where only some activities are used for calibration and the remaining activities used for cross-validation, with the same sample taking part in both parts of the session. Mackintosh et al. (2012) and Welk, Dale, and Schaben (2002) used this approach where accelerometer output was calibrated during a structured protocol and cross-validated during a free-play session. Although this method allows calibration to be conducted on more ecologically valid activities which are different from the calibration activities, it does not allow investigation into the generalisability of the cut points to the wider population. In addition, Pulsford et al. (2011) utilised a unique approach where the validity was established using the same data set that was used for calibration, but using ROC curve analysis to test the classification accuracy.

As a fundamental aspect of cross-validation is investigation into the generalisability of cut points in an independent sample, the split-protocol approach and the statistical approach used by Vanhelst et al. (2011) do not allow the cross-validation of cut points to be investigated. Therefore, to increase the accuracy of the cross-validation analyses, it is important to use activities which vary from the calibration analysis and conduct the cross-validation in a different sample of the same population.

7.2.2 Avoiding the ecological fallacy

As discussed in previous chapters, particularly in relation to the design of a calibration protocol, ecological validity has been an important consideration of study design. Ecological validity in the context of a calibration protocol refers to whether the activities conducted in these small-scale studies are representative of the activities conducted by the population of children with intellectual disabilities in the real world. Including an ecologically valid activity protocol will increase the likelihood that the relationship between counts and activity behaviours, as measured by SOFIT, will be the same in a real-world setting.

A related area for consideration is how applicable this group calibration data is to individuals. The ecological fallacy refers to a deduction-related error where inferences are made on an individual level based on the analysis of group level data (O'Dowd, 2003). Specifically, the ecological fallacy is to assume that a relationship which is observed at a group level, i.e. the relationship between accelerometer counts and SOFIT, will be the same on an individual level. Although this fallacy is generally more related to ecological research, its underlying principles are relevant to accelerometer calibration and understanding the difficulties associated with this area of research.

As discussed in Chapter 1, one method to avoid the ecological fallacy would be to calibrate accelerometry at an individual level. However, in practical terms, this is generally not feasible for larger-scale studies. Therefore, conducting cross-validation on the developed cut points will, to an extent, provide initial empirical evidence regarding the inferences made from these cut points, which will help identify if an ecological fallacy is present.

7.2.3 Need for the developed cut points

The literature is currently saturated with cut points, with nine existing sets of ActiGraph-specific cut points available (Mackintosh et al., 2012; Welk, 2005). This leaves researchers with the “cut point conundrum” of deciding upon which cut points are most valid for their study sample (Trost, 2007a; Trost et al., 2006). Furthermore, another limitation with the large number of exiting cut points is that it limits the scope of comparison and, as highlighted in Chapter 4, can result in significant differences in the intensity estimations based on the chosen cut points.

To prevent further saturation of the literature, investigating the validity of newly developed cut points in comparison with existing cut points is an important aspect of calibration research (Bassett et al., 2012). As a result, if newly developed cut points do not provide a greater level of validity than existing cut points, researchers should recommend and support the use of an existing set of cut points (Bassett et al., 2012; Welk, 2005). Welk (2005) highlighted this point well and discussed the need to develop a standard of care within accelerometer calibration and validation research whereby researchers need to demonstrate the advantages of newly derived methods compared with existing methods. This will create a more standardised approach where the burden is on the original researchers themselves to provide initial evidence that their cut points are more valid than existing cut points.

7.2.4 Research questions

In accordance with recommendations from experts in the field of measurement research, the purpose of this chapter - and this thesis as a whole - is not to further saturate the literature with cut points (Bassett et al., 2012; Welk, 2005).

Instead, the aim of this chapter is to address the gap in the literature for population-specific cut points and investigate validity in both a sub-sample of children with intellectual disabilities and against existing ActiGraph cut points. The specific research questions to be addressed in this chapter are:

RQ 17: Do the developed vertical axis cut points provide a valid estimation of physical activity intensity in a sub-sample of children with intellectual disabilities?

RQ 18: Do the developed vector magnitude cut points provide a valid estimation of physical activity intensity in a sub-sample of children with intellectual disabilities?

RQ 19: Do the developed vertical axis cut points provide a more valid estimation of physical activity intensity in children with intellectual disabilities than existing cut points?

7.3 Method

7.3.1 Ethical considerations

This study was approved by the Medical, Veterinary, and Life Sciences College Ethics Committee, University of Glasgow (Appendix iv). Written informed consent was required from both participants and parents prior to participation. Verbal consent was additionally sought from participants prior to each activity session.

7.3.2 Participants and methods

Participants for this study were recruited as per the procedures reported in the calibration chapter of this thesis (Chapter 6). All participants took part in the school-based activity sessions, as previously described in Section 6.3.3. Prior to each session, two coloured bibs were randomly selected, with the corresponding participant subsequently assigned to the cross-validation group. This resulted in 14 participants in the cross-validation group, which represented 39% of the number of participants in the calibration group. These cross-validation procedures are in accordance with expert recommendations, which suggest cut points are cross-validated in an independent and representative sample during field-based activity (Welk, 2005). Furthermore, randomly selecting two participants from each session will limit the effect of between-session differences.

7.3.3 Management of data

Accelerometer data were downloaded in 10-second epochs using ActiLife version 6.11.5 software at a sampling rate of 30 Hz (ActiGraph LLC, Pensacola). Data were then transformed into 20-second epochs of data to allow accelerometer epochs to be time matched with SOFIT epochs. For research question 19, SOFIT data and vertical axis counts were converted into 60-second epochs to enable comparison with existing cut points, which are in the format of counts per minute. Detailed data management procedures for accelerometer and SOFIT data are presented in Section 6.3.5.

For all validation analyses, data were reclassified into positive binary codes (intensity of interest) and negative binary codes (not intensity of interest) for sedentary, light, moderate, vigorous, and moderate to vigorous intensity. As with the calibration analysis, SOFIT scores of 1 and 2 were recorded as sedentary, code 3 as light, code 4 as moderate, code 5 at vigorous, and codes 4 and 5 as moderate to vigorous. Accelerometer data were converted into binary codes based on cut points, e.g. for the cross-validation analysis for the sedentary cut point, ≤ 169 counts/20-seconds were coded as positive (binary code 1), with > 169 counts/20-seconds coded as negative (binary code 0).

7.3.4 Statistical analysis

All statistical analyses were conducted using SPSS 22 IBM statistical package (SPSS IBM, New York, NY, USA). Descriptive statistics (mean ± SD) were calculated for participant variables (age, sex, height, weight, and BMI) and session variables (counts/20-seconds recorded for each SOFIT category). In addition, independent samples t-tests were conducted to test for significant ($p < .05$) differences in participant and session variables between participants in the calibration and cross-validation groups.

To examine the validity of the developed cut points, classification agreement was investigated between the criterion measure of SOFIT and the calibrated cut points (counts/20-seconds) using sensitivity, specificity, total agreement percentages, and Cohen's kappa scores. Sensitivity was calculated as the percentage of positive cut point epochs which corresponded with a positive SOFIT epoch, i.e. the cut point correctly classified a positive score as positive (true positive). Specificity represented the percentage of negative cut point epochs that corresponded with a negative SOFIT epoch, i.e. the cut point correctly classified a negative score as negative (true negative). Total agreement was calculated as the overall percentage of positive and negative cut point epochs which agreed with the corresponding SOFIT epoch.

Cohen's kappa scores were further used to investigate classification agreement. An advantage of this method is that it accounts for agreements which may occur by chance, unlike total agreement scores. However, this results in kappa scores being a conservative measure of agreement and therefore should be interpreted with caution. Kappa scores were calculated to test the agreement between SOFIT and cut point epochs for each intensity, as described above. In addition to testing the agreement using binary data, kappa scores were also used to test the agreement for total activity, i.e. the inclusion of sedentary, light, moderate, and vigorous cut points. The interpretation of kappa scores is somewhat arbitrary and varies between subject areas; however, for the purposes of this study the following scale will be used as a guide to interpret the kappa statistic (κ): $< .00$ as less than change agreement, .00−.20 as slight agreement, .21−.40 as fair agreement, .41−.60 as moderate agreement, .61−.80 as substantial agreement, and .81−1.00 as almost perfect agreement (Landis & Koch, 1977; Viera & Garrett, 2005).

7.4 Results

7.4.1 Participants

Fourteen children with intellectual disabilities participated in this study. Participant characteristics are presented in Table 7.1. There were no significant differences in age ($t = -.13$, $df = 48$, $p > .01$, 95% CI -.74, .65) height ($t = .08$, $df = 48$, $p > .01$, 95% CI -.05, .06), weight ($t = -.06$, $df = 48$, $p > .01$, 95% CI -6.78, 6.38), or BMI ($t = -.32$, $df = 48$, $p > .05$, 95% CI -2.81, 2.05) between the participants in the calibration and cross-validation groups, suggesting this sample is representative of the calibration group.

Table 7.1. Descriptive statistics (mean ± SD) of cross-validation participants

Characteristic	Boys (n = 9)	Girls (n = 6)	Total (n = 14)
Age (yrs)	9.44 ± 1.13	9.80 ± 1.30	9.57 ± 1.16
Height (m)	1.45 ± .07	1.39 ± .07	1.43 ± .08
Weight (kg)	39.84 ± 6.92	38.82 ± 6.98	39.48 ± 6.69
BMI (kg/m^2)	18.97 ± 2.69	20.09 ± 3.18	19.37 ± 2.81

Mean (± SD) accelerometer counts recorded for each SOFIT category and statistics for the comparison of these scores with those recorded by the calibration group are presented in Table 7.2. With the exception of the light (vector magnitude) and moderate to vigorous (vertical axis) intensity categories, participants in this study recorded significantly different, and mostly lower, mean counts and standard deviations than participants in the calibration group.

Table 7.2. Mean (± SD) counts for each SOFIT category and comparison with calibration group

Intensity	Vertical axis counts (counts/20-seconds)					Vector magnitude counts (counts/20-seconds)				
	Calibration Mean (SD)	Cross-val Mean (SD)	*t*	df	95% CI	Calibration Mean (SD)	Cross-val Mean (SD)	*t*	df	95% CI
Sedentary	170.66 (241.86)	55.48 (85.92)	11.80**	1124.28	96.02, 134.34	516.20 (478.19)	331.48 (325.79)	7.37**	784.68	135.51, 233.94
Light	256.19 (335.57)	172.57 (183.46)	4.93**	778.53	50.30, 116.93	642.80 (497.91)	669.18 (383.57)	-.87	542.69	-86.28, 33.51
Moderate	625.06 (437.32)	552.10 (245.10)	3.63**	945.79	33.55, 112.37	1281.20 (573.79)	1197.32 (360.27)	3.00**	842.02	29.06, 138.71
Vigorous	1294.13 (612.15)	1387.26 (554.51)	-2.26*	979.00	-174.06, - 12.20	2008.50 (689.90)	2205.62 (677.38)	- 4.15**	979.00	-290.42, -103.82
MVPA	925.33 (619.92)	963.40 (601.14)	-1.29	2143.00	-95.94, 19.80	1602.20 (728.44)	1690.08 (748.11)	-2.49*	2143.00	-156.98, -18.77

* p < .05, **p < .01

MVPA; moderate to vigorous intensity physical activity

7.4.2 Research question 17

Do the developed vector magnitude cut points provide a valid estimation of physical activity intensity in a sub-sample of children with intellectual disabilities?

The total classification agreement, sensitivity, specificity, and kappa scores for vector magnitude for the cross-validation group are presented in Table 7.3. The high total agreement and substantial kappa scores for sedentary, vigorous, and moderate to vigorous intensity suggest these cut points are accurate for classifying physical activity intensity in children with intellectual disabilities. This is confirmed by the high sensitivity (82.21%, 89.33%, and 91.27%) and specificity (86.10%, 89.46%, and 83.55%) scores for sedentary, vigorous, and moderate to vigorous intensity, respectively. Therefore, this further demonstrates that these cut points have a high probability of correctly classifying activity intensity and limiting the probability of misclassification.

The total agreement scores for the light and moderate intensity cut points suggest good agreement; however, this is not confirmed with the kappa scores. For the light intensity cut point, the kappa agreement is slight ($\kappa = .20$), although still significant at $p < .001$. The sensitivity result shows only 24.80% of light intensity activity being correctly recorded as light, although the specificity is high (92.84%), which illustrates that sensitivity and specificity are not optimised at a cut point of 622−869 cpm. Similarly, the moderate intensity cut point has good kappa agreement ($\kappa = .51$, $p < .001$) and high specificity (89.42%), but sensitivity agreement of only 59.74% for correctly classifying moderate intensity activity.

These cross-validation results suggest that the sedentary, vigorous and moderate to vigorous intensity vector magnitude cut points provide a valid classification of activity intensity in children with intellectual disabilities. However, the moderate intensity cut point should be used with caution and, due to limited validity, use of the light intensity cut point is not recommended.

Table 7.3. Cross-validation sensitivity, specificity, total agreement, and kappa statistics

Cut point	Total agreement (%)	Sensitivity (%)	Specificity (%)	Kappa (SE)
Vertical axis				
Sedentary	85.18	92.62	83.06	.66* (.02)
Light	80.68	32.93	93.72	.32* (.04)
Moderate	90.29	74.92	95.65	.74* (.02)
Vigorous	94.19	93.00	94.61	.85* (.02)
MVPA	92.63	90.94	95.06	.85* (.02)
Total activity				.79* (.01)
Vector magnitude				
Sedentary	84.92	82.21	86.10	.63* (.03)
Light	77.99	24.80	92.84	.20* (.03)
Moderate	79.03	59.74	89.42	.51* (.03)
Vigorous	89.51	89.33	89.46	.74* (.02)
MVPA	87.35	91.27	83.55	.75* (.02)
Total activity				.72* (.01)

* Significant at $p < .001$

Note: total agreement refers to the classification agreement with SOFIT when the sedentary, light, moderate, and vigorous cut points are simultaneously applied to the data; all other kappa score were calculated using data in binary format.

7.4.3 Research question 18

Do the developed vertical axis cut points provide a valid estimation of physical activity intensity in a sub-sample of children with intellectual disabilities?

The total classification agreement, sensitivity, specificity, and kappa scores for the vertical axis for the cross-validation group are presented in Table 7.3. The vigorous and moderate to vigorous intensity cut points had almost perfect agreement ($\kappa = .85$, $p < .001$) with sensitivity scores of 93.00% and 90.94%, respectively, and specificity scores of 94.61% and 95.06%, respectively, confirming that these cut points are valid. The sedentary and moderate intensity cut points had kappa scores at the higher boundary of substantial ($\kappa = .66$ and .74, respectively), with sensitivity and specificity scores further confirming the accuracy of these cut points. The light intensity cut point shows moderate agreement with the criterion measure ($\kappa = .32$). Similar to the vector magnitude cut points, sensitivity and was disproportionately low (32.93%) in comparison to specificity (93.72%), suggesting that there is a high probability that time spent in light intensity activity will be underestimated using this cut point.

In comparison with the results for vector magnitude cut points, similar trends of classification agreement are present. The vigorous and moderate to vigorous cut points represent the highest classification agreement, with similar low sensitivity and high specificity scores for the light intensity cut points. The kappa agreement for total activity is higher for the vertical axis ($\kappa = .79$) than vector magnitude ($\kappa = .72$), suggesting that the vertical axis cut points are more accurate for the overall classification of activity when all developed cut points are simultaneously applied to the data.

In summary, the sedentary, moderate, vigorous, and moderate to vigorous intensity cut points for the vertical axis provide valid classification of physical activity intensity. The use of the light intensity cut point is not recommended. In addition, the vertical axis cut points for all intensities provide more valid classifications of intensity than the vector magnitude cut points.

7.4.4 Research question 19

Do the developed vertical axis cut points provide a more valid estimation of physical activity intensity in children with intellectual disabilities than existing cut points?

When the validity of existing cut points was investigated against the criterion measure of SOFIT, notable differences in accuracy were found. For all intensities, only one of the 35 existing cut points demonstrated a higher level of validity than the cut points developed in this thesis (current cut points). Full validation statistics are presented in Table 7.4.

For sedentary intensity, the current cut points exhibit higher overall classification accuracy ($\kappa = .66$) than existing cut points, which range from $\kappa = .55–.64$, with the exception of Freedson et al. (2005; $\kappa = .67$). All existing cut points exhibit high sensitivity and specificity scores, although for most cut points, sensitivity is higher than specificity. Therefore, existing sedentary cut points will provide comparable, although slightly lower, validity to the current cut point.

For moderate intensity, the existing cut points generally show very low classification accuracy, with six sets of cut points exhibiting accuracy which is less than chance ($\kappa < .00$). The cut points developed by Freedson et al. (2005) show the highest level of classification accuracy of existing cut points, with sensitivity of 94.06% and specificity of 66.28%. This suggests that the Freedson et al. (2005) moderate cut point will accurately classify 94.06% of moderate intensity activity, but will have a higher probability of false-positive scores. In contrast, the remaining existing cut points show very low sensitivity 0.33–33.33% and disproportionally high specificity (75.44–98.12%), which illustrates that these cut points will not be accurate for classifying moderate intensity activity in children with intellectual disabilities.

Table 7.4. Sensitivity, specificity, total agreement, and kappa statistics for the validation of existing cut points

Study	Puyau (2002)	Treuth (2004)	Freedson (2005)	Mattocks (2007)	Evenson (2008)	Pulsford (2011)	Vanhelst (2011)	Mackintosh (2012)	Jimmy (2013)	Current study
Sedentary										
Sensitivity (%)	98.01	81.54	91.69	81.54	81.54	81.54	85.38	85.27	n/a	85.18
Specificity (%)	76.55	92.29	83.24	92.29	92.29	92.29	84.88	83.27	n/a	92.62
Total agreement (%)	86.14	83.88	85.44	83.88	83.88	83.88	84.92	85.93	n/a	83.06
Kappa (± SE)	.61 ± .02	.55 ± .03	.67 ± .02	.55 ± .03	.55 ± .03	.55 ± .03	.64 ± .02	.64 ± .03	n/a	.66 ± .02
Moderate										
Sensitivity (%)	1.98	2.97	94.06	0.33	10.56	13.20	31.68	19.15	33.33	74.92
Specificity (%)	75.44	80.73	66.28	82.61	81.14	85.19	83.43	75.44	98.12	95.65
Total agreement (%)	56.24	60.31	74.18	61.27	65.08	66.29	69.84	60.66	81.11	90.29
Kappa (± SE)	-.25 ± .02	-.19 ± .02	.48 ± .02	-.19 ± .02	-.05 ± .03	-.02 ± .03	.16 ± .03	-.06 ± .03	.39 ± .03	.74 ± .02

Table 7.4. Continued

Study	Puyau (2002)	Treuth (2004)	Freedson (2005)	Mattocks (2007)	Evenson (2008)	Pulsford (2011)	Vanhelst (2011)	Mackintosh (2012)	Jimmy (2013)	Current study
Vigorous										
Sensitivity (%)	1.33	22.00	47.00	12.00	49.67	53.00	50.67	27.33	92.33	93.00
Specificity (%)	100.00	99.88	99.30	100.00	99.30	99.18	99.30	99.77	94.61	94.61
Total agreement (%)	74.35	79.64	85.70	77.12	86.40	87.18	86.66	80.94	94.02	94.19
Kappa (± SE)	.02 ± .01	.29 ± .03	.58 ± .03	.17 ± .03	.58 ± .03	.61 ± .03	.59 ± .03	.35 ± .03	.85 ± .02	.85 ± .02
Moderate to vigorous										
Sensitivity (%)	35.91	39.21	96.71	30.81	52.72	52.72	63.92	57.00	70.68	90.94
Specificity (%)	99.63	99.63	74.59	99.82	99.09	99.09	98.72	99.09	97.99	95.06
Total agreement (%)	66.12	67.85	86.22	63.52	74.70	74.70	80.42	76.95	83.62	92.63
Kappa (± SE)	.34 ± .02	.38 ± .02	.74 ± .02	.30 ± .02	.51 ± .02	.51 ± .02	.62 ± .02	.55 ± .02	.68 ± .02	.85 ± .02

For vigorous intensity, agreement between existing cut points and SOFIT ranged from κ = .02−.85, with accuracy decreasing in cut points which had a higher count boundary. Furthermore, existing cut points demonstrated very high specificity scores (94.61−100.00%) but low sensitivity scores (1.33−53.00%), with the exception of Jimmy et al. (2013; 92.33%). This suggests that the lower count boundaries are high enough to almost perfectly prevent false-positive scores but are too high to detect most vigorous intensity activity. The cut points developed by Jimmy et al. (2013) produce similar sensitivity and equivalent specificity and kappa scores (κ = .85) to the current cut points. Therefore, only the use of the cut points developed by Jimmy et al. (2013) will provide accurate estimations of vigorous intensity activity in children with intellectual disabilities, producing comparable levels of validity to the current cut points.

For moderate to vigorous intensity, classification accuracy ranged from κ = .30−.74, which is lower in comparison to the current cut point (κ = .85). Similar trends were found to the moderate intensity cut points, as specificity was generally almost perfect (74.59−99.82%) but at the detriment of sensitivity (30.81−96.71). Furthermore, the lower Freedson et al. (2005) and Jimmy et al. (2013) cut points provided the best balance of sensitivity and specificity. However, as sensitivity and specificity were both almost perfect for the current cut points (90.94% and 95.06%, respectively), these cut points provide notably higher accuracy than existing cut points.

In summary, for all intensities, the current cut points provide higher levels of accuracy for the classification of physical activity intensity in children with intellectual disabilities, in comparison with existing cut points. The differences in validity between the existing and current cut points varies between intensities, with smaller differences found for sedentary and more substantial differences found for the physical activity intensities.

7.5 Discussion

The purpose of this study was to cross-validate the cut points calibrated in Chapter 6 in a sub-sample of 14 children with intellectual disabilities. The following sections will compare the cross-validation statistics from the present study to cross-validation statistics reported in previous research which calibrated and conducted cross-validation analysis using ROC curves. Possible cut point-specific and general causes for the differences reported will be discussed. Furthermore, the findings relating to the validation of existing cut points will also be examined. This section will conclude with discussion on possible reasons for the differences in cross-validation results and the future research implications of these findings.

7.5.1 Cross-validation

7.5.1.1 Sedentary cut points

The current sedentary vertical axis cut point demonstrates high sensitivity and specificity for the classification of sedentary behaviours, with a kappa score of $\kappa = .66$. This cut point correctly classified 92.62% of sedentary behaviours and excluded 83.06% of non-sedentary behaviours, suggesting this cut points is accurate. The vector magnitude cut point provides a lower level of accuracy ($\kappa = .63$) but more equally optimises sensitivity (82.21%) and specificity (86.10%), in comparison with the vertical axis cut point. These findings suggest that the vertical axis cut point may be too high, therefore increasing the likelihood that higher intensity activity is incorrectly classified as sedentary. However, there is currently no consensus on whether sensitivity or specificity should be more highly weighted, or if optimising both is most valid.

In comparison with previous research which cross-validated vertical axis cut points, the current cut point exhibits a lower level of accuracy. The ≤ 100 cpm and ≤ 372 cpm cut points developed by Pulsford et al. (2011) and Mackintosh et al. (2012), show almost perfect agreement with an AUC = .98 and kappa score of $\kappa = .97$, respectively. Sensitivity scores of 98% and 99% and specificity scores of 100% and 97% were additionally reported for these cut points, respectively. Furthermore, for a cut points of ≤ 400 cpm, Vanhelst et al. (2011) reports almost perfect classification accuracy when cross-validated ($\kappa = .85$).

Although a slightly lower level of validity was found for the current sedentary cut point in comparison with those established in previous studies, the cross-validation results still suggest that a higher sedentary cut point is required for children with intellectual disabilities. However, it is important to consider possible reasons for the development of a higher cut point, and whether it is a result of participant-factors, the protocol, or a result of measurement error. Previous studies have noted a limitation with direct observation in that it does not account for extraneous movements requiring energy expenditure, such as fidgeting, which could be of particular relevance to sedentary activity. For example, if children were in the seated posture but fidgeting, such as foot tapping, a higher count per epoch would be recorded as opposed to sitting without fidgeting (Spruijt-Metz et al., 2009). Furthermore, these existing studies which report higher classification accuracy all used structured protocols for both calibration and cross-validation activities, which will at least partially account for the higher validity reported. Therefore, in comparison with the free-living protocol used in the present study, the coding of activity would be more valid, both for the direct observation and respiratory gas exchange, suggesting an effect of protocol and criterion measure – however, as these factors will likely effect the validation of each cut point, possible effects will be collectively discussed in Section 7.5.2.

In summary, the cross-validation of the current sedentary cut point exhibits a lower level of validity than that reported in previous studies. However, considering the free-living nature of the present study protocol in comparison with the structured and constant activities used in previous studies, it is likely that the protocols and criterion measures used in previous studies had a positive effect on validity.

7.5.1.2 Light intensity cut points

The vertical axis and vector magnitude cut points for light intensity exhibit low levels of accuracy (κ = .32 and .20, respectively). Similar trends were found for the vertical axis and vector magnitude cut points, with low sensitivity (32.93% and 24.80%, respectively) and disproportionately high specificity (93.72% and 92.84%, respectively). This suggests that the range of the count boundaries is too small to accurately classify all light intensity behaviour, but small enough to limit the probability of false-positives. This finding is consistent with previous validation research, as Trost et al. (2011) found that light intensity cut points are notably less valid than other intensities, with specificity highly outweighing sensitivity.

This low accuracy could be attributed to various factors, but is most likely an effect of the analysis used. The light cut points were the only developed cut points which were not specifically calibrated, i.e. the optimal light intensity cut point was not identified using ROC curve analysis. Instead, the sedentary cut point and lower moderate cut point were used as the boundaries for light intensity activity, which is common practice in the development of light intensity cut points using ROC curves (Mackintosh et al., 2012). Therefore, the validity of the light intensity cut point is somewhat dependant on the validity of the sedentary and lower moderate cut points. As discussed in the preceding section (7.5.1.1), the disproportionately high sensitivity score suggests that the vertical axis sedentary cut point is too high. Therefore, if this cut point was lower, theoretically, a larger proportion of light intensity activity would be correctly classified and subsequently increase the accuracy of the cut point, i.e. reduce the likelihood of false-negative scores. Mackintosh et al. (2012) investigated the effect that lowering the optimal calibration cut point by ± 90 cpm and ± 200 cpm had on validity but found that changing cut points had a negligible effect on validity; an adapted version of their findings is presented in Chapter 6 (Table 6.11). However, as light intensity activity is not a commonly measured intensity in research, with sedentary and moderate to vigorous most widely used, the lower validity of this cut point may not have an important effect on research.

The light intensity cut points developed in this study are less accurate than those developed in previous studies. For the vertical axis, the light intensity cut point calibrated by Jimmy et al. (2013) correctly classified 81% of light intensity activity, and misclassified the remaining 19% as moderate. Unlike the present study, the vector magnitude cut points were more accurate than the vertical axis cut points, and correctly classified 94% of light activity, and misclassified the remaining 6% of light activity as moderate intensity. The vertical axis cut points developed by Pulsford et al. (2011; 100–2240 cpm), achieved sensitivity = 59%, specificity = 83%, and AUC = .61. Higher accuracy was reported for the cut points developed by Vanhelst et al. (2011; 401-1900 cpm), which achieved overall agreement of κ = .72, although this study did not report sensitivity or specificity scores.

In-depth comparability with Jimmy et al. (2013) and Vanhelst et al. (2011) is somewhat limited as neither study reported ROC curve statistics and instead reported only confusion matrices to determine the percentage of correctly classified observations and kappa scores. Furthermore, the cross-validation conducted by Jimmy et al. (2013) included only 88 observations in comparison with the 1154 observations used in the present study; therefore, the validity of these results may be limited by the small data set used. The Pulsford et al. (2011) cut points demonstrate a similar trend in which specificity is disproportionately high to sensitivity, however, the higher scores could be a result of the wider count boundaries, i.e. due to the lower sedentary and higher moderate cut points.

Sedentary behaviours and moderate to vigorous intensity are the most commonly measured activity intensities and the focus of physical activity guidelines, due to the greater associated health benefits. Therefore, it could be argued that the low validity of the light intensity cut points will have a limited effect on future research. However, the focus on ≥ moderate intensity activity has left a gap in our knowledge regarding the health benefits of light intensity activity, with the minimal intensity required for health benefits unknown (Chaput, Carson, Gray, & Tremblay, 2014). Therefore, there is a need for additional research on light intensity activity, both in relation to its independent health effects and use as a transition intensity for increasing moderate to vigorous intensity activity (Carson et al., 2013). This could be of particular relevance to children with intellectual disabilities, as moderate to vigorous intensity activity only accounts for a relatively small portion of this population’s activity, as discussed in Chapter 1. Furthermore, considering the additional heath needs of this population and associated physical limitations, light intensity activities, such as slow walking, may be more widely conducted in this group (Ryan, Forde, Hussey, & Gormley, 2015). As a result, understanding the benefits of movements and activities which are classified as light on the intensity continuum requires further investigation. However, the developed cut points do not provide a valid method to do this.

In summary, the vector magnitude and vertical axis light intensity cut points are not valid, demonstrating poor classification accuracy. Yet, this lack of accuracy is consistent with previous research involving typically developing children. From a public health perspective, there is an increasing focus on measuring and understanding light intensity activity; therefore, the limited validity of these cut points will have wider implications relating to increasing our understanding of light intensity activity in children with intellectual disabilities.

7.5.1.3 Moderate intensity cut points

The vertical axis moderate intensity cut points demonstrate substantial accuracy ($\kappa = .74$), with sensitivity = 74.92% and specificity = 95.65%. However, the vector magnitude cut point exhibits a lower level of accuracy ($\kappa = .51$), with specificity (89.42%) outweighing sensitivity (59.74%). Similar to light intensity, the higher specificity of these cut points could suggest that the boundaries are too small to accurately detect all moderate intensity activity but small enough to limit the probability of false-positives. Again, similar to the light cut point, the upper boundary of moderate is a result of calibrating a cut point for vigorous intensity. Therefore, the validity of the upper boundary will be somewhat dependent on the validity of the vigorous cut point. However, as almost perfect agreement and optimised sensitivity and specificity were found for the vigorous intensity cut points, this suggests that the vigorous cut points do not require alteration. Another possible reason for the lower validation found in the present study for the moderate intensity cut points could be a result of SOFIT. Specifically, ordinary walking was the only behaviour classified as moderate, which required a subjective element with regards to deciding what constituted “ordinary walking” and what was “fast walking”, i.e. very active.

In comparison with previous research, Jimmy et al. (2013) reports similar findings with the moderate intensity cut points exhibiting lower accuracy than other intensities. Furthermore, the vertical axis cut points developed by Jimmy et al. (2013) show higher validity than the vector magnitude cut points, with 51% and 29% of moderate activity correctly classified, respectively. Similarly, the 2241−3840 cpm cut points developed by Pulsford et al. (2011) demonstrated poor validity (AUC = .60), with sensitivity = 60% and specificity = 76%. However, the cut points developed by Vanhelst et al. (2011) and Mackintosh et al. (2012)

show higher validity (κ = .88 and .71, respectively), with Mackintosh et al. (2012) additionally reporting sensitivity and specificity scores of 94% and 75% respectively. Therefore, the trend of lower validity for moderate intensity cut points is consistent with some previous studies. However, the higher validity found by Mackintosh et al. (2012), which used also used a criterion measure of SOFIT, suggests that possible effects associated with classifying walking may be limited.

In summary, the vertical axis moderate intensity cut point demonstrates a similar or higher level of validity in comparison with previous research, with Jimmy et al. (2013) also reporting lower validity for the vector magnitude cut point. Although the classification of SOFIT could possibly account for the slightly lower validity reported for this cut point, the high validity reported by Mackintosh et al. (2012) suggests limited measurement error caused by SOFIT.

7.5.1.4 Vigorous intensity cut points

The current vigorous intensity vertical axis cut point shows almost perfect agreement (κ = .85) with equivalent sensitivity and specificity scores of 93.00% and 94.61%, respectively. The vector magnitude cut point also demonstrates equivalent sensitivity (89.33%) and specificity (89.46%), although lower, yet still substantial, agreement (κ = .74). The high kappa scores and optimised sensitivity and specificity demonstrates that the calibrated cut points provide a valid threshold for classifying vigorous intensity activity. In relation to the previously discussed point that the classification of walking could have contributed to the lower validity of the moderate cut point, the very high specificity scores demonstrates that very little non-vigorous activity, such as “ordinary walking”, was misclassified as vigorous.

The increased validity of the vertical axis cut point contradicts Jimmy et al. (2013), who report higher validity for the vector magnitude cut point (sensitivity = 89%, specificity = 80%, κ = .63) in comparison with the vertical axis cut point (sensitivity = 74%, specificity = 79%, κ = .50). Mackintosh et al. (2012) also report lower validity for their vigorous cut point of ≥ 4807 cpm, with sensitivity = 79%, specificity = 89%, κ = .62. On the other hand, Pulsford et al. (2011) report almost perfect validity (sensitivity = 95%, specificity = 91%, AUC = .98) for a cut point of

≥ 3841 cpm, with comparably high validity found for the cut point of ≥ 3919 cpm developed by Vanhelst et al. (2011; κ = .91). Therefore, the higher validity in the present study further supports the emerging pattern that children with intellectual disabilities require lower cut point boundaries for the accurate classification of physical activity intensity.

In summary, both the vertical axis and vector magnitude cut points exhibit high validity for the classification of vigorous intensity activity. Considering that the vertical axis cut point is lower than some previously calibrated moderate intensity cut points, this high validity further supports the need for lower cut points in children with intellectual disabilities to prevent a systematic underestimation of vigorous intensity activity.

7.5.1.5 Moderate to vigorous intensity cut points

The validity of the moderate to vigorous intensity cut points in the present study is consistent with the previous findings that the vertical axis cut point (sensitivity = 90.94%, specificity = 95.06%, κ = .85) provides a higher level of validity than the vector magnitude cut point (sensitivity = 91.27%, specificity = 83.55%, κ = .75). The disproportionately high sensitivity for the vector magnitude cut point indicates that this boundary is too low and therefore incorrectly classifying a higher number of non-moderate to vigorous epochs as moderate to vigorous (false-positive). On the other hand, the vertical axis cut point will provide a valid measure of moderate to vigorous intensity activity in children with intellectual disabilities.

This is an important and encouraging finding as this intensity is regarded as health-enhancing and required for the attainment of physical activity guidelines. However, there are numerous studies which suggest a cut point of in the range of 2300–3000 cpm should be used for ActiGraph devices to classify moderate to vigorous intensity in children (Clanchy et al., 2011; Ekelund et al., 2004; Guinhouya & Hubert, 2008; Trost et al., 2011; Vanhelst et al., 2011). In contrast, the high validity of the current ≥ 1008 cpm cut point suggests that a much lower cut point is required to accurately classify moderate to vigorous intensity activity in children with intellectual disabilities. Therefore, this raises important issues surrounding the validity of previous research and the generalisation of existing cut points, which will introduce a high level of systematic error and underestimate physical activity intensity in children with intellectual disabilities.

Jimmy et al. (2013) was the only previous study which cross-validated the moderate to vigorous cut point and, consistent with the present study, found the vertical axis cut point (sensitivity = 85%, specificity = .81%, κ = .55) to be more accurate than the vector magnitude cut point (sensitivity = 72%, specificity = 94%, κ = .45). The higher validity for the vertical axis cut point found in the present study and by Jimmy et al. (2013) is an interesting finding. Although the use of vector magnitude is still in its relative infancy, it was envisaged that this method would provide a higher level of validity as, theoretically, the measurement of all three axes should most accurately capture the dynamic nature of children's moderate to vigorous intensity activity. However, based on these findings, the inclusion of three axes reduces validity. Therefore, a possible explanation for this is that children, both typically developing and those with intellectual disabilities, conduct additional, extraneous movements which are captured by the additional axes, therefore introducing random error and reducing validity.

In summary, considering the notably lower moderate to vigorous cut points established in this study, the high validity exemplifies the need for population-specific measurement to prevent further systematic error caused by generalising cut points between populations.

7.5.2 Factors affecting cross-validation

The physical activity cut points cross-validated in this chapter exhibit comparable or higher validity than that reported in previous studies, although the sedentary cut points did show lower validity than previous studies. There are various factors that could have attributed to the differences between studies, some of which have been discussed in the preceding sections. However, there are some additional possible causes for the differences in reported cross-validation that generally apply to all cut points; specifically, the protocol, criterion measure, and participants. The effect of these factors in relation to calibration has previously been discussed in detail in Chapter 6. However, it is also important to consider these possible effects specific to cross-validation.

7.5.2.1 Protocol

The design of the protocols used in these previous studies will have potentially impacted on cross-validation, both in relation to the structure and type of activities used. Pulsford et al. (2011), Vanhelst et al. (2011), and Jimmy et al. (2013) used the same protocol and activities for calibration and cross-validation. Therefore, the cross-validation analysis conducted in these studies is restrictive and gives little indication of the validity of generalising these cut points to free-living physical activity. Furthermore, Pulsford et al. (2011) and Vanhelst et al. (2011) used constant, structured activities, with Jimmy et al. (2013) and Mackintosh et al. (2012) using a combination of structured and semi-structured activities. As a result, the protocols used in these previous studies will limit the validity of their results, as cross-validation should be conducted in different activities from calibration (Welk, 2005). Furthermore, cut points only reflect the energy or intensity demands of the activities in which they were calibrated; therefore, the use of a different activity protocol for cross-validation increases our understanding of the generalisability of the cut points to other activities (Ekelund et al., 2004).

Mackintosh et al. (2012) was the only study that used different activities for calibration and cross-validation. For physical activity intensities, the cut points were calibrated using structured activities and cross-validated using a free-play protocol; therefore, the validity established in this study will be more representative of field-based validity. However, for the calibration of sedentary activity, Mackintosh et al. (2012) used similar, structured activities for calibration (drawing/colouring) and cross-validation (DVD watching). Therefore, for the cross-validation of sedentary cut points, all previous studies used a structured protocol which involved constant sitting and/or lying down. As a result, due to the constant and structured nature of the included activities for the cross-validation, it can be assumed that this increased the validity found for the sedentary cut points. In comparison, as cross-validation in the present study was conducted on free-living activities, this increases the likelihood of different intensities being included in an epoch coded as sedentary.

7.5.2.2 Criterion measure

The use of a structured protocol for cross-validation will increase the accuracy and reduce the measurement error associated with the criterion measure. For SOFIT, the classification of behaviours will not be affected by transitions and the subjective nature of classifying walking will be less. On the other hand, for physiological criterion measures, the use of structured activity will enable the use of steady state measurement, therefore increasing validity (McClave et al., 2003). Pulsford et al. (2011) and Jimmy et al. (2013) both specified that the oxygen uptake data used for calibration was steady state. Furthermore, Jimmy et al. (2013) only included 1 or 2 minutes of the total measurement period (which ranged from 3 minutes 15 seconds to 4 minutes 15 seconds) in the analysis. Yet, although this enabled the use of steady state measurements, it reduced the available data and subsequently statistical power and, as a result, Jimmy et al. (2013) only included 88 measurement epochs in the cross-validation analysis.

7.5.2.3 Sample

Another limitation with the studies conducted by Pulsford et al. (2011) and Mackintosh et al. (2012) is that cross-validation was conducted on the same sample in which the cut points were calibrated. Therefore, as an important aspect of cross-validation is investigating classification accuracy in a different sample, the reported validity may be higher in these studies as this analysis did not account for between-sample differences (Welk, 2005). On the other hand, Vanhelst et al. (2011) and Jimmy et al. (2013) conducted cross-validation in a sub-group of the recruited sample. In these studies, the use of a different sample did not seem to affect cross-validation, as comparable or higher validity was reported, in comparison with Mackintosh et al. (2012) and Pulsford et al. (2011). Furthermore, as the use of cut points in an independent sample is more representative of the conditions in which the developed cut points will be used, the validation of the current cut points may be more ecologically sound.

Another factor that could have affected cross-validation is between-participant differences. Specifically, as highlighted in Table 7.2, very large standard deviations were recorded for each intensity. This suggests that a wide range of counts per epoch were recorded for the same intensity, which could be a result of participant differences, such as the force generated during certain movements. Individuals with disabilities are known to exhibit high within- and between-participant differences, which can make the attainment of stable criterion scores difficult (Rikli, 1997). However, the high standard deviation scores could also include measurement error associated with SOFIT, whereby the intensity code given is not representative of the activity conducted throughout the entire epoch. Although, as the cross-validation group recorded notably lower standard deviations for almost all intensities, it could be argued that these high scores are not entirely a result of SOFIT but could partially be attributed to a greater variance between participants in the calibration group.

7.5.3 Validation of existing cut points

The validation of existing cut points against the criterion measure of SOFIT produced many interesting findings which will have important implications on the interpretation of past research and how future research is conducted. The large differences found for the validation of existing cut points demonstrates the very low levels of accuracy of generalising cut points calibrated in typically developing children to children with intellectual disabilities. As discussed in the previous section, some of these cut points exhibited high levels of validity when cross-validated as part of the original calibration studies. Therefore, when the low levels of validity are compared to the validity of the current cut points, this further highlights the need to conduct population-specific measurement research.

7.5.3.1 Sedentary cut points

The sedentary cut points all exhibited moderate to substantial classification accuracy, with the cut points that have higher boundaries producing moderate, and comparable, levels of validity to the current cut points (Freedson et al., 2005; Mackintosh et al., 2012; Puyau et al., 2002; Vanhelst e al., 2011). This therefore suggests that use of these cut points will provide a valid estimation of sedentary time in children with intellectual disabilities. On the other hand, the lower validity for the remaining cut points demonstrates that a lower count boundary is not valid in children with intellectual disabilities and will result in an underestimation of sedentary time. This is in contrast to the emerging research recommending the use of a 100 cpm threshold for sedentary behaviour, which has demonstrated high validity in both children and toddlers (Janssen et al., 2013; Trost et al., 2011). However, as the importance of understanding sedentary behaviour as a construct independent of physical activity has increased in recent years, it is vital that this type of behaviour can be accurately measured so that risk factors can be identified and effective interventions developed (Owen, Healy, Matthews, & Dunstan, 2010).

7.5.3.2 Moderate and vigorous intensity cut points

The validation results for moderate intensity activity are low, with six sets of existing cut points demonstrating a kappa score of < .00, suggesting that the classification accuracy of these cut points is less than chance. The cut points developed by Freedson et al. (2005) and Jimmy et al. (2013) show the highest validity (κ = .48 and .39, respectively), although this is still notably lower than the substantial agreement of the current cut point (κ = .74). When the sensitivity and specificity scores are considered, the disproportionally high specificity suggests that these cut points are too high to capture moderate intensity activity in children with intellectual disabilities. Furthermore, the mean count data for moderate intensity activity, as presented in Table 7.2, is 552.10 (SD = 245.10) counts/20-seconds, which equates to 1656.30 cpm. Therefore, as the cut points which exhibit a validity of κ< .00 range from 2241–3200 cpm, it is clear that these cut points are too high for children with intellectual disabilities.

A similar trend is seen for the vigorous intensity cut points where specificity scores generally outweigh sensitivity, suggesting these cut point are also too high to measure vigorous physical activity in children with intellectual disabilities. This is confirmed by the mean 4161.78 cpm recorded for children with intellectual disabilities, as the existing typically developing cut points demonstrating low validity ($\kappa \leq$ .35) range from 4807–8200 cpm. However, the cut points which show a higher level of validity ($\kappa \geq$ .58) range from 2316–4012 cpm.

These findings for the moderate and vigorous intensity cut points are somewhat consistent with previous research in typically developing children, which also found that specificity outweighs sensitivity, thus increasing the likelihood of false-negative scores (Trost et al., 2011). However, Trost et al. (2011) - which was the only identified study that investigated the validity of moderate and vigorous intensity cut points independently - reports very different classification accuracy scores. For moderate intensity, the Treuth et al. (2004), Freedson et al. (2005), and Evenson et al. (2008) cut points exhibit fair classification accuracy (AUC = .71−.79), with the Puyau et al. (2002) and Mattocks et al. (2007) cut points showing poor validity (AUC = .56−.63). For the vigorous cut points, the Evenson et al. (2008) cut point demonstrated the highest accuracy (AUC = .84) with Freedson et al. (2005) and Treuth et al. (2004) showing fair classification accuracy (AUC = .77 and .73, respectively). Although, similar to the present study, the Puyau et al. (2002) and Mattocks et al. (2007) cut points, which have the highest cut point boundaries (≥ 8200 and ≥ 6130 cpm, respectively), demonstrate poor classification accuracy (AUC = .54−.66) and very low sensitivity (7.50% and 31.50% respectively).

The higher validity reported by Trost et al. (2011) suggests that, with the exception of the Puyau et al. (2002) and Mattocks et al. (2007) vigorous intensity cut points, the remaining existing cut points provide a valid measure of physical activity in typically developing children. When this is compared to the very low validity found in the present study, it supports the inference that existing cut points – although valid in a population of typically developing children - are too high and therefore not valid for children with intellectual disabilities. Furthermore, the poor accuracy reported by Trost et al. (2011) for the Puyau et al. (2002) and Mattocks et al. (2007) vigorous cut points suggests that as these count boundaries are also too high for typically developing children. Therefore, not all the error in classification accuracy can be attributed to differences between children with intellectual disabilities and typically developing children, and may be affected by study specific factors, such as protocol. This is also relevant to the study by Trost et al. (2011), which used a structured protocol and therefore the results may not be fully representative of the validity of existing cut points during unstructured, field-based activity.

7.5.3.3 Moderate to vigorous intensity cut points

For existing moderate to vigorous intensity cut points, classification agreement ranged from κ = .30−.74, in comparison with the almost perfect agreement found for the current cut point κ = .85. Due to the higher count thresholds, all existing cut points underestimated time spent at a moderate to vigorous intensity, with specificity outweighing sensitivity for all cut points, except Freedson et al. (2005). The higher validity reported for the Freedson et al. (2005) cut point is a result of the very low 500 cpm cut point; however, this will also increase the likelihood of false-positive scores, i.e. misclassifying light intensity as moderate to vigorous intensity. Therefore, based on these results, no existing cut points provide a valid method of estimating moderate to vigorous intensity activity in children with intellectual disabilities.

These findings are similar with previous validation research which found that the cut points developed by Mattocks et al. (2007) and Treuth et al. (2004) underestimate moderate to vigorous intensity activity by 39%−74%, against a criterion measure of SOFIT (McClain et al., 2008). Trost et al. (2011) also report similar findings in that the Freedson et al. (2005) and Evenson et al. (2008) cut points show high levels of validity (AUC = .90), with the remaining cut points underestimating activity, although demonstrating fair to good classification accuracy (AUC = .77−.85). It is important to note, however, that not all of the cut points included in the present study were included in these previous studies.

Moderate to vigorous intensity activity is primarily important as this intensity threshold is used to distinguish health-enhancing levels of activity and the attainment of physical activity guidelines. To ensure that our understanding of sedentary and physical activity behaviours is accurate, it is vital that the measurement method used is valid. It has long been acknowledged that the method used to interpret data will have an effect on the estimates of physical activity intensity (Riddoch & Boreham, 1995). Therefore, ensuring valid data interpretation is paramount. However, the use of existing cut points in children with intellectual disabilities will not provide valid measurements of moderate to vigorous intensity activity and, as a result, will underestimate physical activity intensity and produce clinically meaningful differences.

7.5.4 Strengths and limitations

This study was the first to validate population-specific accelerometer cut points for children with intellectual disabilities. As existing cut points have been generalised to children with intellectual disabilities in previous research, the major strength of this study is that it provides the first stages of establishing valid methods to interpret accelerometer output for the estimation of physical activity intensity in children with intellectual disabilities. The methods used for cross-validation were also in line with best practice recommendations, in that validation was conducted during a field-based protocol in a sample independent of the calibration group (Bassett et al., 2012; Welk, 2005). Furthermore, in addition to cross-validating the developed cut points, this study also validated all existing child-specific ActiGraph cut points, and subsequently demonstrated the superiority of the developed cut points over existing cut points.

The limitations of this study primarily relate to possible sampling and analysis errors, which were also limitations of the calibration study. Session duration and time spent in each intensity varied, with longer session durations and higher percentages of time spent at moderate and vigorous intensities occurring in sessions which were observed to include participants with lower levels of intellectual disabilities. Therefore, it is possible that the cross-validation results are based on a data-set which is more representative of children with milder intellectual disabilities. More generally, a limitation with cross-validation is that it only estimates how valid the developed cut points will be in a similar sample to the calibration sample (Staudenmayer et al., 2012). Therefore, as cut points are generally age-specific, investigation into the effect of age on validity is important (Trost et al., 2011). However, as the age range of children in the present study was relatively small, it was not possible to make inferences regarding the validity of these cut in a sample of children with intellectual disabilities who are younger or older than the included sample.

7.5.5 Conclusions

This study was the first to examine the validity of population-specific cut points for children with intellectual disabilities and empirically demonstrate the superior validity of these cut points in comparison with existing cut points calibrated in typically developing children. The cut points cross-validated in this study exhibit high classification accuracy, giving promise to the wider application of these cut points for use in children with intellectual disabilities. The cross-validation results show that the calibrated cut points exhibit good to excellent overall classification accuracy for vertical axis and vector magnitude cut points. Furthermore, the identified trends in the calibration analyses are consistent with the cross-validation findings, such as vertical axis being more accurate than vector magnitude and the cut points more accurately measuring moderate to vigorous activity, in comparison with sedentary behaviours. Furthermore, the current cut points exhibit a substantially greater level of accuracy than existing cut points.

The large number of existing cut points and the high variance in thresholds between these cut points is negatively affecting the quantification, interpretation, and the real world application of data collected using accelerometers. To address this, researchers have aimed to create standardised cut points, specifically 100 cpm for sedentary and 2300–3000 cpm for moderate to vigorous intensity. However, the lower population-specific cut points validated in this study suggests that children with intellectual disabilities require a higher cut point for the classification of sedentary behaviour but lower cut points for the classification of physical activity. Therefore, the standardisation of physical activity measurement cannot be generalised to children with intellectual disabilities.

However, the lack of generalisability between cut points does not prevent physical activity being directly compared between studies. One important method of increasing clarity and comparability between studies is to additionally present accelerometer data in the form of counts, such as the mean (SD) data presented in Table 7.2. Furthermore, it is important to acknowledge that this study only provides the first stages of validation; therefore, a future longitudinal study investigating and comparing classification accuracy over time would add valuable knowledge to this emerging area of research (Trost et al., 2011).

Chapter 8 – General discussion

8.1 Introduction

Physical inactivity is a growing public health concern and is one the leading causes of worldwide mortality, with 1 million deaths per year in Europe (10%) attributed to physical inactivity (WHO, 2010; WHO, 2015). However, as discussed in Chapter 1, physical activity has a correlational and casual relationship with many physical and mental health benefits and a reduction in risk factors, such as reduced risk of cancer, chronic heart disease, and depression. Therefore, the promotion and advocacy of active lifestyles is ever-increasing. Furthermore, as physical activity levels in childhood are a strong determinant of physical activity in adulthood, there is a growing research need to better understand physical activity levels and behaviours in children, and to develop effective interventions (Telama, 2009; Telama et al., 2005).

The promotion of increased physical activity is primarily important in children with intellectual disabilities, as recent studies show that many of this population group do not meet the recommended levels of health enhancing physical activity and, therefore, are at a higher risk of obesity and its associated risk factors (Boddy et al., 2015; Einarsson et al., 2015; Maiano, 2010). As a result, young adults with intellectual disabilities have levels of cardiorespiratory fitness, muscular strength, and physical activity similar to older adults without intellectual disabilities, which contributes to the increased rates of obesity, falls and diabetes described in the literature (Balogh, Lake, Lin, Wilton, & Lunsky, 2015; Graham & Reid, 2000; Melville et al., 2008). Valid measurement is the first stage of identifying health-related outcomes that could be improved with increased physical activity, i.e. based on the dose-response relationship. From this, target outcomes can be identified which can be addressed through interventions; however, the lack of measurement-specific research and valid data interpretation methods for children with intellectual is hindering the advancement of this area of research. Therefore, the purpose of the research described within this thesis was to increase our understanding of accelerometer use and develop effective data interpretation methods specific to children with intellectual disabilities.

8.2 Summary of principal findings

This thesis presents the first research to develop valid and population-specific methods of interpreting accelerometer output in children with intellectual disabilities. The innovative nature of this research has generated various new findings and made a substantial contribution to our knowledge regarding physical activity measurement in children with intellectual disabilities. Prior to this research, the knowledge-base on physical activity measurement in children with intellectual disabilities was very limited. However, this research has increased our knowledge of how accelerometers are used, the effects that accelerometer use decisions, such as choice of cut points, have on results, and the feasibility of laboratory-based methods and measures. Therefore, as this area of research is in its relative infancy, the specific findings and knowledge generated from this thesis are very important to the development and progression of this field of research. Furthermore, the calibration of intensity cut points has not only provided researchers with the first valid method of interpreting accelerometer output in children with intellectual disabilities, but has empirically demonstrated that children with intellectual disabilities require substantially lower cut points than typically children, which has previously only been theorised (Frey et al., 2008; Hinckson & Curtis, 2013).

In relation to the five research aims of this thesis, the principal findings were:

1. *To systematically review accelerometer use in field-based physical activity research involving children with intellectual disabilities.*

Accelerometer use varied greatly between studies involving children with intellectual disabilities, with only a small percentage of use decisions being conducted in line with best practice recommendations (12–47%). Therefore, the majority of accelerometer use decisions were negatively impacting on the validity of results and/or clarity of reporting. This systematic review identified numerous areas of accelerometer use which required further investigation; however, considering the fundamental importance of valid data interpretation, this area was deemed most important as the focus for this thesis.

2. *To examine the effect of accelerometer cut points on the estimation of physical activity intensity in children with intellectual disabilities.*

Many significant and clinically meaningful differences were found for each intensity between existing accelerometer cut points that were calibrated in typically developing children. The consolidation of all previous child-specific ActiGraph cut points within this study provided a wide scope for discussion on factors that potentially affect calibration. This provided valuable findings relating to furthering our understanding of measurement research and the design of a future calibration protocol. From this, the importance of the following factors was highlighted: cut points cannot be generalised between ActiGraph devices, the effects and limitations of the criterion measure used need to be understood, and ROC curve analysis provides a more valid method of data interpretation than regression analysis.

3. *To develop an effective and feasible accelerometer calibration protocol for children with intellectual disabilities.*

A laboratory-based design provides direct physiological data for the purposes of calibration, which was deemed an important area of study as this is the first calibration research conducted in children intellectual disabilities. The feasibility of five general areas was investigated: recruitment, activities, resting energy expenditure measurements, treadmill-based graded exercise test, and breath by breath respiratory gas exchange. Furthermore, the relationship between accelerometry and $\dot{V}O_2$ was investigated. This study provided a vast amount of data from which the subsequent studies were based, although this study raised as many questions as it answered. With the exception of the free-living activity protocol, feasibility and validity issues were identified with all other areas investigated. Therefore, the design of a field-based protocol was recommended.

4. *To calibrate population-specific accelerometer cut points for the estimation of physical activity intensity in children with intellectual disabilities.*

Based on a field-based semi-structured activity protocol and using ROC curve analysis, cut points were calibrated for the vertical axis and vector magnitude, both in the form of counts per minute and counts per 20-second. For the vertical axis, the developed cut points were: sedentary ≤ 507 cpm, light 508–1007 cpm, moderate 1008–2300 cpm, and vigorous ≥ 2301 cpm. For vector magnitude, the cut points were; sedentary ≤ 1863 cpm, light 1864–609 cpm, moderate 2610–4214 cpm, and vigorous ≥ 4215 cpm. The vertical axis and vector magnitude cut points exhibited excellent classification accuracy (AUC = .87–.94 and .86–.92, respectively), with higher sensitivity and specificity found for moderate and vigorous intensity activity in comparison with the sedentary cut points. Furthermore, an interesting finding was the slightly higher accuracy of the vertical axis cut points in comparison with the vector magnitude cut points.

5. *To cross-validate the developed accelerometer cut points in children with intellectual disabilities.*

The cross-validation findings were consistent with the calibration results in that the vertical axis cut points provided a slightly higher level of accuracy than the vector magnitude cut points (κ = .32–.85 and κ = .20–.75, respectively). Furthermore, the accuracy of the developed cut points increased with intensity, suggesting increased validity for higher intensity physical activity in comparison with sedentary behaviours. This study also enabled the validity of the combined moderate to vigorous intensity cut point to be investigated, with the vector magnitude cut point demonstrating substantial agreement (κ = .75, $p < .001$) and the vertical axis cut point showing almost perfect agreement (κ = .85, $p < .001$).

An additional and highly relevant finding of this study was the validity of existing cut points against the criterion measure of SOFIT. For the measurement of sedentary behaviour, the accuracy of existing cut points (κ = .55–.67) was similar to the developed cut point (κ = .66). However, for the physical activity cut points, notable differences were found at each intensity, with numerous cut points demonstrating a lower level of validity than chance (κ < .00). This raises serious questions regarding the use of typically developing cut points in children with intellectual disabilities and reiterates the need for the research described within this thesis.

8.3 Relevance and implications of findings

8.3.1 Bridging the research gap

Physical activity research involving children with intellectual disabilities is a neglected area of study (Frey et al., 2008). Therefore, a lot of research being conducted in this area is lagging behind the knowledge and technological advances seen in typically developing populations. To highlight the slow progress in intellectual disabilities research in comparison with typically developing research, the first study reporting accelerometer calibration in a typically developing population was published 1983 and used an almost identical methodology to that reported in Chapter 5 of this thesis, i.e. respiratory gas exchange was measured whilst participants conducted treadmill-based and free-living activities (Montoye et al., 1983). Therefore, the field of accelerometer calibration in populations with intellectual disabilities is over three decades behind research in typically developing populations. This demonstrates that although physical activity has been widely measured in children with intellectual disabilities, and the methods and technology required to validate accelerometry has been around for over 30 years, this thesis is the first to address the need for valid data interpretation in this population.

The importance of population-specific cut points, and the low validity associated with generalising data interpretation techniques that are based on parameters of energy expenditure, are known (Freedson et al., 2005; Staudenmayer et al., 2012). However, with the advancement of accelerometer technology and the greater depth of data that can be collected and stored, new techniques for interpreting data are now being investigated. Therefore, the development and use of intensity cut points is no longer recommended as a method for interpreting accelerometer output. Freedson et al. (2012) specified that:

"for data interpretation, researchers should discontinue development and use of cut point methods to define intensity categories. Alternative analytic techniques, such as pattern recognition methods, that use more features of the raw acceleration signal and reduce the likelihood of extreme over- or underprediction of energy expenditure are recommended." pg. S2

However, this recommendation has to be interpreted relative to the current state of both intellectual disabilities and typically developing measurement research. Firstly, pattern recognition - which aims to identify patterns in the raw acceleration signal to detect which type of activity is being conducted - is still in an early developmental stage. Therefore, regardless of the future measurement potential of this technique, it is currently not an effective or valid method for measuring physical activity. On the other hand, specific to intellectual disabilities measurement research, prior to the research in this thesis being conducted, there were no population-specific methods to interpret accelerometer output. Therefore, due to the validity issues associated with generalising cut points, it was deemed more important to firstly address this gap in intellectual disabilities research and develop valid, population-specific intensity cut points.

The use of the developed cut points will provide a more valid method of data interpretation for research involving children with intellectual disabilities in the short-term. Nonetheless, it is important that measurement research in this population continues to progress. The current state of measurement research in typically developing children is that the interpretation of accelerometer data using energy expenditure regression equations and intensity cut points is no longer recommended, with no valid alternative currently widely available. Therefore, the stall in accelerometer measurement research specific to data interpretation provides the ideal opportunity for intellectual disabilities measurement research to continue to close this gap and keep abreast of the developments made with regards to pattern recognition techniques.

One of the potential advantages of pattern recognition is increased validity as the error associated with physiological parameters will be less because this method primarily focusses on the type of activity being conducted. As a result, the criterion measure for pattern recognition calibration and validation studies will be direct observation during physical activity (Freedson et al., 2012). Therefore, re-analysis of the data collected in Chapter 6 and 7 of this thesis has the potential to be used for initial pattern recognition analysis in children with intellectual disabilities once the required statistical methods and algorithms have been developed.

8.3.2 Importance of feasibility testing

Continuing the theme within this thesis of not generalising findings from other populations to children with intellectual disabilities, the feasibility of methods and measures should not be generalised between populations. However, although it is important to consider potential participant-related factors that could affect data collection, the need to automatically adapt testing procedures for populations with disabilities should not be assumed (Rikli, 1997). Therefore, feasibility testing is an important component in the design of studies to increase our understanding of what procedures are feasible in terms of completion rates, reliability, and validity, with protocol adaptations only investigated if the methods are ineffective.

There are two primary reasons for a protocol not providing sufficient data, thus affecting reliability and validity: the “floor effect” and the “ceiling effect”. The floor effect refers to participants not being able to meet the minimum requirements of a protocol or test. Therefore, to ensure that a representative sample of participants across the spectrum of intellectual disabilities are included in data collection, it is important that the procedures are not too difficult resulting in participant exclusion. The ceiling effect more specifically refers to tests whereby the standard of the testing procedure is too low, which may skew or effect the results; for example, in a test where a high number of participants achieve a perfect score, effects of physical activity may not be accurately detected and measured over time.

In this thesis, testing the feasibility of the laboratory-based protocol was an important aspect of this programme of research. It was important to develop a protocol which limited the floor effect due to the physical and skill requirements of the activities. Furthermore, it was also important to investigate the floor and ceiling effects of the treadmill-based graded exercise test to ensure valid and representative measurements. The protocol-specific findings of this feasibility study, which suggest this protocol was generally not feasible, highlights the need for feasibility testing to be conducted.

In addition to investigating protocol-specific feasibility, there is also a need to investigate feasibility in relation to recruitment. This is primarily important as recruitment of children with intellectual disabilities to health-related research is low, therefore, there is a need to generate an evidence base to inform the development of effective recruitment strategies (Maiano et al., 2014). Therefore, to ensure a heterogeneous and representative sample of children with intellectual disabilities are recruited, it is important that future studies investigate and report the feasibility of new testing procedures and recruitment strategies.

8.3.3 Involvement of stakeholders

Ensuring that physical activity protocols and procedures are feasible for children with intellectual disabilities is not entirely dependent on conducting feasibility testing. The involvement of stakeholders within this thesis, specifically teachers and Active Schools staff, provided valuable input into the protocol design and the delivery of the activity sessions reported in Chapters 6 and 7. Furthermore, the Active Schools coordinators and teachers provided assistance and feedback relating to recruitment.

In future studies, it would be advantageous to involve stakeholders in the early stages of study development to continue to ensure that physical activity studies are not only effectively designed from a research perspective but also are enjoyable for the children who participate. This stakeholder involvement could be included in various ways, e.g. an official advisory group or unofficial communications with involved parties, as with the present study. Furthermore, for future studies which involve field-based designs, such as interventions with multiple days of measurement, the inclusion of parents could be of additional benefit.

In research involving adults with intellectual disabilities, there is an increasing understanding of the value that can be gained from including adults and their carers in the processes of designing and conducting health-related research. There is a growing trend of moving away from a passive participatory approach, as described in this thesis, to an active inclusive approach where adults and carers work in a research capacity and have an active role in shaping the research agenda, as opposed to the traditional “researcher/researched” roles. Although this has many advantages, such as breaking down barriers, it is often the case that the focus on including adults with intellectual disabilities in the processes of research has a negative impact on the quality of the research output (Walmsley, 2004).

When deciding upon the extent of user and stakeholder involvement, it is important to consider the advantages and disadvantages specific to the nature of the research. For example, as the research included within this thesis is very measurement-focussed, the type of protocol, possible methods of measurement, and intended outcomes used within this research are somewhat limited. Therefore, active inclusion may be more effective and feasible at the other end of the research compendium, such as in the design and implementation of interventions (Turk et al, 2012). Furthermore, children with intellectual disabilities may be too young to be effectively included in an active inclusive role. However, as the effectiveness of participation within this thesis was dependent on parents consenting to their child's participation and time/travel requirements (specific to laboratory-based study described in Chapter 5), and teachers assisting with recruitment and hosting sessions, the involvement of these stakeholders in the early processes of designing studies could have been beneficial.

In relation to the research conducted in this thesis, areas of study design which would benefit from parent, teacher, or Active Schools involvement would be:

- The design and wording used in parent and participant information sheets

- The design of physical activity-related protocols

- Input on the most appropriate methods of conducting data collection, e.g. how to appropriately explain the procedures and methods

- The organisation of parent sessions during recruitment to allow a first-hand explanation of the study, which could help improve the clarity of the research being conducted and develop ongoing relationships with teachers and parents.

8.3.4 Implications on past and future research

The findings of this thesis support previous recommendations that physical activity measurement research needs to be population-specific to account for between-group differences (Freedson et al., 2005; Freedson et al., 2012; Frey et al., 2008). The present research demonstrates that children with intellectual disabilities require lower accelerometer cut points for the estimation of physical activity intensity than typically developing children. Although these findings require additional validation, this raises questions on the conclusions made in previous research. As discussed in Section 1.3.4, there is a general consensus that children with intellectual disabilities are an inactive population and participate in less physical activity than their typically developing peers (Einarsson et al., 2015; Frey et al., 2008; Stanish & Mozzochi, 2000; Tyler et al., 2014). However, these conclusions are partially based on the use of existing cut points, and the application of the same cut points to children with intellectual disabilities and typically developing children (Einarsson et al., 2015; Tyler et al., 2014). Therefore, the lower physical activity levels of children with intellectual disabilities reported in these studies could be a result of systematic measurement error due to the use of cut points which are too high to accurately classify physical activity intensity in children with intellectual disabilities.

This could have important implications for future research as the findings from observational studies are used to inform longitudinal research, the development of interventions, and the subsequent translation into policy and practice (Biddle et al., 2015). As the findings from this thesis raise questions on whether children with intellectual disabilities are as inactive as described in the existing literature, this could lead to future research with limited validity and relevance being conducted. Therefore, it is important to develop a body of evidence focussed on physical activity in children with intellectual disabilities that is based on valid population-specific measurements. It would also be beneficial to conduct additional research on the prevalence and effects of light intensity activity in this population; however, with the light intensity cut points demonstrating low validity, other measurement methods need to be investigated, or light intensity cut points specifically calibrated.

As previously discussed, the generalisability of cut points is limited to the original calibration population. Furthermore, due to the energy expenditure changes associated with maturation, cut points calibrated for children are also age-specific. Therefore, the cut points developed within this thesis only provide a valid estimation of physical activity intensity in children with mild to moderate intellectual disabilities aged 8 to 11 years. Therefore, there are currently no valid methods to estimate physical activity intensity in younger children or adolescents with intellectual disabilities. Furthermore, although this thesis discussed possible physiological and biomechanical causes for the lower cut points calibrated for children with intellectual disabilities, making definitive conclusions is outwith the scope of this thesis. Therefore, it is not known whether the causes of these differences will continue into adulthood, thus raising further questions on the validity of accelerometer data interpretation in adults with intellectual disabilities and the need for further measurement research in this population.

This thesis was the first to empirically identify differences in the calibration of accelerometer cut points between children with intellectual disabilities and typically developing children. These findings have many wider implications as it raises questions on the validity of previous research and what is “known” in this field of research. However, understanding the causes of these differences and addressing the wider implications was outwith the scope of this thesis; therefore, numerous areas for future research have been identified.

8.4 Future research

Based on the findings of this thesis, the following areas of study are recommended for future research:

- Investigate the most valid and reliable accelerometer wear criteria, e.g. number of monitoring days and wear time required, for multiple days of monitoring in children with intellectual disabilities

- Investigate the adherence and compliance rates of multiple days of accelerometer wear in children with intellectual disabilities

- Investigate the low participation rates of children with intellectual disabilities in physical activity research, and develop effective recruitment strategies

- Investigate biological and biomechanical differences in parameters of physical activity, such as energy expenditure and ground reaction forces, between children with intellectual disabilities and typically developing children

- Conduct additional validation research on the cut points established in this thesis

- Conduct additional research into the physical activity levels of children with intellectual disabilities using population-specific cut points

- Calibrate and validate pattern recognition techniques in children with intellectual disabilities

- Conduct measurement research in other populations with intellectual disabilities, i.e. younger children, adolescents, and adults.

- Increase our understanding of light intensity activity and its health benefits in children with intellectual disabilities

8.5 General strengths and limitations

Specific strengths and limitations relating to each study have been discussed in the preceding chapters. However, there are some general strengths and limitations of the research presented within this thesis that are noteworthy.

One of the primary strengths of this research is that is addresses a substantial gap in our knowledge relating to effective and valid measurement of physical activity in children with intellectual disabilities. The studies reported in this thesis were the first to develop valid methods of interpreting accelerometer output for estimating physical activity intensity in children with intellectual disabilities. As valid measurement is a fundamental requirement of high quality research, this provides a starting point for the collection of more in-depth and accurate data in this population. Furthermore, as a trend in intellectual disabilities research is to generalise methodologies from typically developing populations, the research presented was based on substantial feasibility testing and discussions with stakeholders to ensure that feasible and valid methods were used.

Finally, although low recruitment rates were reported in Chapter 5, with this recruitment strategy deemed ineffective (research question 6), this limitation with recruitment was overcome in Chapters 6 and 7. In comparison with the recruitment strategy in Chapter 5, recruitment in Chapters 6 and 7 involved contacting a greater number of schools, including those outwith the Glasgow area. Furthermore, recruitment was primarily conducted in the autumn and, unlike the recruitment in Chapter 5, was not affected by school holidays. The support of Active Schools - which had existing relationships with schools - and their advocacy for this research, was also observed to increase the willingness of schools to be involved. This resulted in more schools being involved in recruitment, even though the research burden, in terms of data collection, was greater for schools. Within this recruitment strategy, there was also an additional focus on asking teachers to advise on recruitment, which helped ensure appropriate content and language was used in the information sheets given to parents. This highlights that the inclusion of teachers and other organisations can help the effectiveness of recruitment. Therefore, this provides initial data that can inform the development of future recruitment strategies for physical activity research involving children with intellectual disabilities.

The general limitations of this thesis relate to the study samples and the effect of disability type and severity. Firstly, no data were collected on the aetiology or severity of the intellectual disabilities of participants. Therefore, it was not possible to investigate disability-specific effects. Furthermore, no background information was collected relating to any medications being taken by participants which could affect heart rate or energy expenditure (Durstine & Moore, 2003). Considering the difficulties associated with recruiting children with intellectual disabilities, as discussed in previous research, and the need to recruit large and representative samples to calibration/validation studies, the decision not to collect this data was based on feasibility reasons (Maïano et al., 2014). Specifically, the additional burden on children, parents, and schools to collect this type of data, e.g. conducting standardised IQ tests and parent questionnaires, was deemed to be something that may limit participation.

Secondly, parents are a vital component in the recruitment of children to research. However, as discussed in Chapter 5, some parents involved in this research questioned their child's capability of completing the protocol. As a result, it is possible that parents of children with more severe disabilities declined participation due to the activity-focussed designs of the protocols used within this thesis. Therefore, this could have resulted in study samples which were not fully representative of the target population, thus raising questions on the validity of generalising cut points to children with moderate or higher levels of intellectual disabilities.

Thirdly, specific to the field-based protocol reported in Chapters 6 and 7, the session duration – and the subsequent volume of data included in the analysis – varied between sessions. From observations, the sessions which included children with more moderate levels of intellectual disabilities required more instruction and, therefore, less time spent on the activities – this was particularly apparent in session 7. Furthermore, two students who were exhibiting behavioural problems were removed from sessions (one prior to a session and one during a session). Therefore, the data and participants included in this research may not be fully representative of the target population of children with mild to moderate intellectual disabilities. Furthermore, as a larger volume of data were included from participants who were observed to have milder intellectual disabilities, the generalisability of these findings to children with higher levels of intellectual disabilities may not be valid.

Finally, this thesis did not investigate the sensitivity to change of the wGT3X+. Although an important aspect of measurement research, it was outwith the scope of this thesis. However, as sensitivity to change relating to accelerometers is device-specific and is not affected by data handling or interpretation methods, such as epoch or cut points used, the real-world effects may be limited when using the cut points calibrated within this thesis (Montoye et al., 2014).

8.6 Research and personal skills developed

The research described within this thesis was motivated by the need to address the lack of measurement-specific research in children with intellectual disabilities. At the point of commencing this PhD, my knowledge of research and physical activity measurement was almost entirely theoretical, with little knowledge or understanding of how to develop and conduct research, especially in a population of children with intellectual disabilities. Therefore, throughout the course of this PhD, I developed my understanding and abilities to design and conduct research using various methodologies, collect and analyse a wide range of data measured using multiple techniques, and write-up findings for both academic and non-academic audiences. In addition, I learned to view my research out with the scope of a single study, which allowed me to develop a coherent body of work.

Furthermore, considering the nature of the research described within this thesis, I also learned the practical skills required to not only recruit participants but to engage and build relationships with parents, teachers, external organisations, and researchers. During the periods of recruitment and data collection, my communication and organisation skills greatly improved, as did my ability to adapt and overcome unexpected events in the course of conducting this research. My communication skills also developed beyond the scope of working with participants as I learned the importance of communicating my research to the wider academic environment, through presentations and publications. From this engagement with the wider academic community, I was faced with many researchers who overlooked the relevance and importance of conducting physical activity measurement research in children with intellectual disabilities. As a result, one of the most valuable learning experiences I will take from this PhD is the importance of effective communication and advocacy for this important area of research.

Appendices

Appendix i – Feasibility study ethical approval

University of Glasgow | College of Medical, Veterinary & Life Sciences

29/05/13

Dear Dr Melville

MVLS College Ethics Committee

Project Title: Calibration of accelerometer cut-points in children with intellectual disabilities: A feasibility study
Project No: 200120045

The College Ethics Committee has reviewed your application and has agreed that there is no objection on ethical grounds to the proposed study. They are happy therefore to approve the project, subject to the following conditions:

- The research should be carried out only on the sites, and/or with the groups defined in the application.
- Any proposed changes in the protocol should be submitted for reassessment, except when it is necessary to change the protocol to eliminate hazard to the subjects or where the change involves only the administrative aspects of the project. The Ethics Committee should be informed of any such changes.
- If the study does not start within three years of the date of this letter, the project should be resubmitted.
- You should submit a short end of study report to the Ethics Committee within 3 months of completion.

Yours sincerely

Dorothy McKeegan

Dr Dorothy McKeegan
College Ethics Officer

Appendix ii – Feasibility study information sheets

University of Glasgow | College of Medical, Veterinary & Life Sciences

Child Information Sheet

Can you help?

We would like your help. Before you choose to help, we want you to know exactly what we're doing. Please read this and talk about it with your parents. Ask us if you have any questions.

Why do we want your help?

We would like you to take part in this study to help us choose what exercises and activities are the most fun for children to do. We would also like to measure how fit you are.

What will happen when you take part?

If you and your parents/guardian are happy to take part, we would like you to come to the University of Glasgow to take part in some activities. We would like you to come along on 2 different days, for about 2 hours each time.

During these 2 sessions, you will take part in different exercises and games. Some of these will be done on the floor and some will be done on a treadmill, which is a type of walking machine.

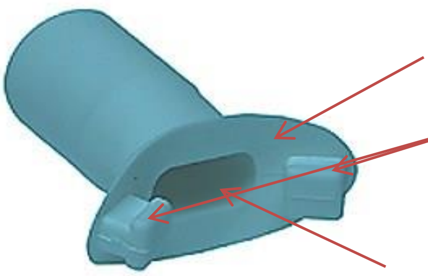
Before you do any exercises on the treadmill, we will show you how to walk properly on it and give you plenty of time to practice. We would then like you to walk on the treadmill at different speeds and then gently jog. Some speeds will be slow, like when you are gently walking about the shops, and some speeds will be faster, like how you walk when you are in a hurry.

We would also like you to do some fun activities, such as: watching a DVD, playing Xbox 360 Kinect, drawing, passing a football, hula hoop, and much more!

The last thing we would like to do is test how fit you are. We will do this on the treadmill. We will ask you to pick a speed you are comfortable with, and we will increase the treadmill angle while you walk. This will make it feel like you are walking up a hill. Once you feel out of breath we will stop the test.

Whilst you are doing the treadmill exercises, games and fitness test, we would like you to wear a special breath mouthpiece and a small mask to help keep it in place. The mouthpiece is attached to a tube which runs into a machine. This will let us measure the air you breathe out. Wearing this mouthpiece won't hurt you at all, but it might feel a bit strange at first.

The mouthpiece will look like this:



This circle-shaped piece will go in your mouth and sit between your teeth and lips

You will bite onto these to help hold the mouthpiece in place

The air you breathe out will go out this hole, through a tube, and into the machine.

What else will we be doing?

We would also like to know how tall you are and how much you weigh. We will keep this information private.

We will measure your height like this:

We will measure your weight like this:

Are there any risks?

We have aimed to make this as risk-free as possible for you. Because you may not be used to walking on a treadmill, we will give you plenty of time to practice. The treadmill also has hand rails than you can use to support yourself. Also, because this study involves you doing exercise, your muscles might feel a little tired the day after you take part.

Will it help me?

This study is part of our university work. Although it won't help you directly, we hope that you will have fun doing it.

Who will know I have taken part?

All information we get during the study will be kept very secret. If we want to show our work to other people, no one will know your name or that it was you who helped us.

Will I be told the results?

Yes, if you want. We can send you a letter to let you know what we found.

Financial arrangements

Researchers at the University of Glasgow will run this study - we have been given some money to help us do this.

Do I have to take part?

No. You don’t have to take part if you don’t want to. Also, if you decide to take part and then change your mind, that’s not a problem, you can stop at any point.

Who has checked the study?

This study has been checked over by the College Ethics Committee of the University of Glasgow to make sure it’s safe for you to take part.

University of Glasgow | College of Medical, Veterinary & Life Sciences

Parent Information Sheet

Study title

Investigation into exercise activities in children with intellectual disabilities.

Invitation paragraph

Your child is being invited to take part in a research study. Before you and your child decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like any more information. Take time to decide whether or not you wish your child to take part.

Once you have read the information and if you are happy for your child to take part, please sign and return the attached consent forms provided in the information pack.

What is the purpose of the study?

Physical activity rates in children, particularly those with intellectual disabilities, are low. These low physical activity levels are a contributory factor to childhood obesity and health problems in later life. Accurately measuring physical activity in children is vital to increase our knowledge of their activity behaviours. This will allow for strategies to be developed to help increase activity levels and, ultimately, help improve overall health. Accelerometers are widely used to measure the physical activity behaviours of children. However, there are issues surrounding the accuracy of these devices. To try to ensure their accuracy, they can be calibrated during various exercises, whilst measuring the amount of energy a child expends. We can measure how much energy a child is using during activities by measuring the air they breathe out.

This research study is the first stage of a wider project aiming to investigate the use of accelerometers for the measurement of physical activity in children with intellectual disabilities. However, before we do this, we want to ensure that the best possible methods are used. Therefore, the aim of this study is to test the practicality of measuring expired air and aerobic fitness, and to investigate the use of various exercise activities.

Why has your child been chosen?

Your child has been invited to take part as they are aged between 8-14 years and attend a school or sports club in the Glasgow City area that caters specifically for children with learning disabilities. For this study, we aim to recruit 10 children that meet the above criteria.

Does your child have to take part?

No. The study is completely voluntary and it is up to you and your child to decide whether or not to take part. If your child does decide to take part, you will be given this information sheet to keep and you and your child will be asked to sign a consent form. Following consent, your child is still free to withdraw at any time, without any consequences, and without giving a reason.

What will happen to my child if he/she takes part?

If you and your child agree to take part, we will ask you attend an exercise laboratory at the University of Glasgow on two separate occasions. The laboratory sessions will be organised with you directly to arrange a day and time which suit you and your child. We will reimburse all travel expenses you incur throughout your child's participation in this study. Also, to thank your child for their time, after participation we will give them a £30 High Street gift voucher.

We estimate that both sessions will last approximately 2 hours and we invite you to be present in the laboratory during these sessions. The first session will initially involve familiarising your child with the laboratory and surrounding environment. We will show your child the equipment to be used and simply explain its purpose. Your child will then be given a training time in which they will practice walking on a treadmill and wearing the breath analysis equipment. This equipment involves your child wearing a mouthpiece, attached with a child-specific mask, which allows the collection of expired air through a tube attached to the mouthpiece. This equipment is not sore to wear, although it may initially feel strange. Your child will be asked to wear this during all the activities and the fitness test. Due to the nature of this equipment, we will show your child simple hand signals to help them communicate whilst wearing the mouthpiece, for example, thumbs up for 'okay'.

If we feel your child would benefit from an additional session to further practice using the equipment or to further familiarise them with the laboratory, we will invite them back before we conduct any data collection. If your child is comfortable with the equipment and walking on the treadmill, the latter part of this session will involve the first phase of data collection. We will initially take height and weight measurements and then conduct the treadmill-based activities. After a simple warm-up, your child will be asked to walk on the treadmill, at 4 different speeds (ranging from 3 to 8 km/h), each for 5 minutes. These speeds resemble a slow walk through to a gentle jog.

The second session will involve the final phase of activity data collection and the aerobic fitness test. For the first part of this session, your child will be asked to complete the following activities, each for 5 minutes: watching a DVD, drawing, passing a football, throwing/catching a ball, playing Xbox 360 Kinect, step aerobics, hula hoop, and jumping jacks.

The final part of this session will involve a treadmill-based graded exercise test to measure aerobic fitness level. Your child will choose a walking speed that they feel comfortable with, and then the gradient will increase by 2.5% every 2 minutes until your child feels too tired to continue.

It is important to note, however, that your child does not have to be fully-competent in all these activities to be able to participate in this study.

What will happen to my child if he/she takes part?

Outwith the laboratory sessions, participation in this study will have no wider implications for your child.

What are the possible disadvantages and risks of taking part?

This study has been designed to limit any risk to your child, with all procedures having been extensively conducted in previous studies and approved by the College Ethics Committee. All researchers involved are also experienced in this type of testing. The only foreseeable disadvantage for your child's participation in this study is the possible muscle fatigue in the day following testing. The only anticipated risks associated with this study involve the use of the treadmill and the fitness test. However, as previously noted, your child will have ample treadmill training time and will only participate in the data collection phase when the researchers are confident that they can safely walk on the treadmill. The treadmill also has hand rails which your child can use to further support themselves. As the fitness test involves your child exercising to a high intensity, it is possible they may feel slightly light-headed at the end of this test, although this should only last a few moments.

What are the possible benefits of taking part?

Although there are no direct benefits for your child's participation in this study, the information collected will help develop methods to better understand the physical activity behaviours of children with intellectual disabilities. From the information gathered, we will be able to provide you with information on your child's fitness level and their BMI. We also hope your child will have fun taking part.

Will my child's taking part in this study be kept confidential?

Yes. All information which is collected about your child during the course of the research will be kept strictly confidential. Your child will be identified by an ID number, and any information about your child will have their name removed so that they cannot be recognised from it.

What will happen to the results of the research study?

The primary use of the results from this study will be to inform the methods of a future study. If any of the findings are presented at a scientific conference or published in scientific literature, your child will not be identifiable from the results.

Who is organising and funding the research?

This study is being organised by researchers at the University of Glasgow and is funded by the Scottish Government Health and Social Care Directorate.

Who has reviewed the study?

This study has been reviewed and approved by the Medicine, Veterinary and Life Sciences College Ethics Committee, University of Glasgow.

Contact for Further Information

Arlene McGarty 1st floor Admin Building Mental Health & Wellbeing Gartnavel Royal Hospital 1055 Great Western Road Glasgow G12 0XH	Dr. Craig Melville 1st floor Admin Building Mental Health & Wellbeing Gartnavel Royal Hospital 1055 Great Western Road Glasgow G12 0XH	Victoria Penpraze 239 West Medical Building University of Glasgow Glasgow G12 8QQ
Email: a.mcgarty.1@research.gla.ac.uk Telephone: 0141 211 0210	Email: Craig.Melville@glasgow.ac.uk Telephone: 0141 211 3878	Email: Victoria.Penpraze@glasgow.ac.uk Telephone: 0141 330 2456

Thank you for taking the time to read this information sheet. If you have any questions, please do not hesitate to contact us.

If your child wishes to take part in this study, please sign the parent and child consent forms and fill out the contact details section. You can then either post these to us in the freepost envelope provided or you can return them to the school/sports club where your child received this information pack. We will then contact you to discuss your child's participation. If you or your child would like additional time to consider participation, please just let us know. Please keep this information sheet and a copy of you and your child's consent form.

Appendix iii – Feasibility study consent forms

University of Glasgow | College of Medical, Veterinary & Life Sciences

Centre Number:
Project Number:
Subject Identification Number for this trial:

CHILD CONSENT FORM

Title of Project: Investigation into exercise activities in children with intellectual disabilities.

Name of Researcher(s): Arlene McGarty
Dr. Craig Melville
Victoria Penpraze

Please initial box

I confirm that I have read and understand the information sheet dated __________ (version _____) for the above study and have had the opportunity to ask questions. ☐

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my legal rights being affected. ☐

I agree to take part in the above study. ☐

Your Name	Date	Signature
Parent / Guardian / Witness	Date	Signature
Researcher	Date	Signature

(1 copy for subject; 1 copy for researcher)

University of Glasgow | College of Medical, Veterinary & Life Sciences

Centre Number:
Project Number:
Subject Identification Number for this trial:

PARENT CONSENT FORM

Title of Project: Investigation into exercise activities in children with intellectual disabilities.

Name of Researcher(s): Arlene McGarty
Dr. Craig Melville
Victoria Penpraze

Please initial box

We confirm that we have read and understand the information sheet dated __________ (version ______) for the above study and have had the opportunity to ask questions. ☐

We understand that our child's participation is voluntary and that we are free to withdraw at any time, without giving any reason, without our legal rights being affected. ☐

We agree to take part in the above study. ☐

______________________	__________	______________________
Name of subject	Date	Signature
______________________	__________	______________________
Parent / Guardian / Witness	Date	Signature
______________________	__________	______________________
Researcher	Date	Signature

Contact Details

Parent Name . . . .

Child Name . . . .

Email . . . .

Telephone . . . .

(1 copy for subject; 1 copy for researcher)

Appendix iv – Calibration & cross-validation study ethical approval

University of Glasgow | College of Medical, Veterinary & Life Sciences

20 May 2014

Dr Craig Melville

Academic Unit for Mental Health & Wellbeing
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow G12 0XH

Dear Dr Melville

MVLS College Ethics Committee

Project Title: **Validation of the ActiGraph accelerometer in children with intellectual disabilities**
Project No: **200130128**

The College Ethics Committee has reviewed your application and has agreed that there is no objection on ethical grounds to the proposed study. It is happy therefore to approve the project, subject to the following conditions:

- Project end date: 31 January 2015.
- The research should be carried out only on the sites, and/or with the groups defined in the application.
- Any proposed changes in the protocol should be submitted for reassessment, except when it is necessary to change the protocol to eliminate hazard to the subjects or where the change involves only the administrative aspects of the project. The Ethics Committee should be informed of any such changes.
- You should submit a short end of study report to the Ethics Committee within 3 months of completion.

Yours sincerely

Professor William Martin
College Ethics Officer

Approval200130128.docx

Appendix v – Calibration & cross-validation study information sheets

University of Glasgow | College of Medical, Veterinary & Life Sciences

Child Information Sheet

Can you help?

We would like your help. Before you choose to help, we want you to know exactly what we're doing. Please read this and talk about it with your parents. Ask us if you have any questions.

Why do we want your help?

We want to look at different ways to measure how much physical activity children do. To be able to do this, we need children, like you, who are aged between 8 and 11 years to take part in a fun activity session.

What will happen when you take part?

We would like you to take part in a fun activity session at your school. This session will include lots of different games and activities, from gentle stretching to dancing. During the session we would like you to wear a special waistband which measures how much activity you do. We will also video record the session so we can see how active you are. The session will last about 45 minutes.

What else will we be doing?

We would also like to know how tall you are and how much you weigh. We will keep this information private.

We will measure your height like this:

We will measure your weight like this:

Are there any risks?

We have made this session as fun and risk-free as possible.

Will it help me?

This study is part of our university work. Although it won't help you directly, we hope that you will enjoy taking part.

Who will know I have taken part?

All information we get during the study will be kept very secret. If we want to show our work to other people, no one will know your name or that it was you who helped us. The video recordings of the session will only be used to measure how active you are. Only the researchers involved in the study will have access to it.

Will I be told the results?

Yes, if you want. We can send you a letter to let you know what we found.

Financial arrangements

Researchers at the University of Glasgow will run this study - we have been given some money to help us do this.

Do I have to take part?

No. You don't have to take part if you don't want to. Also, if you decide to take part and then change your mind, that's not a problem, you can stop at any point.

Who has checked the study?

This study has been checked over by the University of Glasgow College of Medical Veterinary and Life Sciences ethics committee to make sure it's safe for you to take part.

Appendix vi – Calibration & cross-validation study consent forms

Centre Number:
Project Number:
Subject Identification Number for this trial:

CHILD CONSENT FORM

Title of Project: Measuring physical activity in children with intellectual disabilities.

Name of Researcher(s): Arlene McGarty
Dr. Craig Melville
Victoria Penpraze

Please initial box

I confirm that I have read and understand the information sheet dated __________ (version _____) for the above study and have had the opportunity to ask questions. ☐

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my legal rights being affected. ☐

I agree to the use of video recording during the activity session. ☐

I agree to take part in the above study. ☐

Your Name	Date	Signature
Parent / Guardian / Witness	Date	Signature
Researcher	Date	Signature

University of Glasgow | College of Medical, Veterinary & Life Sciences

Centre Number:
Project Number:
Subject Identification Number for this trial:

PARENT CONSENT FORM

Title of Project: Measuring physical activity in children with intellectual disabilities.

Name of Researcher(s): Arlene McGarty
Dr. Craig Melville
Victoria Penpraze

Please initial box

We confirm that we have read and understand the information sheet dated ___________ (version _______) for the above study and have had the opportunity to ask questions. ☐

We understand that our child's participation is voluntary and that we are free to withdraw at any time, without giving any reason, without our legal rights being affected. ☐

We agree to the use of video recording during the activity session. ☐

We agree to take part in the above study. ☐

Name of participant	Date	Signature
Parent / Guardian / Witness	Date	Signature
Researcher	Date	Signature

Appendix vii – Publications arising from this thesis

Research in Developmental Disabilities 35 (2014) 973–981

Contents lists available at ScienceDirect

Research in Developmental Disabilities

Review article

Accelerometer use during field-based physical activity research in children and adolescents with intellectual disabilities: A systematic review

Arlene M. McGarty [a], Victoria Penpraze [b], Craig A. Melville [a,*]

[a] Institute of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, Scotland, United Kingdom
[b] School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, Scotland, United Kingdom

ARTICLE INFO

Article history:
Received 31 October 2013
Received in revised form 5 February 2014
Accepted 14 February 2014
Available online 12 March 2014

Keywords:
Physical activity
Accelerometers
Intellectual disabilities

ABSTRACT

Many methodological questions and issues surround the use of accelerometers as a measure of physical activity during field-based research. To ensure overall research quality and the accuracy of results, methodological decisions should be based on study research questions. This paper aims to systematically review accelerometer use during field-based research in children and adolescents with intellectual disabilities. Medline, Embase, Cochrane Library, Web of Knowledge, PsycINFO, PubMed, and a thesis database (up to May 2013) were searched to identify relevant articles. Articles which used accelerometry-based monitors, quantified activity levels, and included ambulatory children and adolescents (≤18 years) with intellectual disabilities were included. Based on best practice guidelines, a form was developed to extract data based on 17 research components of accelerometer use. The search identified 429 articles. Ten full-text articles met the criteria and were included in the review. Many shortcomings in accelerometer use were identified, with the percentage of review criteria met ranging from 12% to 47%. Various methods of accelerometer use were reported, with most use decisions not based on population-specific research. However, a lack of measurement research, e.g., calibration/validation, for children and adolescents with intellectual disabilities is limiting the ability of field-based researchers to make to the most appropriate accelerometer use decisions. The methods of accelerometer use employed can have significant effects on the quality and validity of results produced, which researchers should be more aware of. To allow informed use decisions, there should be a greater focus on measurement research related to children and adolescents with intellectual disabilities.

Contents

* Corresponding author at: Mental Health and Wellbeing, Institute of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, Gartnavel Royal Hospital, 1055 Great Western Road, Glasgow G12 0XH, Scotland, United Kingdom. Tel.: +44 0141 211 3878; fax: +44 0141 357 4899.
E-mail address: Craig.Melville@glasgow.ac.uk (C.A. Melville).

http://dx.doi.org/10.1016/j.ridd.2014.02.009
0891-4222/© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Physical activity can promote physical and mental health benefits for children and adolescents, including cardiovascular and musculoskeletal health (Janssen & LeBlanc, 2010; Strong et al., 2005), reductions in stress (Norris, Carroll, & Cochrane, 1992), improved concentration (Caterino & Polak, 1999) and maintenance of healthy weight (Dencker et al., 2006; Ness et al., 2007). The majority of children and adolescents do not meet the consensus physical activity guidelines that recommend at least 60 min of activity per day (Chief Medical Officers, 2011, US Department of Health & Human Services, 2008; World Health Organization, 2010). Furthermore, children with intellectual disabilities have been shown to be even less active than their typically developing peers (Borremans, Rintala, & McCubbin, 2010; Foley & McCubbin, 2009; Kozub, 2003).

The concern over low levels of physical activity has led to a research focus on quantifying activity to aid in the development of interventions to increase overall levels of physical activity (Brown, Hume, Pearson, & Salmon, 2013; Metcalf, Henley, & Wilkin, 2012). This has highlighted the need for accurate methods of measurement to assess the effectiveness of interventions and estimation of dose-response effects (Wareham & Rennie, 1998; Warren et al., 2010).

Physical activity can be measured via direct observation, subjective reports, and portable monitors (Troiano, 2005). The measurement of physical activity has been the topic of several review articles. These have included general (Reilly et al., 2008; Warren et al., 2010) and population-specific (Corder, Ekelund, Steele, Wareham, & Brage, 2008; Hinckson & Curtis, 2013) reviews of subjective and objective measures. The findings from these reviews suggest objective measures, such as accelerometers, as the preferred method for measuring physical activity.

In recent years, the use of accelerometers to measure physical activity has been ever-increasing (Freedson, Bowles, Troiano, & Haskell, 2012). Accelerometers are small devices which can measure acceleration of the body on up to three planes (vertical, mediolateral, anterior-posterior) during movement (Chen & Bassett, 2005). These acceleration signals are converted into quantifiable activity counts which can then be interpreted by equations or cut points (Hinckson & Curtis, 2013). This allows the quantification of activity without the high researcher burden required for direct observation or the high costs of doubly labelled water.

Viewed by many researchers as a gold standard measure, the use of accelerometers to measure physical activity in children and young people has been the topic of several review articles (Murphy, 2009; Youngwon, Beets, & Welk, 2012). However, the complexity of accelerometers poses many methodological choices and challenges for the physical activity researcher, e.g. in relation to protocols and data analysis (Ward, Evenson, Vaughan, Rodgers, & Troiano, 2005; Warren et al., 2010). Although the broad issues of measuring physical activity in children and adolescents with intellectual disabilities have been described (Hinckson & Curtis, 2013), no review to date has specifically examined the use of accelerometers in children and adolescents with intellectual disabilities.

This paper aimed to systematically review accelerometer use during field-based research in children and adolescents with intellectual disabilities. The objectives of this paper were to review (1) the methods of accelerometer use, and (2) the possible impact of accelerometer use decisions on results.

2. Method

The PRISMA guideline on systematic reviews was used as the basis for this review (Moher, Liberati, Tetzlaff, & Altman, 2009).

2.1. Search strategy

A literature search was conducted to establish papers that used accelerometry-based monitors to quantify physical activity in children and adolescents with intellectual disabilities. Six electronic databases were searched (Medline, Embase, Cochrane Library, Web of Knowledge, PsycINFO, and PubMed), from 1990 up to, and including, May 2013. This search was limited to 1990 onwards as accelerometers were still in a developmental stage during the late 1990s and were not widely used (Troiano, 2005). The search strategy focussed on truncated population (e.g., developmental disability, intellectual disability, learning disability, and mental retardation) and accelerometer terms (e.g., accelerometry, activity monitor, ActiGraph, ActivPal). Searches were limited to children and adolescents (0–18 years), English language, and human. For the full Embase search strategy, which was adapted for other databases, see Appendix A. Reference lists of retrieved studies and thesis databases were hand searched for additional relevant studies.

2.2. Study selection

Once potentially relevant articles were retrieved, the inclusion/exclusion criteria were applied to produce a final list of studies to be included in the review. The inclusion criteria were studies which:

- used accelerometry-based monitors to measure physical activity levels
- aimed to quantify levels of activity based on intensity, duration, or frequency in free-living settings
- included participants with intellectual disabilities
- included participants aged <18 years.

The exclusion criteria were studies which:

- aimed to calibrate/validate accelerometry-based monitors
- included a population with developmental disabilities, such as autism, but did not specify if intellectual disabilities were present
- included a non-ambulatory population.

Due to the limited research within this area, studies which met the design and measurement criteria, but only a portion of the sample met the population criteria, e.g., only some participants had intellectual disabilities, were included. Furthermore, as this review does not aim to investigate the impact of best practice guidelines on accelerometer use, studies meeting the above criteria, but published before the guidelines, i.e., pre-2005, were included.

2.3. Data extraction

Best practice guidelines relating to accelerometer use were developed at a scientific meeting held in 2004, entitled "Objective Monitoring of Physical Activity: Closing the Gaps in the Science of Accelerometry" (Ward et al., 2005). A data extraction form based on these best practice guidelines by Ward et al. (2005) was developed to obtain relevant data for review. The guidelines were based around five research themes, with 18 specific components to be considered when designing and conducting physical activity research using accelerometers. However, one research component, analysing data, was not included in the criteria as it was deemed too specific to monitor calibration and therefore not relevant to the aims of this review. Based on this, 17 research components were used; these are presented in Table 1 with a summary of each specific research component and review criteria. The summary and criteria were formulated primarily from the information presented by Ward et al. (2005), however, the specific research articles were used for clarification and to ensure the accuracy of the guidelines reported. Data was extracted based on whether studies met the review criteria for each research component.

3. Results

3.1. Study selection

Overall, 428 references were retrieved from the search of electronic databases (Medline, $n = 16$; Embase, $n = 14$; Cochrane Library, $n = 0$; Web of Knowledge, $n = 341$; Psycinfo, $n = 37$; PubMed, $n = 20$), and one article was retrieved through a thesis database search. After title and abstract screening, thirty full text articles were assessed. Twenty articles did not meet the inclusion criteria due to inappropriate population or the collection of physical activity data to measure variability and reliability, not to quantify activity levels during field-based research. Based on this, ten articles met the inclusion criteria and were included in the review. The full search and selection strategy is presented in Fig. 1.

3.2. Study characteristics

Eight studies were carried out in the United States, one in Australia, and one in the United Kingdom. Participants ($N = 677$) aged from three to 70 years, with reported intellectual disabilities ranging from mild to severe. All studies, with the exception of Phillips and Holland (2011) (12–70 years), included only child and adolescent participants (three-17 years). Seven studies quantified overall physical activity during daily free-living (Esposito, MacDonald, Hornyak, & Ulrich, 2012; Foley, Bryan, & McCubbin, 2008; Lloyd, 2008; Phillips & Holland, 2011; Shields, Dodd, & Abblitt, 2009; Ulrich, Burghardt, Lloyd, Tiernan, & Hornyak, 2011; Whitt-Glover, O'Neill, & Stettler, 2006), two studies quantified school recess and classroom

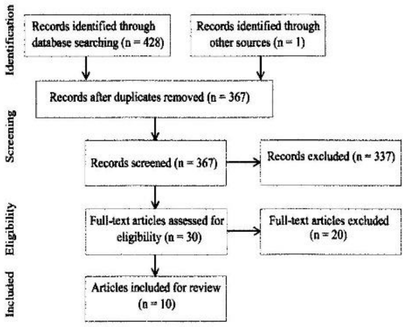


Fig. 1. Flow chart of study selection process.

based physical activity (Horvat and Franklin, 2001Lorenzi, Horvat, & Pellegrini, 2000), and one quantified after school physical activity (Foley & McCubbin, 2009). Seven accelerometer brands were used: ActiGraph (Phillips & Holland, 2011), Actical (Esposito et al., 2012; Lloyd, 2008; Ulrich et al., 2011), Actitrac (Whitt-Glover et al., 2006), Actiwatch (Foley & McCubbin, 2009; Foley et al., 2008), Caltrac (Lorenzi et al., 2000), TriTrac R3D (Horvat & Franklin, 2001), and RT3 (Shields et al., 2009).

The ratings of the ten papers against the best practice guidelines are presented in Table 2.

3.3. *Monitor selection, quality, and dependability*

Five studies clearly reported selecting an accelerometer based on their research question (see Table 2). Three studies presented a general rationale for the benefits of accelerometry use, citing the capability of the device to detect movement across various planes (Esposito et al., 2012; Horvat & Franklin, 2001) and advantages of accelerometry over other methods of physical activity assessment, e.g., measurement not affected by stress or self-report bias, and non-compliance can be easily quantified (Shields et al., 2009). Two studies referenced previous use of accelerometry in children with intellectual disabilities as a rationale for accelerometer use (Foley & McCubbin, 2009; Foley et al., 2008). However, no study stated any measure of instrument quality or dependability that was established in children and adolescents with intellectual disabilities.

3.4. *Monitor use protocols*

Nine studies used a single accelerometer, without providing a rationale for doing so. Lorenzi et al. (2000) used two monitors to enable comparison between constant and individualised methods of device programming relating to participant characteristics, but only used one device for data analysis. Accelerometer monitoring periods varied from 16 min (Horvat & Franklin, 2001; Lorenzi et al., 2000), to 4 days (Lloyd, 2008), 5 days (Foley & McCubbin, 2009), and 7 days (Esposito et al., 2012; Foley et al., 2008; Phillips & Holland, 2011; Shields et al., 2009; Ulrich et al., 2011; Whitt-Glover et al., 2006), with only two studies providing a rationale for wearing days, based on previous use (Foley et al., 2008) and reliability (Shields et al., 2009).

The hip was the most popular placement, with four studies attaching the accelerometer to the right hip (Esposito et al., 2012; Lloyd, 2008; Phillips & Holland, 2011; Ulrich et al., 2011), one the right and left hip (Lorenzi et al., 2000), and three studies not specifying right or left hip (Horvat & Franklin, 2001; Shields et al., 2009; Whitt-Glover et al., 2006). Two studies placed the device on the wrist of the participant's non-dominant hand, in line with manufacturer's recommendations (Foley & McCubbin, 2009; Foley et al., 2008).

Foley and McCubbin (2009) and Foley et al. (2008) reported establishing field practices by replacing the accelerometer twice during the monitoring period to reduce measurement error associated with individual devices. Furthermore, Foley et al. (2008) reported testing device calibration before and after data collection. Strategies employed to aid compliance were the use of daily monitoring logs (Esposito et al., 2012; Phillips & Holland, 2011; Shields et al., 2009; Ulrich et al., 2011), familiarisation and orientation sessions (Horvat & Franklin, 2001; Lorenzi et al., 2000), and instructing parents to regularly check monitor wear and placement (Whitt-Glover et al., 2006).

3.5. *Monitor calibration*

Of the studies which aimed to predict energy expenditure, five used equations or cut points that had been calibrated against indirect calorimetry, with only one study (Whitt-Glover et al., 2006) not providing sufficient evidence of a gold standard measurement for calibration. However, no study used equations that were specific for the study population, i.e., that had been calibrated specifically for children or adolescents with intellectual disabilities. The studies which only reported physical activity data in counts (Foley & McCubbin, 2009; Foley et al., 2008; Lorenzi et al., 2000) and counts with vector magnitude (Horvat & Franklin, 2001) all stated the error associated with energy expenditure calculations as a rationale.

Table 1
Best practice guidelines (Ward et al., 2005), summary, and review criteria.

Best practice guidelines	Summary	Review Criteria
Monitor selection, quality, and dependability		
1. Selecting instruments	With the variability of available accelerometers, e.g., in size, cost, data processing and data storage, and with no device viewed as superior, choice of device should be based on research purpose.	Rationale provided for choice of device.
2. Assessing instrument quality and dependability	Instrument variance (i.e., coefficient of variability), reliability and validity should be tested before and after use.	Population-appropriate coefficient of variability, validity and/or reliability evidence provided for device used.
Monitor use protocols		
3. Using multiple monitors	Additional information and accuracy of using multiple monitors should be considered in relation to study population and participant burden.	Rationale provided for number of monitors used.
4. Defining wearing days	Days of monitoring required should be based on population-specific calculations, e.g. based on ICC, with setting, resources, and research question considered.	Seven days of monitoring required or appropriate justification for a shorter monitoring period provided.
5. Determining monitor placement	Placement should be decided based on existing calibration equations, e.g., which placements do pre-existing equations exist for the study population, participant comfort, and manufacturer recommendation should be considered when deciding placement.	Rationale provided for monitor placement.
6. Establishing field practices	Quality control measures should be employed throughout the period of accelerometer use, and during distribution and collection.	Inter-unit variation controlled and/or face-to-face accelerometer distribution and collection employed.
7. Ensuring compliance	Investigator- or participant-based compliance strategies can encourage accelerometer wear.	At least one of the following compliance techniques employed: log diary, reminder calls, information on proper wear, relapse prevention model, visual prompts, wear information given to teachers, coaches, etc., participant shown example of data output which indicates when device is not worn, or incentives offered.
Monitor calibration		
8. Predicting energy expenditure	Equations employed should be calibrated against a gold-standard measure and should account for different patterns of activity.	Cut points or equations calibrated against gold standard measure of energy expenditure (calorimetry, doubly-labelled water or direct observation) during various patterns of activity.
9. Using individual calibration equations	In small scale studies, individual calibration equations should be used.	Individual calibration equations developed for each participant.
10. Constructing group calibration equations	In larger-scale studies, equations should be used which were calibrated in similar population with a representative sample, with population appropriate activities conducted.	Equations calibrated in population matched for participant characteristics during various activities.
11. Determining epoch length	Epoch length should be determined based on the study population and their activity characteristics, e.g., children generally engage in short bouts of high intensity activity.	Rationale provided for epoch length used.
Analysis of accelerometer data		
12. Defining a day	The time period of monitoring required to constitute a day can vary, e.g., from 12 to 24 h, with participant age, weekday/weekend monitoring, and activities to be considered.	Definition of what constitutes a day of measurement for inclusion in analysis.
13. Handling incomplete data	Activity is not always measured over a consistent time period and data can be missing, therefore decisions should be made to try to prevent under/overestimation of activity.	Method specified for dealing with missing or incomplete data, e.g., imputation.
14. Creating reporting standards	Due to a lack of standardised procedures for data processing and reduction, decision rules need to be clearly stated.	Clarity of data collected and assumptions made, e.g., identifying wearing period, minimum wear required for valid day, spurious data, computing variables and aggregating days, extracting bouts of MVPA

Table 1 (*Continued*)

Best practice guidelines	Summary	Review Criteria
15. Determining bouts	Reporting bouts of activity can allow for more accurate estimation of MVPA and understanding/ comparison of activity patterns.	Measurement of total activity duration and number and length of bouts per day.
16. Handling spurious data	Data should be cleaned for implausible data points with accelerometers checked for malfunction, error, or participant tampering.	Comment made on how spurious data is dealt with, e.g., setting data points to "missing".
Integration with other data sources		
17. Integration with other data sources	Use of multiple technologies can increase the quality and breadth of data collected.	Use of other data source, e.g., HR or GPS.

The most commonly used epoch length was 15 s (Esposito et al., 2012; Foley & McCubbin, 2009; Foley et al., 2008; Lloyd, 2008; Ulrich et al., 2011), with the use of five second (Phillips & Holland, 2011), 30 s (Whitt-Glover et al., 2006), and 1 min epochs also reported (Horvat & Franklin, 2001; Lorenzi et al., 2000; Shields et al., 2009). With the exception of two studies (Esposito et al., 2012; Lloyd, 2008), which both chose a 15 s epoch based on children's activity tempo and behaviours, no other study stated a rationale for epoch length used.

3.6. *Analysis of accelerometer data*

To be included in analysis, three studies specified a minimum of 10 h of wear per day (Esposito et al., 2012; Phillips & Holland, 2011; Ulrich et al., 2011), with Esposito et al. (2012) and Ulrich et al. (2011) further specifying that this had to achieved on four out of seven days, including one weekend day. Shields et al. (2009) did not report a required number of hours of wear per day, but specified that six out of seven days of monitoring was required. For dealing with missing or

Table 2
Standard of accelerometer use based on best practice guidelines.

Best Practice Guidelines	Study									
	Esposito et al. (2012)	Foley & McCubbin (2009)	Foley et al. (2008)	Horvat & Franklin (2001)	Lloyd (2008)	Lorenzi et al. (2000)	Phillips & Holland (2011)	Shields et al. (2009)	Ulrich et al. (2011)	Whitt-Glover et al. (2006)
Monitor selection, quality, and dependability										
1. Selecting instruments	•	•	•	•	–	–	–	•	–	–
2. Assessing instrument quality and dependability	–	–	–	–	–	–	–	–	–	–
Monitor use protocols										
3. Using multiple monitors	–	–	–	–	–	•	–	–	–	–
4. Defining wearing days	•	–	•	–	–	–	•	•	•	–
5. Determining monitor placement	–	–	–	–	–	–	–	–	–	–
6. Establishing field practices	–	•	•	–	–	–	–	–	–	–
7. Ensuring compliance	•	–	–	•	–	•	•	•	•	•
Monitor calibration										
8. Predicting energy expenditure	•	N/A	N/A	N/A	•	N/A	•	•	•	–
9. Using individual calibration equations	–	N/A	N/A	N/A	–	N/A	–	–	–	–
10. Constructing group calibration equations	–	N/A	N/A	N/A	–	N/A	–	–	–	–
11. Determining epoch length	•	–	–	–	•	–	–	–	–	–
Analysis of accelerometer data										
12. Defining a day	•	–	–	•	–	•	•	•	•	–
13. Handling incomplete data	–	–	•	–	–	–	•	–	–	•
14. Creating reporting standards	•	•	–	–	–	–	•	•	•	•
15. Determining bouts	–	N/A	N/A	N/A	–	N/A	•	•	–	•
16. Handling spurious data	–	•	–	–	–	–	•	–	–	–
Integration with other data sources										
17. Integration with other data sources	–	•	–	•	–	•	–	–	–	–

• Study met the best practice guidelines for specific research component.
– Study did not meet the best practice guidelines for specific research component.
N/A Research component not applicable to study.

incomplete data, two studies (Foley et al., 2008; Whitt-Glover et al., 2006) excluded the measurements from any participants who returned incomplete data, whereas one study (Phillips & Holland, 2011) asked participants to wear the accelerometer for an additional seven days.

Of the studies which met the criteria for creating reporting standards, four noted wear time (Esposito et al., 2012; Phillips & Holland, 2011; Shields et al., 2009; Ulrich et al., 2011), which ranged from 12.38 to 14.23 h per day. Two studies also noted the average number of wear days as 6.3 days (Whitt-Glover et al., 2006) and 6.8 days (Shields et al., 2009) for a seven-day monitoring period. Two studies (Foley & McCubbin, 2009; Phillips & Holland, 2011) reported checks for spurious data, with Phillips and Holland (2011) further specifying that non-wear time was defined as ten or more minutes of zero counts.

3.7. *Integration with other data sources*

To increase the breadth of data collected, two studies combined accelerometry with the SOAL observational checklist and heart-rate measurements (Horvat & Franklin, 2001; Lorenzi et al., 2006) and one study used a proxy activity log based on the Activitygram to determine which activities children engaged in and when they did so (Foley & McCubbin, 2009).

4. Discussion

This review illustrates the variance of accelerometer use in physical activity research for children and adolescents with intellectual disabilities. A variety of methods were reported in each of the five themes, with an array of use decisions described within each research component. However, the majority of these did not meet the criteria based on best practice guidelines (Ward et al., 2005). A lack of measurement specific research involving children and adolescents with intellectual disabilities was also highlighted. This is limiting the scope for decisions to be made based on population-specific research, therefore negatively impacting on the reliability and validity of results.

Although there is an abundance of devices that enable the quantification of physical activity, there is no gold standard method that will accurately measure all activity components (Warren et al., 2010). For this reason, it is the responsibility of the researcher to choose the most appropriate device and methods of data collection and data analysis, based on study research questions (Trost, McIver, & Pate, 2005).

There was a clear lack of identified studies that met the criteria for best practice guidelines. The percentage of review criteria met ranged from 12% (Lloyd, 2008) to 47% (Phillips & Holland, 2011), which highlights the variability and overall scarcity of research being conducted in line with best practice guidelines. A contributing factor to this is that many studies consistently failed to provide a rationale for accelerometer use decisions. These use decisions, however, can have considerable research implications; for example, choice of cut points or epoch used can produce significantly ($p < .01$) different results (Nilsson, Ekelund, Yngve, & Sjöström, 2002; Reilly et al., 2008). Within this review, the use of five different sets of cut points was reported, with no study providing an appropriate rationale. Furthermore, four different epoch lengths were used, with only two studies providing an appropriate rationale for this decision. To limit the effect these decisions could have on the accuracy of results, the use of appropriate and valid methods is vital.

Many studies in this review, however, failed to report appropriate validity evidence as justification for decisions, e.g., in relation to: monitor selection, placement, epoch, and group calibration equations employed. Of the studies that did report validity evidence, none of this was established in children and adolescents with intellectual disabilities. It is important that validity is established in the situation and population of interest, as generalising validity coefficients can have a negative impact on research quality (Lincare, 2000; Yun & Ulrich, 2002). Previous reviews of physical activity in children and adolescents with intellectual disabilities have also noted this lack of reporting population appropriate validity evidence (Cervantes & Porretta, 2010; Hinckson & Curtis, 2013). With closer examination of the research, however, it is apparent that there is a distinct lack of physical activity research focusing on validating devices and components of accelerometer use in children and adolescents with intellectual disabilities. To date, only one study has aimed to validate accelerometry in youth with intellectual disabilities, but found a weak relationship between counts and the prediction of percent heart rate (Esposito, 2012). This therefore limits the ability of the physical activity researcher to make informed decisions on accelerometer use.

Within physical activity research, a distinction can be made between the measurement researcher and the end user (Freedson et al., 2012). The measurement researcher has a focus on measurement science and the specifics of accelerometry, e.g., validation/calibration, whereas the end users interests lie in the outcome measures of accelerometry in relation to surveillance, intervention and epidemiological based research. Although this review focuses on the accelerometer use of end users, many of the shortcomings noted, e.g., relating to the lack of validity evidence in this population, is in fact due to the deficiency of measurement research conducted in children and adolescents with intellectual disabilities. Furthermore, this lack of measurement research may limit the awareness of end users to the effect these decisions have on the research conducted.

To illustrate this, within the theme of monitor calibration (an area which is the focus of the measurement researcher), no studies employed calibration equations or cut points that were established in children and adolescents with intellectual disabilities. All studies which reported an outcome measure based on energy expenditure, with the exception of Whitt-Glover et al. (2006), specified the use of cut points that had been established in a different population, but against a gold standard measure of indirect calorimetry. However, due to variability between populations, for example children and adolescents with intellectual disabilities have lower reported levels of fitness (Pitetti, Yarmer, & Fernhall, 2001), the generalising of these cut-points could result in levels of physical activity being misreported.

The practical limitation of directly measuring energy expenditure in field-based research results in the use of a proxy measure, such as accelerometry, to quantify activity. Assumptions relating to energy expenditure are therefore based on the interpretation of accelerometry data, which is primarily done through the use of equations and cut points. However, due to many of the difficulties previously discussed, recent research suggests the use of cut points should cease, although no valid alternative is currently available (Freedson et al., 2012). This lack of measurement-specific research is limiting the ability of end users to produce results based on valid measures, and is resulting in methodologies which may have a negative impact on the reliability of findings described in the literature.

Best practice guidelines, and the criteria developed within this review, provide end users with simple recommendations on the many methodological decisions related to accelerometer use. Although it may not be feasible to meet all of these guidelines, end users should be aware of the implications of their decisions and ensure the most appropriate decisions are made, where possible, with these decisions accurately reported in publications. To enable this, it is also recommended that measurement researchers validate specific components of accelerometer use in children and adolescents with intellectual disabilities.

This systematic review is the first to focus on accelerometer use in children and adolescents with intellectual disabilities. Although more recent recommendations have been published (Freedson et al., 2012), the specificity of accelerometer use guidelines presented within Ward et al. (2005) was most relevant to the aims of this review. This allowed the development of simple criteria which not only enabled investigation into accelerometer use, but could also guide researchers on the importance of decision making and reporting.

The wide scope of this review – in relation to the number of research components included – has prevented a more thorough review and discussion on accelerometer use and the research implications of use decisions. The criteria developed, although a useful summary for end users, is considerably generalised in comparison to the depth of recommendations presented within the best practice guidelines (Ward et al., 2005).

5. Conclusion

From this review, many shortcomings have been identified in accelerometer use for children and adolescents with intellectual disabilities. The lack of measurement specific research is enforcing limitations on intervention and epidemiological research, thus raising questions on the reliability and validity of results. With the abundance of methodological questions facing the physical activity researcher and variance of accelerometer use reported, investigation into the effect of these use decisions – although problematic – is vital for informing future accelerometer use and monitoring protocols.

Conflicts of interest

None declared.

Appendix A. Embase search strategy

1. (developmental adj (disab$ or disorder or difficult$)).tw.
2. (intellectual adj (disab$ or disorder or difficult$)).tw.
3. (learning adj (disab$ or disorder or difficult$)).tw.
4. (mental$ adj (retard$ or deficiency)).tw.
5. Acceleromet*.tw.
6. Accelerometry-based monitors.tw.
7. Physical activity measurement.tw.
8. Activity monitor.tw.
9. ActiGraph.tw.
10. MTI.tw.
11. CSA.tw.
12. Actical.tw.
13. Actiheart.tw.
14. Activpal.tw.
15. Tritrac.tw.
16. Uniaxial.tw.
17. Dualaxial.tw.
18. Triaxial.tw.
19. MVPA.tw.
20. Or/1–4
21. Or/5–19
22. 20 and 21

Limit 22 to (human and English language and child <unspecified age>.

References

Borremans, E., Rintala, P., & McCubbin, J. A. (2010). Physical fitness and physical activity in adolescents with Asperger syndrome Asperger S: A comparative study. *Adapted Physical Activity Quarterly, 27*(4), 308–320.

Brown, H., Hume, C., Pearson, N., & Salmon, J. (2013). A systematic review of intervention effects on potential mediators of children's physical activity. *BMC Public Health, 13*, 165.
Caterino, M. C., & Polak, E. D. (1999). Effects of two types of activity on the performance of second-, third-, and fourth-grade students on a test of concentration. *Perceptual and Motor Skills, 89*(1), 245–248.
Cervantes, C. M., & Porretta, D. L. (2010). Physical activity measurement among individuals with disabilities: A literature review. *Adapted Physical Activity Quarterly, 27*(3), 173–190.
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), S490–S500.
Chief Medical Officers. (2011). *Start active, stay active: A report on physical activity for health from the four home countries'*, Chief Medical Officers. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_127931. Accessed 03.01.13.
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.
Dencker, M., Thorsson, O., Karlsson, M. K., Lindén, C., Eiberg, S., Wollmer, P., et al. (2006). Daily physical activity related to body fat in children aged 8–11 years. *Journal of Pediatrics, 149*(1), 38–42.
Esposito, P. M. (2012). *Validating accelerometry and skinfold measures in youth with Down syndrome*, MI, USA: University of Michigan. (Unpublished doctoral dissertation).
Esposito, P. M., MacDonald, M., Hornyak, J., & Ulrich, D. A. (2012). Physical activity patterns of youth with Down syndrome. *Intellectual and Developmental Disabilities, 50*(2), 109–119.
Foley, J. T., & McCubbin, J. A. (2009). An exploratory study of after-school sedentary behaviour in elementary school-age children with intellectual disability. *Journal of Intellectual and Developmental Disability, 34*(1), 3–9.
Foley, J. T., Bryan, R. R., & McCubbin, J. A. (2008). Daily physical activity levels of elementary school-aged children with and without mental retardation. *Journal of Developmental and Physical Disabilities, 20*(4), 365–378.
Freedson, P., Bowles, H. R., Troiano, R., & Haskell, W. (2012). Assessment of physical activity using wearable monitors: Recommendations for monitor calibration and use in the field. *Medicine and Science in Sports and Exercise, 44*(1), S1–S4.
Hinckson, E. A., & Curtis, A. (2013). Measuring physical activity in children and youth living with intellectual disabilities: A systematic review. *Research in Developmental Disabilities, 34*(1), 72–86.
Horvat, M., & Franklin, C. (2001). The effects of the environment on physical activity patterns of children with mental retardation. *Research Quarterly for Exercise and Sport, 72*(2), 189–195.
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.
Kozub, F. M. (2003). Explaining physical activity in individuals with mental retardation: An exploratory study. *Education and Training in Developmental Disabilities, 28*, 302–313.
Lincare, J. M. (2000). New approaches to determining reliability and validity. *Research Quarterly for Exercise and Sport, 71*(2), 129–136.
Lloyd, M. C. (2008). *Physical activity of preschool age children with and without Down syndrome*, MI, USA: University of Michigan. (Unpublished doctoral dissertation).
Lorenzi, D. G., Horvat, M., & Pellegrini, A. D. (2000). Physical activity of children with and without mental retardation in inclusive recess settings. *Education and Training in Mental Retardation and Developmental Disabilities, 35*(2), 160–167.
Metcalf, B., Henley, W., & Wilkin, T. (2012). Effectiveness of intervention on physical activity of children: Systematic review and meta-analysis of controlled trials with objectively measured outcomes. *British Medical Journal, 345*, e6320.
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*(4), 264–269.
Murphy, S. L. (2009). Review of physical activity measurement using accelerometers in older adults: Considerations for research design and conduct. *Preventive Medicine, 48*(2), 108–114.
Ness, A. R., Leary, S. D., Mattocks, C., Blair, S. N., Reilly, J. J., Wells, J., et al. (2007). Objectively measured physical activity and fat mass in a large cohort of children. *PLoS Medicine, 4*(3), e97.
Nilsson, A., Ekelund, U., Yngve, A., & Sjöström, M. (2002). Assessing physical activity among children with accelerometers using different time sampling intervals and placements. *Pediatric Exercise Science, 14*(1), 87–96.
Norris, R., Carroll, D., & Cochrane, R. (1992). The effects of physical activity and exercise training on psychological stress and wellbeing in an adolescent population. *Journal of Psychosomatic Research, 36*(1), 55–65.
Phillips, A. C., & Holland, A. J. (2011). Assessment of objectively measured physical activity levels in individuals with intellectual disabilities with and without Down's syndrome. *PLoS ONE, 6*(12), e28618.
Pitetti, K. H., Yarmer, D. A., & Fernhall, B. (2001). Cardiovascular fitness and body composition of youth with and without mental retardation. *Adapted Physical Activity Quarterly, 18*(2), 127–141.
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.
Shields, N., Dodd, K. J., & Abblitt, C. (2009). Do children with Down syndrome perform sufficient physical activity to maintain good health? A pilot study. *Adapted Physical Activity Quarterly, 26*(4), 307–320.
Strong, W. B., Malina, R. M., Blimkie, C. J. R., Daniels, S. R., Dishman, R. K., Gutin, B., et al. (2005). Evidence based physical activity for school-age youth. *Journal of Pediatrics, 146*(6), 732–737.
Troiano, R. P. (2005). A timely meeting: Objective measurement of physical activity. *Medicine and Science in Sports and Exercise, 37*(11), S487–S489.
Trost, S. G., McIver, K. L., & Pate, R. R. (2005). Conducting accelerometer-based activity assessments in field-based research. *Medicine and Science in Sports and Exercise, 37*(11), S531–S543.
Ulrich, D. A., Burghardt, A. R., Lloyd, M., Tiernan, C., & Hornyak, J. E. (2011). Physical activity benefits of learning to ride a two-wheel bicycle for children with Down syndrome: A randomized trial. *Physical Therapy, 91*(10), 1463–1473.
US Department of Health and Human Services. (2008). *Physical activity guidelines for Americans*, Hyattsville, MD: US Department of Health and Human Services.
Ward, D. S., Evenson, K. R., Vaughan, 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), S582–S588.
Wareham, N. J., & Rennie, K. L. (1998). The assessment of physical activity in individuals and populations: Why try to be more precise about how physical activity is assessed? *International Journal of Obesity and Related Metabolic Disorders, 22*(2), S30–S38.
Warren, J. M., Ekelund, U., Besson, H., Mezzani, A., Geladas, N., & Vanhees, L. (2010). Assessment of physical activity – A review of methodologies with reference to epidemiological research: A report of the exercise physiology section of the European Association of Cardiovascular prevention and rehabilitation. *European Journal of Preventive Cardiology, 17*(2), 127–139.
Whitt-Glover, M. C., O'Neill, K. L., & Stettler, N. (2006). Physical activity patterns in children with and without Down syndrome. *Pediatric Rehabilitation, 9*(2), 158–164.
World Health Organization. (2010). *Global recommendations on physical activity for health*, Geneva: World Health Organization.
Yun, J., & Ulrich, D. (2002). Estimating measurement validity: A tutorial. *Adapted Physical Activity Quarterly, 19*(1), 32–47.
Youngwon, K., Beets, M. W., & Welk, G. J. (2012). Everything you wanted to know about selecting the "right" accelerometer cut-points for youth, but...: A systematic review. *Journal of Science and Medicine in Sport, 15*(4), 311–321.

McGarty *et al. Pilot and Feasibility Studies* (2015) 1:18
DOI 10.1186/s40814-015-0014-2

PILOT AND FEASIBILITY STUDIES

RESEARCH Open Access

Feasibility of a laboratory-based accelerometer calibration protocol for children with intellectual disabilities

Arlene M. McGarty[1], Victoria Penpraze[2] and Craig A. Melville[1*]

Abstract

Background: Accelerometry has not been calibrated for the estimation of physical activity in children with intellectual disabilities (ID), raising questions regarding the validity of interpreting accelerometer data in this population. Various protocols and criterion measures have been used in calibration studies involving typically developing (TD) children; however, the suitability of these activities and measures for children with ID is unknown. Therefore, this study aimed to test the feasibility of a laboratory-based calibration protocol for children with ID. Specifically, the feasibility of activities, measurements, and recruitment was investigated.

Methods: Five children with mild to moderate ID (10.20 ± .98 years) and a comparative sample of five TD children (12.40 ± .01 years) participated in this study. Participants performed a free-living and treadmill-based activity protocol during two laboratory-based sessions. Activities were performed for 5 min and ranged from sedentary to vigorous intensity. Treadmill activities ranged from 3 to 8 km/h, and free-living activities included watching a DVD, passing a football, and jumping jacks. Resting energy expenditure was measured, and a graded exercise test was used to assess cardiorespiratory fitness. Breath-by-breath respiratory gas exchange and accelerometry were continually measured during all activities. Feasibility was assessed using observations, activity completion rates, and respiratory data.

Results: All TD participants and one participant with ID completed the protocol. The physical demands of the treadmill activities affected the completion rate for participants with ID. No participant met the maximal criteria for the graded exercise test or attained a steady state during the resting measurements. Limitations were identified with the usability of respiratory gas exchange equipment and the validity of measurements. The school-based recruitment strategy was not effective, with a participation rate of 6 %. A significant ($z = 13.21$, $p < .0001$) difference in the relationship of $\dot{V}O_2$ and accelerometry was identified between ID and TD participants.

Conclusions: Due to issues with the usability and validity of breath-by-breath respiratory gas exchange and recruitment, a laboratory-based calibration protocol is currently not feasible for children with ID. An alternative field-based protocol with a non-invasive criterion measure should be considered for future studies.

Keywords: Physical activity, Accelerometer, Calibration, Intellectual disabilities, Children

* Correspondence: Craig.Melville@glasgow.ac.uk
[1]Institute of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, Gartnavel Royal Hospital, 1055 Great Western Road, Glasgow G12 0XHScotland, UK
Full list of author information is available at the end of the article

Background

Accelerometers are a widely used objective measure that enables the quantification of physical activity levels. Accelerometers measure raw acceleration of the body in gravitational units (g), which is converted into an arbitrary unit (count). Counts can then be used to estimate physical activity, such as energy expenditure or time spent in moderate to vigorous intensity, through the application of prediction equations or cut points. These cut points and equations are developed by calibrating activity counts against a known biological measure, which is a form of validity-based research referred to as "value calibration" [1].

However, there is limited research investigating the validity and reliability of accelerometers for the estimation physical activity in children with ID [2]. This is impacting on the ability of researchers to accurately measure physical activity levels within this population [3]. Accelerometer counts have previously been validated in children with ID [4, 5], although no calibration studies have been conducted. To ensure the validity of measurements in children with ID, the development of population-specific cut points is urgently needed.

Calibration is a complex process, and there are many challenges in deriving a biological meaning from a raw biomechanical measure of acceleration. Furthermore, additional difficulties face calibration involving children, due to the relationship between energy expenditure and body mass and the influence of maturation [6]. In a review of accelerometer calibration in children, Freedson et al. [6] suggested that the development of population-specific cut points could limit the effect of these biological differences.

When differences between typically developing (TD) children and children with ID are considered, the validity of generalizing cut points is questionable [7]. For example, children with ID have lower reported levels of cardiorespiratory fitness than their TD peers [8–10]. Due to the effect that cardiorespiratory fitness has on energy expenditure and oxygen uptake ($\dot{V}O_2$) during activity, the generalizing of cut points calibrated in a population with higher fitness could lead to an underestimation or misclassification of activity intensity for a population with lower fitness. Due to validity issues such as these, Freedson et al. [11] discussed the importance of investigating and classifying fitness for calibration studies, which will provide information on relative activity and health.

The protocol of a calibration study fundamentally involves the concurrent measurement of accelerometry and a gold standard measure of activity. The use of a biological criterion measure of energy expenditure, specifically indirect calorimetry measured through respiratory gas exchange, is deemed most appropriate and has been widely used in studies involving children [12]. A laboratory-based protocol should include at least six free-living and/or treadmill-based activities which are representative of activities conducted by the study population [1, 2]. Previous calibration studies have employed a wide range of activities, including treadmill speeds ranging from 4 to 10 km/h, watching a DVD, playing catch, hopscotch, basketball dribbling, and martial arts exercises [13–15]. However, no standardized guidelines exist for the type of activities to be conducted or the amount of time for which these activities should be conducted.

To ensure the appropriateness of a calibration protocol for children with ID, the feasibility of the methods of measurement and the activity protocol should initially be tested. Respiratory gas exchange is not a measurement which has been widely conducted in children with ID. Furthermore, many of the activities used in previous studies involve sport-specific skills, co-ordination, and concentration and therefore may not be suitable for this population. Similarly, Oortwijn et al. [16] tested the feasibility of calibration protocol in a whole-room calorimeter in a sample of five children.

The aims of this study are to 1) test the feasibility of recruiting children from additional support needs (ASN) schools to a laboratory-based study, 2) test the feasibility of activities and measures, 3) test the feasibility of using breath-by-breath respiratory gas exchange, and 4) compare the relationship between accelerometry and $\dot{V}O_2$ between ID and TD participants.

Methods

Ethical consideration

This study was approved by the Medical, Veterinary, and Life Sciences College Ethics Committee, University of Glasgow. Written informed consent was obtained from both the participants and parents.

Protocol

This study was conducted in three phases over two laboratory sessions: 1) familiarization, 2) preparation, and 3) data collection. The familiarization and preparation phases were conducted during session one. During the familiarization phase, the participant was introduced to the laboratory environment, with the aim of reducing participant anxiety in the data collection phase. The preparation phase allowed the participant to become familiar with the equipment and procedures and to ensure they were physically capable of safely completing the protocol. Data collection was conducted over the two sessions and consisted of anthropometric and resting energy expenditure (REE) measurements, treadmill-based and free-living activities, and a graded exercise test. Only participants with ID completed the REE measurements. Throughout the protocol, activity was measured using accelerometry and $\dot{V}O_2$ measured through breath-by-breath respiratory gas exchange.

The protocol used within this study is described in Table 1. The activity protocols of previous laboratory-based calibration studies in TD children, which included both treadmill and free-living activities, informed the development of the present protocol (Appendix). The activities chosen ranged from sedentary to vigorous intensity, were treadmill-based and free-living, and were skill and non-skill specific. This variety and number of included activities will enable the most appropriate and effective activities to be included in a full-scale calibration study.

Table 1 Protocol of measurements and activities performed

Activity	Description
Session 1	
Rest	Sitting in reclined position watching DVD
Treadmill-based activities	
Light intensity	
3 km/h	Walking at 3 km/h at zero gradient
Moderate intensity	
6 km/h	Jogging at 6 km/h at zero gradient
5 km/h at 5 %	Walking briskly at 5 km/h at 5 % gradient
Vigorous intensity	
8 km/h	Running at 8 km/h at zero gradient
Session 2	
Free-living activities	
Sedentary	
Sitting playing computer game	Sitting playing handheld Nintendo DS
Watching DVD	Sitting watching DVD
Drawing	Sitting drawing
Light intensity	
Passing football	Passing a football with a researcher
Playing catch	Standing throwing/catching a ball with a researcher
Standing playing computer game	Standing playing handheld Nintendo DS
Moderate intensity	
Step aerobics	Continual stepping on and off aerobic step
Hula hoop	Continual twirling of hula hoop around the waist
Interactive computer game	Playing interactive bowling on an Xbox Kinect
Vigorous intensity	
Jumping jacks	Continual jumping jacks/star jumps
Graded exercise test	Treadmill-based incremental fitness test

Recruitment strategy

Participants with ID were recruited from ASN schools in Glasgow, Scotland. A researcher visited two schools, explained the study to children, and handed out information packs. If children were interested in participating, parents were asked to return a parent and child consent form to the researcher to allow discussion regarding participation. A convenience sample of TD children was recruited from the Glasgow area. All participant and parent travel expenses were reimbursed. Furthermore, children received a £30 high street voucher after completion of the study.

The exclusion criteria for participation were as follows: i) having a physical disability, ii) being non-ambulatory, or iii) being outwith the age range of 8–14 years.

Measures

Anthropometric

Height was measured to the nearest 0.1 cm using a stadiometer (Seca Scales, Hamburg, Germany), and weight was measured to the nearest 0.1 kg using digital scales (Seca Scales, Hamburg, Germany). Measurements were conducted twice to produce a mean value whilst participants were wearing light clothing and no shoes.

Resting energy expenditure

Respiratory gas exchange was measured continually for 15 min to allow REE to be established. Throughout this measurement, participants sat in a reclined position and watched an age-appropriate DVD.

Activities

Participants were asked to complete four treadmill and ten free-living activities, each for 5 min (Table 1). These types of activities have been extensively conducted in calibration studies involving TD children [12] and were based on previously defined intensity classifications [15]. Prior to the activities, participants completed a 2-min treadmill-based warm-up. Additionally, rest periods were given between activities to allow measurements to return to within a resting range.

Graded exercise test

Participants walked on the treadmill at a constant and self-selected pace. The gradient was increased from zero in increments of 2.5 % every 2 min until the participant was unable to continue. Concurrent with previous protocols in adolescents with ID [17], walking speed was self-determined. The use of self-determined pace for exercise testing in individuals with disabilities may prevent the test being discontinued due to participant anxiety [18].

The primary criterion for the attainment of maximal oxygen consumption ($\dot{V}O_{2max}$) was a plateau in $\dot{V}O_2$

with increased workload. A secondary criteria of an increased respiratory exchange ratio (RER) >1.0 and a heart rate within 10 bpm of an age-adjusted estimate of maximal heart rate were also used [19, 20]. Predicted maximal heart rate was calculated using the equation: maximal heart rate = 220 – age.

Instrumentation

Accelerometry

The ActiGraph wGT3X+ (ActiGraph LLC, Pensacola, FL, USA) is a small triaxial accelerometer which measures acceleration of the body across the vertical, horizontal, and perpendicular axes during movement. Prior to the session, the accelerometers were initialized in accordance with manufacturer specifications. The device was worn around the waist, positioned at the hip (at the iliac crest). In line with manufacturer guidelines for use in children, the device was attached using an elastic belt.

Heart rate

Heart rate was measured using a chest-worn heart rate monitor (Vantage, Polar Electro). The sensor was attached directly to the skin using an elastic belt, and measurements (beats per minute (bpm)) were recorded on the device receiver which was held by the researcher. Heart rate was recorded every minute during the graded exercise test and at the termination of the test.

Respiratory gas exchange

Respiratory gas exchange was measured using the Ultima CPX (Medical Graphics, MN, USA) which analyses expired gases on a breath-by-breath basis. Prior to each test, airflow, ventilatory volume, and gas analysers were calibrated using standard measures in accordance with manufacture guidelines. Participants wore a preVent (Medical Graphics, MN, USA) material mask which covered their nose and mouth. This was attached directly to a bidirectional flow meter, a sampling line, and measurement sensor. Data were initially recorded using standard threshold settings of minimum 50 mL $\dot{V}O_2$ and carbon dioxide production ($\dot{V}CO_2$), minimum 180 mL tidal volume, and RER between 0.5 and 2.6. These standard threshold settings are specific to the Ultima CPX system and are used to reduce error in the measurements; however, thresholds can be altered or removed when data are downloaded.

Management of data

Respiratory gas exchange data were initially downloaded using standard threshold settings. In this unaveraged format, there were periods of missing data; it was hypothesized that these measurements were outwith the threshold settings and therefore excluded. Data were additionally downloaded with no threshold settings applied to allow comparison. Data were time averaged into 10-s intervals, to reduce variability and random error [21]. Time averaging data reduced the number of missing data points. The remaining missing data with the standard threshold settings were imputed from the data with no threshold settings, which had no missing data points.

Oxygen uptake (mL/kg/min) was extracted for each activity, RER during the graded exercise test, and $\dot{V}CO_2$ (mL/min) during the measurement of REE. As steady-state measurements provide a more valid representation of the respiratory and metabolic requirements of activity, only minutes 2–4 for each activity was included in the analysis [22]. The final minute of data was excluded from the analysis as some participants became fatigued and agitated towards the end of the 5-min measurement period.

Accelerometer data were sampled at a rate of 30 Hz and was post-processed and reduced to 10-s epochs of data. This duration of epoch was chosen because of the intermittent movements used within the free-living activities, where a shorter epoch will more accurately capture sporadic activity [23]. Data were downloaded into Excel where count data for all activities and measures was extracted. As vertical axis and vector magnitude data are used for calibration, only these measurements were included in the analysis. Accelerometer data were then time matched to the corresponding 10-s epoch of respiratory data. Accelerometer and $\dot{V}O_2$ data were organized for total activity, with $\dot{V}O_2$ data additionally organized for each individual participant. Accelerometer data are presented as counts per 10-s epoch (counts/10-s).

Statistical analysis

All statistical data were analysed using SPSS 21 IBM statistical package (SPSS IBM, New York, NY, USA). Normality was assessed for all variables. For data that were not normally distributed, logarithmic and square root transformations were separately applied to the data and normality was retested. If transformations were not effective in producing normally distributed data, non-parametric tests were used.

Descriptive statistics were calculated for age, sex, height, weight, and body mass index (BMI), with means, standard deviations (SD), and 95 % confidence intervals (95 % CI) reported. Additionally, independent two-sample t tests were used to compare differences in these variables between ID and TD participants. The feasibility of activities and REE was assessed from observations and percentage completion rates. The attainment of steady state, defined as a coefficient of variation <10 %, was additionally used to test the feasibility of measuring REE. The feasibility of the graded exercise test was based on attainment of a $\dot{V}O_{2max}$ score, with differences between ID and TD participants investigated using independent two-sample t tests. The effect of threshold settings was investigated using dependent two-sample t tests or the Wilcoxon signed-rank test for data that were not normally distributed. The relationship between $\dot{V}O_2$ and counts was investigated using Spearman's correlation coefficients and z-scores.

Results

Participant characteristics

Five children with mild to moderate ID (four males, one female) and five TD children (one male, four females) participated in this study. Descriptive statistics are presented in Table 2.

Recruitment

Seventy-eight children with ID met the inclusion criteria and received an information pack. Ten (12.82 %) initial consent forms were returned, which resulted in a final participation rate of 6 %. Of the parents who returned consent forms but whose child did not participate, reasons given were the need to travel to the laboratory and insufficient time to organize two sessions.

Activities and measures

Resting energy expenditure

Four participants completed the REE measurement for 15 min. One participant became agitated due to wearing the mask and did not complete the measurement. No participant achieved a steady state. The mean coefficient of variation for $\dot{V}O_2$ and $\dot{V}CO_2$ for the final 10 min was 24.38 and 28.61 %, respectively. As illustrated by the higher mean $\dot{V}CO_2$ scores compared to $\dot{V}O_2$, this measurement caused participants to hyperventilate.

Treadmill-based activities

Activity completion rates varied greatly between ID and TD participants. Only one participant with ID completed all activities for the required time. In comparison, all TD participants completed the protocol. The physical demands of the treadmill speeds prevented participants with ID from attempting or completing activities.

Table 2 Descriptive statistics for all participants

Description	ID	TD	95 % CI
Age (years)	10.20 ± 1.10	12.40 ± 1.12*	−3.83, −.57
Weight (kg)	36.60 ± 8.08	47.52 ± 5.78*	−21.16, −.68
Height (m)	1.45 ± .08	1.60 ± .09*	−.27, −.03
BMI (kg/m²)	17.34 ± 2.85	18.68 ± 2.50	−5.24, 2.57

Data are presented as means, standard deviations, and 95 % confidence intervals for between-group differences
*Significantly ($p < .05$) different from participants with ID

Feedback from participants with ID was that the treadmill-based activities were most enjoyable. An interesting observation was the views of parents', who were present throughout, in relation to their children's ability to complete the treadmill-based activities. In general, parents underestimated the competence and ability of their child to complete activities. For example, one parent suggested that her child did not participate in the 6 or 8 km/h activities as she had never seen him run and assumed he was not capable of doing so; however, this participant (ID2) completed the 6 km/h activity for 5 min and 8 km/h for 3.5 min.

Free-living activities

Three participants with ID completed all free-living activities. Two participants with ID opted out of activities which they did not perceive to be enjoyable. Table 3 shows the activities completed and not completed by participants with ID.

Graded exercise test

Four participants with ID performed the graded exercise test, and one participant opted out due to fatigue. All TD participants performed the test, however, due to a system error, no respiratory gas exchange measurements were

Table 3 Activities completed by each participant with intellectual disabilities

Activities and measures	ID 1	ID 2	ID 3	ID 4	ID 5
Session 1					
REE	-	-	-	X	-
Treadmill activities					
3 km/h at 0 %	-	-	-	-	-
6 km/h at 0 %	-	-	-		X
5 km/h at 5 %	-	-	-	-	X
8 km/h at 0 %	-	X	-	X	X
Session 2					
Free-living activities					
Sitting playing DS	-	-	-	-	-
Watching DVD	-	-	-	-	-
Drawing	X	X	-	-	-
Passing football	-	-	-	-	-
Throw/catch	-	X	-	-	-
Standing playing DS	-	-	-	-	-
Step aerobics	-	-	-	-	-
Hula hoop	-	X	-	-	-
Xbox	-	X	-	-	-
Jumping jacks	-	X	-	-	-
Graded exercise test	-	X	-	-	-

- participant completed the activity for the required 5 min, X participant did not complete the activity recorded for TD1. Individual test data are presented in Table 4. No participant met the primary criteria of a plateau in $\dot{V}O_2$; therefore, results are presented as the peak scores attained during the test. There were no significant differences between ID and TD participants for $\dot{V}O_2$ (mL/kg/min), $t(6) = 1.30$, $p > .05$; 95 % CI, −14.67, 4.52, HR (bpm), $t(7) = -1.61$, $p > .05$; 95 % CI, −50.67, 9.67, or RER, $t(6) = -1.08$, $p > .05$; 95 % CI, −.25, .98.

Breath-by-breath respiratory gas exchange

Usability issues were identified with the respiratory gas exchange equipment. This was due to participant anxiety and the weight of the mask when the bidirectional flow meter and sampling line were attached. All participants expressed anxiety about wearing the mask, although only during the longer duration measure of REE did anxiety affect one participant's ability to complete the activity. The level of anxiety experienced due to the mask was high; three participants recorded respiratory exchange ratios greater than one, indicating hyperventilation, and one became very upset. Methods employed to reassure participants who were experiencing higher levels of anxiety was a researcher talking to them and a researcher also wearing a mask. However, reported anxiety caused by the mask reduced as the session progressed.

During dynamic movements, the weight of the breath-by-breath valve attached to the mask caused it to slip down, leaving the nose and mouth partially uncovered. All participants were asked to wear a nose clip to limit this effect on the amount of expired gas captured, but no participant agreed to the nose clip. To prevent the mask coming off or slipping off the nose, a researcher held the sample line to reduce the weight the mask had to support. The preVent (Medical Graphics, MN, USA) mask used was the smallest size, and no alternative masks were suitable.

Investigation into the effect of threshold settings showed that $\dot{V}O_2$ was significantly higher when no thresholds were applied for both ID ($z = -12.43$, $p < .001$, $r = -.27$) and TD participants ($z = -4.29$, $p < .001$, $r = -.09$). Furthermore, when the effect of threshold settings was examined on an individual level, ID participants had a greater variance than TD participants, with percentage change scores ranging from −7.63 to 14.61 % and −.39 to .74 %, respectively.

Accelerometry and $\dot{V}O_2$

There was a significant difference in the relationship between $\dot{V}O_2$ and vertical axis counts ($z = 13.21$, $p < .0001$) and $\dot{V}O_2$ and vector magnitude counts ($z = 14.23$, $p < .0001$) between ID and TD participants. For participants with ID, $\dot{V}O_2$ (mean = 11.64, SD = 7.93 mL/kg/min) was significantly correlated with vertical axis (mean = 200.45, SD = 374.35 counts/10 s; $r_s = .70$, $p < .001$) and vector magnitude counts (mean = 328.42, SD = 433.45 counts/10 s; $r_s = .73$, $p < .001$), with no significant differences between these correlation coefficients ($z = -1.79$, $p > .05$). Similarly, $\dot{V}O_2$ (mean = 13.89, SD = 8.41 mL/kg/min) was significantly correlated with both vertical axis (mean = 334.25, SD = 529.47 counts/10 s; $r_s = .29$, $p < .001$) and vector magnitude counts (mean = 479.20, SD = 569.18 counts/10 s; $r_s = .31$, $p < .001$) for TD participants, with no significant differences between correlation coefficients ($z = .59$, $p > .05$).

Discussion

The purpose of this study was to investigate the feasibility of a laboratory-based accelerometer calibration protocol

Table 4 Individual scores attained during graded exercise test

Participant	$\dot{V}O_{2peak}$ (mL/kg/min)	HR_{peak} (bpm)	Age-predicted HR_{max} (bpm)	RER_{peak}	Speed (km/h)	Time (min)
ID1	34.30	195	209	1.06	5.50	21.50
ID3	30.40	172	209	0.92	5.00	9.00
ID4	25.70	152	211	0.97	4.00	16.00
ID5	31.60	135	211	0.89	3.50	7.00
TD 1	-	177	208	-	6.00	20.50
TD 2	29.40	172	209	1.04	5.50	15.50
TD 3	43.80	199	206	1.21	6.00	14.00
TD 4	30.20	179	208	0.96	6.00	12.00
TD 5	38.90	193	207	0.94	5.50	14.00
ID mean	30.50	163.50	215.00	0.96	4.50	13.38
(SD)	(3.59)	(25.88)	(6.93)	(.07)	(.91)	(6.65)
TD mean	35.57	184.00	207.60	1.04	5.80*	15.20
(SD)	(6.97)	(11.45)	(1.14)	(.12)	(.27)	(3.21)

*TD participants significantly ($p < .05$) different from ID participants in children with ID. This section will discuss the results in relation to the four feasibility aims of this study.

Recruitment

Our initial aim was to recruit ten participants to both the ID and TD groups. Although school-based recruitment strategies have been effective for recruiting TD children to exercise-related studies [24], the school-based strategy used within this study was ineffective, with low response and recruitment rates. The initial response rate of 12.82 % is notably lower than the 83 % response rate reported by Oortwijn and colleagues [16]. There is limited research relating to effective recruitment strategies for children with ID; however, for adults with ID, recruitment strategies involving direct contact with participants are most effective [25]. Furthermore, recruitment of adults with ID is lower for studies involving more invasive measures and physical tests [26].

Small sample sizes and an over-representation of boys are common limitations in health-related research involving children with ID [27]. The over-representation of boys within this study could be partially attributed to the higher prevalence of ID in boys compared to girls [28]. Additionally, as boys generally participate in more physical activity than girls, the activity-focussed protocol may have been of less interest to girls, which could have further limited their recruitment. Another important consideration for the low recruitment rate is the time and travel demands of the study, which were noted by parents as factors which prevented participation. The development of a shorter, single-session protocol would reduce the time requirements of participation, which could subsequently benefit recruitment; however, it is important to consider the effect this could have on participants, such as levels of anxiety and fatigue, and the influence this could have on the quality of data collected.

Despite these recruitment difficulties, it is important that pilot testing is conducted, even with smaller sample sizes, to ensure the feasibility of protocols and measures in this population [29]. Although five participants is not a suitable number for a calibration study, this sample size can still provide meaningful findings relating to the feasibility of calibration protocols and measurements [16]. In future studies, the low response rate needs to be accounted for, although the inclusion of a greater number schools and service organizations could provide the required number of participants for a full-scale calibration study. It is also important to increase our understanding of why girls with ID are frequently underrepresented and to develop girl-focussed recruitment strategies. However, as invasive measures and physical tests are a barrier for adults with ID [26], further investigation is needed to determine whether this low recruitment rate was a direct result of an ineffective recruitment strategy or whether this type of study is one in which children and specifically girls with ID did not want to participate.

Activities and measures

Resting energy expenditure

REE was the first physiological measurement conducted within this protocol. Although participants were given a practice time using the respiratory gas exchange equipment during the preparation phase, wearing the mask for this extended measurement caused increased anxiety and hyperventilation. This affected the attainment of a steady state, which optimizes results and is primarily important for resting metabolic measures [30]. REE is required for deriving metabolic equivalent (MET), therefore necessary in studies aiming to derive prediction equations for activity energy expenditure. However, studies aiming to calibrate intensity cut points do not require a measure of REE, although Freedson et al. [6] noted that activity METs should still be presented in all calibration studies. REE can be approximated through age-specific estimates, therefore, a direct measurement is not essential for the estimation of METs. Based on the difficulties identified within this study for the direct measurement of REE in children with ID, the use of age-specific estimates is deemed most appropriate.

Treadmill-based activities

The treadmill speeds were not appropriate for this sample, as the physical demands were too high to allow completion of 5 min for all activities. Previous studies have aimed to ensure the suitability of activities by proposing speeds per age group or within a range of speeds. Puyau et al. [15] included vigorous activities that were age specific; furthermore, Puyau et al. [31] proposed moderate and vigorous activities within speeds of 3.5–4 mph and 4.5–7 mph, respectively. The moderate (5 km/h at 5 % and 6 km/h) and vigorous (8 km/h) speeds in this current study are within the lower ranges used within Puyau et al. [15] and nearest to the vigorous speed for 8–10 years within Puyau et al. [31]. The high completion rate for TD participants could therefore be due to age, as they were significantly older than ID participants. However, three ID participants were still unable to complete the treadmill speeds which were deemed age appropriate in these previous studies, suggesting slower speeds should be used within this population. Employing a range of speeds or age-specific speeds for children with ID could therefore increase the rate of completion.

Free-living activities

In contrast, generalizing free-living activities to a calibration study involving children with ID may be more appropriate. The higher completion rates for the free-living activities could be partially due to the intensity not being fixed, as in the treadmill-based activities. Participants were able to complete activities at an intensity that was comfortable for them and could intermittently stop when fatigued. Although this could have a positive effect on completion, it is important that activities are at least semi-structured to ensure participants reach the desired intensity for the purposes of calibration. Due to the limited time that children spend in moderate and vigorous intensity during free play, the use of unstructured activities could negatively impact on the calibration of higher intensity activity [29, 32].

Graded exercise test

No participant with ID reached $\dot{V}O_{2max}$, which could be due to a number of factors. Firstly, test duration ranged from 7–21.5 min. It is suggested that an exercise test should be completed within 8–12 min to prevent premature termination of the test due to localized muscle fatigue, rather than the attainment of $\dot{V}O_{2max}$ [33]. Additionally, the protocol of incrementally increasing gradient can cause calf muscle and lower back discomfort, which limits the participant's ability to continue with a test [34]. From observation, as the gradient increased, participants became unstable which may have further contributed to the termination of the test before $\dot{V}O_{2max}$.

The attainment of $\dot{V}O_{2max}$ can be difficult in children who have no prior experience of strenuous exercise and the physical effects and discomfort associated with an exercise test [35]. Furthermore, as none of the sample in this study had prior experience of a treadmill, an alternative test could limit this effect. Field-based tests, including the 1-mile walk test [36], 600-yard walk/run test and 20-m and modified 16-m shuttle run test [37], have shown reliability and concurrent validity in children with ID. These tests could therefore be considered within a calibration protocol as an alternative to a maximal test.

Breath-by-breath respiratory gas exchange

Measurement issues were identified with the use of the Ultima CPX breath-by-breath system (Medical Graphics, MN, USA), specifically in relation to the use of thresholds. As $\dot{V}O_2$ is a criterion measure, it is essential that this measurement is accurate to ensure the quality of calibration and prevent systematic error. However, there is currently no universally accepted method for processing breath-by-breath $\dot{V}O_2$ data, which is impacting on the validity of data processing and interpretation [21]. Freedson et al. [6] discussed the complexity and difficulties associated with using and interpreting a biological criterion measure in children and suggested that a behavioural measure, specifically direct observation, was an effective alternative criterion measure.

Generalizing cut points is partially based on the assumption that the relationship between $\dot{V}O_2$ and accelerometer counts is the same between groups. However, a significant ($p < .0001$) difference in the relationship of counts and $\dot{V}O_2$ between ID and TD participants was identified in this study. Therefore, the prediction of intensity classification for ID children based on TD cut points will introduce systematic error and validity issues. Although the validity of generalizing cut points between TD and ID children has been previously discussed [7], this is the first study which aimed to compare the fundamental relationship between $\dot{V}O_2$ and counts. This further supports the need for a calibration study to be conducted specifically in a population of children with ID; however, the feasibility issues identified need further consideration before a full-scale laboratory-based calibration study.

Strengths and limitations

This was the first study which aimed to address the lack of population-specific cut points for children with ID. Furthermore, to ensure the suitability and effectiveness of a gold standard laboratory-based protocol, this study investigated the feasibility of activities and measurements. The wide range of treadmill and free-living activities included within this study increased our knowledge relating to the appropriateness of generalizing activities used with TD children to children with ID. The findings from this study can therefore be used to inform the development of an appropriate protocol.

The design of the present protocol had an effect on completion rates and data collection, as participants became fatigued during the latter stages of the session, in particular the second session. Therefore, the conclusions regarding the appropriateness of activities could be affected by study design. It is also important to note that as all hypothesis testing is based on a small sample size, results should be interpreted with caution. Although the small sample size highlights the difficulties with recruitment, it prevented direct comparison with a group of matched TD participants which would have enabled further investigation into physiological differences. Furthermore, as no in-depth data was collected regarding the aetiology of ID, it is not possible to discuss possible effects of ID type or severity on participation and activity completion rates.

Conclusions

Findings from this study suggest that the methods used within a calibration protocol for TD children cannot be fully generalized to children with ID. Although additional research is required before definitive conclusions can be made regarding feasibility, initial methodological recommendations for the design of a calibration study involving children with ID are as follows: 1) Treadmill activities should not be generalized from protocols involving TD children; instead, speeds should be self-selected or age-appropriate speeds developed. 2) Free-living activities, which can be successfully generalized from TD protocols, should be incorporated due to the high completion rates. 3) REE and $\dot{V}O_{2max}$ should be estimated using validated non-invasive methods. In terms of future research, it is recommended that the suitability and validity of breath-by-breath respiratory gas exchange measurements is further investigated. Furthermore, an effective recruitment strategy has to be developed, and reasons for the low recruitment rate of girls need to be better understood.

Until the limitations identified within this study have been addressed, the use of a laboratory-based calibration protocol is not feasible for children with ID. As these limitations are specific to a laboratory-based protocol, consideration should therefore be given to an alternative protocol. A field-based calibration study which is conducted in the participants' environment, e.g. school, and which uses a non-invasive criterion measure, such as direct observation, could be an effective alternative to a laboratory-based study.

Appendix

Table 5 Activity protocols of previous child calibration studies which included treadmill and free-living activities

Study	Participant age range (years)	Treadmill activities	Time per treadmill activity (min)	Free-living activities	Time per free-living activity (min)
Eston et al. [13][a]	8–10	Walk 4 km/h	4	Playing catch	4
		Walk 6 km/h	4	Hopscotch	4
		Run 8 km/h	4	Sitting crayoning	4
		Run 10 km/h	4		
Evenson et al. [14][b]	6–8	Light		Sedentary	
		Walk 2 mph	7	Rest	15
		Moderate		Sitting watching DVD	7
		Walk 3 mph	7	Sitting colouring books	7
		Vigorous		Moderate	
		Run 4 mph	7	Stair climb	7
				Dribble basketball	7
				Vigorous	
				Cycle on stationary bike	7
				Jumping jacks	7
McMurray et al. [35][b]	8–18	Walk 4 km/h	10	Standing arcade game	10
		Walk 5.6 km/h	10	Stretching	10
		Run 8 km/h	10	Sweeping	10
				Vacuuming	10
				Shovelling	10
				Stair climb (88 steps/min)	10
				Rope skipping	10
Puyau et al. [15][c]	6–16	Light		Sedentary	
		Walk 2.5 mph	10	Rest	20
		Moderate		Sitting playing Nintendo	20
		Walk 3.5 mph (6–7 years)	10	Arts and crafts	20
		Walk 4 mph (8–16 years)	10	Free play sitting, e.g. cards, lego, miniature cars	20
		Vigorous		Light	
		Jog 4.5 mph (6–7 years)	10	Aerobic warm up	10
		Jog 5 mph (8–10 years)	10	Moderate	

Table 5 Activity protocols of previous child calibration studies which included treadmill and free-living activities *(Continued)*

		Jog 6 mph (11–16 years)	10	Tae Bo martial arts exercises	10
				Free play standing, e.g. hula hoop, throwing ball, jumping jacks	10
				Vigorous	
				Jump rope	3
				Walk—self-selected pace	5
				Skip	3
				Jog—self-selected pace	3
				Soccer	3
Puyau et al. [28][c]	7–18	Walk 2 mph	7	Rest	30
		Walk 3.5–4 mph	7	Sitting playing handheld Nintendo	20
		Jog/run 4.5–7 mph	7	Sitting playing computer	20
				Cleaning (dusting)	10
				Aerobic exercises	12
				Ball toss	10
Vanhelst et al. [23][a]	10–16	Light	15 min consecutive	Sedentary	15 min consecutive
		Walk 1.5 km/h, 3 % gradient		Lying in bed watching TV	
		Moderate		Sitting reading	
		Walk 3 km/h, 3 % gradient		Sitting playing computer game	
		Run 4 km/h, 3 % gradient		Light	
				Standing drawing	
		Vigorous		Passing football	
		Run 6 km/h, 3 % gradient			

[a]Respiratory gas exchange was measured using a stationary metabolic system, which restricts movement within a laboratory environment
[b]Respiratory gas exchange was measured using a portable metabolic system, which allows almost unrestricted movement within a designed indoor or outdoor area
[c]Respiratory gas exchange was measured using a whole room calorimeter, which allows unrestricted body movement within a confined room

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
AMM, VP, and CAM were involved in the study design, data collection, and editing of the manuscript. AMM also conducted the statistical analysis and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements
This research was funded by the Scottish Government's Equally Well Fund.

Author details
[1]Institute of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, Gartnavel Royal Hospital, 1055 Great Western Road, Glasgow G12 0XHScotland, UK. [2]School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, Scotland, UK.

Received: 30 December 2014 Accepted: 14 May 2015
Published online: 24 May 2015

References

1. Welk GJ. Principles of design and analysis for the calibration of accelerometry-based activity monitors. Med Sci Sport Exer. 2005;37 Suppl 11:501–11.
2. Hinckson EA, Curtis A. Measuring physical activity in children and youth living with intellectual disabilities: a systematic review. Res Dev Disabil. 2013;34:72–86.
3. McGarty AM, Penpraze V, Melville CA. Accelerometer use during field-based physical activity research in children and adolescents with intellectual disabilities: a systematic review. Res Dev Disabil. 2014;35:973–81.
4. Capio CM, Sit CH, Abernethy B. Physical activity measurement using MTI (Actigraph) among children with cerebral palsy. Arch Phys Med Rehabil. 2010;91:1283–90.
5. Esposito PM. Validating accelerometry and skinfold measures in youth with Down syndrome. University of Michigan: PhD thesis; 2012.
6. Freedson P, Pober D, Janz KF. Calibration of accelerometer output for children. Med Sci Sport Exer. 2005;37 Suppl 11:523–30.
7. Frey GC, Stanish HI, Temple VA. Physical activity of youth with intellectual disability: review and research agenda. Adapt Phys Act Q. 2008;25:95–117.
8. Horvat M, Franklin C. The effects of the environment of physical activity patterns of children with mental retardation. Res Q Exercise Sport. 2001;72:189–95.
9. Pan CY, Frey GC. Physical activity patterns in youth with autism spectrum disorders. J Autism Dev Disord. 2006;36:597–606.
10. Pitetti KH, Yarmer DA, Fernhall B. Cardiovascular fitness and body composition of youth with and without mental retardation. Adapt Phys Act Q. 2001;18:127–41.
11. Freedson P, Bowles HR, Troiano R, Haskell W. Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. Med Sci Sport Exer. 2012;44 Suppl 1:1–4.
12. Youngwon K, Beets MW, Welk GJ. Everything you wanted to know about selecting the "right" Actigraph accelerometer cut-points, but…: a systematic review. J Sci Med Sport. 2012;15:311–21.

13. Eston RG, Rowlands AV, Ingledew DK. Validity of heart rate, pedometry, and accelerometry for predicting the energy cost of children's activities. J Appl Physiol. 1998;84:362–71.
14. Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG. Calibration of two objective measures of physical activity for children. J Sport Sci. 2008;26:1557–65.
15. Puyau MR, Adolph AL, Vohra FA, Butte NF. Validation and calibration of physical activity monitors in children. Obes Res. 2002;10:150–7.
16. Oortwijn AWMC, Plasqui G, Reilly JJ, Okely AD. Feasibility of an activity protocol for young children in a whole room indirect calorimeter: a proof-of-concept study. J Phys Activ Health. 2009;6:633–7.
17. Pitetti KH, Jongmans B, Fernhall B. Feasibility of a treadmill test for adolescents with multiple disabilities. Adapt Phys Act Q. 1999;16:362–71.
18. American College of Sports Medicine. ACSM's guidelines for exercise testing and prescription. 9th ed. Philadelphia: Lippincott Williams & Wilkins; 2013.
19. Howley ET, Bassett DR, Welch HG. Criteria for maximal oxygen uptake: review and commentary. Med Sci Sport Exer. 1995;27:1292–301.
20. Powers SK, Howley ET. Exercise physiology, theory and application to fitness and performance. 6th ed. New York: McGraw Hill; 2007.
21. Robergs RA, Dwyer D, Astorino T. Recommendations for improved data processing from expired gas analysis indirect calorimetry. Sports Med. 2010;40:95–111.
22. Compher C, Frankenfield D, Kim N, Roth-Yousey L. Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review. J Am Diet Assoc. 2006;106:881–903.
23. Vanhelst J, Beghin L, Rasoamanana P, Theunynck D, Meskini T, Iliescu C,

List of References

ActiGraph. (2012a). *Knowledge base: what's the difference among the cut points available in ActiLife?* Retrieved from https://help.theactigraph.com/entries/21452826-What-s-the-difference-among-the-Cut-Points-available-in-ActiLife-

ActiGraph. (2012b). *Low frequency extension explained*. Retrieved from https://help.theactigraph.com/entries/21767838-Low-Frequency-Extension-Explained

Adolph, A. L., Puyau, M. R., Vohra, F. A., Nicklas, T. A., Zakeri, I. F., & Butte, N. F. (2012). Validation of uniaxial and triaxial accelerometers for the assessment of physical activity in preschool children. *Journal of Physical Activity and Health*, *9*(7), 944-953.

Ahn, S., & Fedewa, A. L. (2011). A meta-analysis of the relationship between children's physical activity and mental health. *Journal of Pediatric Psychology, 36*(4), 385-397.

American College of Sports Medicine. (2013). *Guidelines for exercise testing and prescription*. Philadelphia: Lippincott Williams and Wilkins.

Anderson, C. B., Hagstromer, M., & Yngve, A. (2005). Validation of the PDPAR as an adolescent diary: effect of accelerometer cut points. *Medicine and Science in Sports and Exercise, 37*(7), 1224-1230.

Astrand, P. O. (1986). Why exercise? An evolutionary approach. *Acta Medica Scandinavica*, *220*(711), S241-S242.

Astrand, P. O. (1994). Physical activity and fitness: evolutionary perspective and trends for the future. In C. Bouchard, R. J. Shepard, & T. Stephens (Eds.), *Physical activity, fitness, and health* (pp. 98-105). Champaign, IL: Human Kinetics.

Badger, D., Nursten, J., Williams, P., & Woodward, M. (2000). Should all literature reviews be systematic? *Evaluation and Research in Education, 14*(3-4), 220-230.

Bailey, R. C., Olson, J., Pepper, S. L., Porszasz, J., Barstow, T. J., & Cooper, D. M. (1995). The level and tempo of children's physical activities: an observational study. *Medicine and Science in Sports and Exercise, 27*(7), 1033-1041.

Balady, G. J., Arena, R., Sietsema, K., Myers, J., Coke, L., Fletcher, G. F.,…Milani, R. V. (2010). Clinician's guide to cardiopulmonary exercise testing in adults: a scientific statement from the American Heart Association. *Circulation, 120*, 191-225.

Balogh. R. S., Lake, J. K., Lin, E., Wilton, A., & Lunsky, Y. (2015). Disparities in diabetes prevalence and preventable hospitalizations in people with intellectual and developmental disability: a population-based study. *Diabetic Medicine, 32*(2), 235-242.

Baranowski, T., Dworkin, R. J., Cieslik, C. J., Hooks, P., Clearman, D. R., Ray, L.,…Nader, P. (1984). Reliability and validity of self-report of aerobic activity: family health project. *Research Quarterly for Exercise and Sport, 55*(4), 309-317.

Baranowski, T., Masse, L. C., Ragan, B., & Welk, G. (2008). How many days was that? We're still not sure, but we're asking the question better! *Medicine and Science in Sports and Exercise*, *40*(7), S544-S549.

Bassett, D. R., Rowlands, A., & Trost, S. G. (2012). Calibration and validation of wearable monitors. *Medicine and Science in Sport and Exercise, 44*(1), S32-S38.

Baynard, T., Pitetti, K. H., Guerra, M., Unnithan, V. B., & Fernhall, B. (2008). Age-related changes in aerobic capacity in individuals with mental retardation: a 20-yr review. *Medicine and Science in Sports and Exercise, 40*(11), 1984-1989.

Beets, M. W., & Pitetti, K. H. (2011). Using pedometers to measure moderate-to-vigorous physical activity for youth with an intellectual disability. *Disability and Health Journal*, *4*(1), 46-51.

Beets, M. W., Combs, C., Pitetti, K. H., Morgan, M., Bryan, R. R., & Foley, J. T. (2007). Accuracy of pedometer steps and time for youth with disabilities. *Adapted Physical Activity Quarterly*, *24*(3), 228-244.

Bertelli, M. O., Carulla, L. S., Scuticchio, D., Varrucciu, N., Leal, R. M., Cooper, S-A.,…Walsh, C. (2014). Moving beyond intelligence in the revision of ICD-10: specific cognitive functions in intellectual developmental disorders. *World Psychiatry, 13*(1), 93-94.

Biddle, S. J. H., & Asare, M. (2011). Physical activity and mental health in children and adolescents: a review of reviews. *British Journal of Sports Medicine, 45*, 886-895.

Biddle, S. J. H., Gorely, T., Marshall, S. J., Murdey, I., & Cameron, N. (2004). Physical activity and sedentary behaviours in youth: issues and controversies. *The Journal of the Royal Society for the Promotion of Health, 124*(1), 29-33.

Biddle, S. J. H., Mutrie, N., & Gorely, T. (2015). *Psychology of physical activity: determinants, well-being and interventions*. London: Routledge.

Biddle, S. J. H., Sallis, J., & Cavill, N. (1998). *Young and active: physical activity guidelines for young people in the UK*. London: Health Education Authority.

Bjornson, K. F. (2005). Physical activity monitoring in children and youth. *Pediatric Physical Therapy, 17*, 37-45.

Bluechardt, M. H., & Shephard, R. J. (1995). Using an extracurricular physical activity program to enhance social skills. *Journal of Learning Disabilities, 28*(3), 160-169.

Boddy, L. M., Downs, S. J., Knowles, Z. R., & Fairclough, S. J. (2015). Physical activity and play behaviours in children and young people with intellectual disabilities: a cross-sectional observational study. *School Psychology International, 36*(2), 154-171.

Bois, J. E., Sarrazin, P. G., Brustad, R. J., Trouilloud, D. O., & Cury, F. (2005). Elementary school children's perceived competence and physical activity involvement: the influence of parent's role modelling behaviours and perceptions of their child's competence. *Psychology of Sport and Exercise, 6*(4), 381-397.

Borremans, E., Rintala, P., & McCubbin, J. A. (2010). Physical fitness and physical activity in adolescents with Asperger Syndrome: a comparative study. *Adapted Physical Activity Quarterly, 27*(4), 308-320.

Bouten, C. V., Sauren, A. A., Verduin, M., & Janssen, J. D. (1997). Effects of placement and orientation of body-fixed accelerometers on the assessment of energy expenditure during walking. *Medical and Biological Engineering and Computing*, *35*(1), 50-56.

Bradley, E. A., Thompson, A., & Bryson, S. E. (2002). Mental retardation in teenagers; prevalence data from the Niagara region, Ontario. *Canadian Journal of Psychiatry, 47*(7), 652-659.

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.

Brage, S., Wedderkopp, N., Andersen, L. B., & Froberg, K. (2003). Influence of step frequency on movement intensity predictions with the CSA accelerometer: a field validation study in children. *Pediatric Exercise Science*, *15*(3), 277-287.

Brown, W. H., Pfeiffer, K. A., McIver, K. L., Dowda, M., Almeida, J. M., & Pate, R. R. (2006). Assessing preschool children's physical activity: the observational system for recording physical activity in children-preschool version. *Research Quarterly for Exercise and Sport*, *77*(2), 167-176.

Bull, F. C., Biddle, S., Buchner, D., Ferguson, R., Foster, C., Fox, K.,… & Watts, C. (2010). *Technical report. Physical activity guidelines in the UK: review and recommendations*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213743/dh_128255.pdf

Burchinal, M. R., Roberts, J. E., Hooper, S., & Zeisel, S. A., (2000). Cumulative risk and early cognitive development: a comparison of statistical risk models. *Developmental Psychology, 36*(6), 793-807.

Caballero, B., Clay, T., Davis, S. M., Ethelbah, B., Rock, B. H., Lohman, T., & Stevens, J. (2003). Pathways: a school-based, randomized controlled trial for the prevention of obesity in American Indian schoolchildren. *The American Journal of Clinical Nutrition*, *78*(5), 1030-1038.

Cain, K. L., Sallis, J. F., Conway, T. L., Van Dyck, D., & Calhoon, L. (2013). Using accelerometers in youth physical activity studies: a review of methods. *Journal of Physical Activity & Health, 10*, 437-50.

Capio, C. M., Sit, C. H., & Abernethy, B. (2010). Physical activity measurement using MTI (Actigraph) among children with cerebral palsy. *Archives of Physical Medicine and Rehabilitation, 91*, 1283-1290.

Carnaby, S. (2007). *Learning disability today: key issues for providers, managers, practitioners and users* (2nd Ed.) Brighton: Pavilion Publishing.

Carson, V., Ridgers, N. D., Howard, B. J., Winkler, E. A., Healy, G. N., Owen, N.,...Salmon, J. (2013). Light-intensity physical activity and cardiometabolic biomarkers in US adolescents. *PloS One*, *8*(8), e71417.

Carulla, L. S., Reed, G. M., Vaez-Azizi, L. M., Cooper, S-A., Leal, R. M., Bertelli, M.,…Saxena, S. (2011). Intellectual developmental disorders: towards a new name, definition and framework for "mental retardation/intellectual disability" in ICD-11. *World Psychiatry, 10*(3), 175-180.

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness, definitions and distinctions for health-related research. *Public Health Reports, 100*(2), 126-131.

Catellier, D. J., Hannan, P. J., Murray, D. M., Addy, C. L., Conway, T. L., Yang, S., & Rice, J. C. (2005). Imputation of missing data when measuring physical activity by accelerometry. *Medicine and Science in Sports and Exercise, 37*(11), S555-S562.

Cavagna, G. A., Franzetti, P., & Fuchimoto, T. (1983). The mechanics of walking in children. *The Journal of Physiology*, *343*(1), 323-339.

Cavill, N., Biddle, S., & Sallis, J. F. (2001). Health enhancing physical activity for young people: statement of the United Kingdom Expert Consensus Conference. *Pediatric Exercise Science*, *13*(1), 12-25.

Cervantes, C. M., & Porretta. D. L. (2010). Physical activity measurement among individuals with disabilities: a literature review. *Adapted Physical Activity Quarterly, 27*(3), 173-190.

Chalmers, I., Hedges, L. V., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation & the Health Professions, 25*(1), 12-37.

Chaput, J. P., Carson, V., Gray, C. E., & Tremblay, M. S. (2014). Importance of all movement behaviors in a 24 hour period for overall health. *International Journal of Environmental Research and Public Health*, *11*(12), 12575-12581.

Chen, K. Y., & Bassett, D. R. (2005). The technology of accelerometry-based monitors: current and future. *Medicine and Science in Sports and Exercise, 37*(11), S490-S500.

Chief Medical Officer. (2009). *Annual report of the Chief Medical Officer*. Retrieved from http://www.sthc.co.uk/Documents/CMO_Report_2009.pdf

Chief Medical Officers. (2011). *Start active, stay active: a report on physical activity for health from the four home countries' Chief Medical Officers*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216370/d h_128210.pdf

Chinapaw, M. J. M., Proper, K. I., Brug, J., van Mechelen, W., & Singh, A. S. (2011). Relationship between young peoples' sedentary behaviour and biomedical health indicators: a systematic review of prospective studies. *Obesity Reviews, 12*(7), 621-632.

Chu, B. B., Lawson, D., & Naughton, G. (2003). Validation of the Computer Science and Applications (CSA) activity monitor as an objective measure of activity energy expenditure in Vietnamese adolescents. *Pediatric Exercise Science, 15*(1), 56-66.

Clanchy, K. M., Tweedy, S. M., Boyd, R. N., & Trost, S. G. (2011). Validity of accelerometry in ambulatory children and adolescents with cerebral palsy. *European Journal of Applied Physiology*, *111*(12), 2951-2959.

Cleaver, S., Ouellette-Kuntz, H., & Sakar, A. (2010). Participation in intellectual disability research: a review of 20 years of studies. *Journal of Intellectual Disability Research, 54*(3), 187-193.

Clemes, S., Edwardson, C., Connelly, J., Konstantinidis, T., Koivula, R., Yates, T.,… & Biddle, S. (2012). Validity of the ActiGraph GT3X+ inclinometer and different counts per minute cut-points for the assessment of sedentary behaviour. *Journal of Science and Medicine in Sport*, *15*, S68.

Cliff, D. P., Reilly, J. J., & Okley, A. D. (2009). Methodological considerations in using accelerometers to assess habitual physical activity in children aged 0–5 years. *Journal of Science and Medicine in Sport, 12*, 557-567.

Coe, D., & Pivarnik, J. M. (2001). Validation of the CSA accelerometer in adolescent boys during basketball practice. *Pediatric Exercise Science, 13*(4), 373-379.

Cohen, J. (1997). *Statistical power analysis for the behavioural sciences*. New York: Academic Press.

Compher, C., Frankenfield, D., Kim, N., Roth-Yousey, L. (2006). Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review. *Journal of the American Dietetic Association, 106*(6), 881-903.

Corbin, C. B., Pangrazzi, R. P., & Welk, G. J. (1994). Toward an understanding of appropriate physical activity levels for youth. President's Council of physical activity and sport. *Physical Activity and Fitness Research Digest, 1*(8), 1-8.

Corder, K., Brage, S., Mattocks, C., Ness, A., Riddoch, C., Wareham, N. J., & Ekelund, U. (2007). Comparison of two methods to assess PAEE during six activities in children. *Medicine and Science in Sports and Exercise, 39*(12), 2180-2188.

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.

Croen, L. A., Grether, J. K., & Selvin, S. (2001). The epidemiology of mental retardation of unknown cause. *Pediatrics, 107*(6), e86.

Cronbach, L. (1971). Test validation. In R. Thorndike (Ed.), *Educational measurement* (pp. 443-507). Washington, DC: American Council on Education.

Crouter, S. E., Schneider, P. L., & Bassett, D. R. (2005). Spring-levered versus piezo-electric pedometer accuracy in overweight and obese adults. *Medicine and Science in Sports and Exercise, 37*, 1673-1679.

Curry, C. J., Stevenson, R. E., Aughton, D., Byrne, J., Carey, J. C., Cassidy, S.,…Opitz, J. (1997). Evaluation of mental retardation: recommendations of a consensus conference. *American Journal of Medical Genetics, 72*(4), 468-477.

Davies, C. T. M. (1980). Metabolic cost of exercise and physical performance in children with some observations on external loading. *European Journal of Applied Physiology and Occupational Physiology*, *45*(2-3), 95-102.

De Bock, F., Menze, J., Becker, S., Litaker, D., Fischer, J., & Seidel, I. (2010). Combining accelerometry and HR for assessing preschoolers' physical activity. *Medicine and Science in Sports & Exercise, 42*(12), 2237-2243.

De Decker, E., De Craemer, M., Santos-Lozano, A., Van Cauwenberghe, E., De Bourdeaudhuij, I., & Cardon, G. (2013). Validity of the ActivPAL™ and the ActiGraph monitors in pre-schoolers. *Medicine and Science in Sports and Exercise, 45*(10), 2002-2011.

de Vries, S. I., Bakker, I., Hopman-Rock, M., Hirasing, R. A., & van Mechelen, W. (2006). Clinimetric review of motion sensors in children and adolescents. *Journal of Clinical Epidemiology*, *59*(7), 670-680.

DeJaeger, D., Willems, P. A., & Heglund, N. C. (2001). The energy cost of walking in children. *European Journal of Physiology*, *441*(4), 538-543.

Dencker, M., & Andersen, L. B. (2011). Accelerometer-measured daily physical activity related to aerobic fitness in children and adolescents. *Journal of Sports Sciences*, *29*(9), 887-895.

Dobbins, M., Husson, H., DeCorby, K., & LaRocca, R. L. (2013). School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18. The *Cochrane Database of Systematic Reviews, 2013*(2), 1-259.

Durstine, J. L., & Moore, G. E. (2003). *ACSM's exercise management for persons with chronic disease and disabilities* (2nd ed.). Champaign, IL: Human Kinetics.

Egger, M., Smith, G. D., & Altman, D. (2001). *Systematic reviews in health care: meta-analysis in context* (2nd ed.). London: BMJ Books.

Eiholzer, U., Nordmann, Y., L'allemand, D., Schlumpf, M., Schmid, S., & Kromeyer-Hauschild, K. (2003). Improving body composition and physical activity in Prader-Willi Syndrome. *The Journal of Pediatrics, 142*(1), 73-78.

Einarsson, I. O., Olafsson, A., Hinriksdóttir, G., Jóhannsson, E., Daly, D., & Arngrímsson, S. A. (2015). Differences in physical activity among youth with and without intellectual disability. *Medicine and Science in Sports and Exercise, 47*(2), 411-418.

Ekelund, U., Sardinha, L. B., Anderssen, S. A., Harro, M., Franks, P. W., Brage, S.,…Froberg, K. (2004). Associations between objectively assessed physical activity and indicators of body fatness in 9- to 10-y-old European children: a population-based study from 4 distinct regions in Europe (the European youth heart study). *American Journal of Clinical Nutrition, 80*(3), 584-590.
Esliger, D. W., & Tremblay, M. S. (2007). Physical activity and inactivity profiling: the next generation. *Canadian Journal of Public Health, 98*(2), S195-S207.

Eslinger, D. W. & Tremblay, M. S. (2007). Physical activity and inactivity profiling: the next generation. *Applied Physiology, Nutrition, and Metabolism, 32*(2), S195-S207.

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

Esposito, P. E., MacDonald, M., Hornyak, J. E., & Ulrich, D. A. (2012). Physical activity patterns of youth with Down syndrome. *Intellectual and Developmental Disabilities, 50*(2), 109-119.

Eston, R. G., Rowlands, A. V., & Ingledew, D. K. (1998). Validity of heart rate, pedometry, and accelerometry for predicting the energy costs of children's activities. *Journal of Applied Physiology, 84*, 362–371.

Evenson, K. R., Catellier, D. J., Gill, K., Ondrak, K. S., & McMurray, R. G. (2008). Calibration of two objective measures of physical activity for children. *Journal of Sports Sciences, 26*(14), 1557-1565.

Faison-Hodge, J., & Porretta, D. L. (2004). Physical activity levels of students with mental retardation and students without disabilities. *Adapted Physical Activity Quarterly*, *21*(2), 139-152.

Fedewa, A. L., & Ahn, S. (2011). The effects of physical activity and physical fitness on children's achievement and cognitive outcomes: a meta-analysis. *Research Quarterly for Exercise and Sport, 82*(3), 521-535.

Fernhall, B. (1993). Physical fitness and exercise training of individuals with mental retardation. *Medicine & Science in Sports & Exercise, 25*(4), 442–450.

Fernhall, B., & Tymeson, G. (1987). Graded exercise testing of mentally retarded adults: a study of feasibility. *Archives of Physical Medicine and Rehabilitation, 68*(6), 363-367.

Fernhall, B., McCubbin, J. A., Pitetti, K. H., Rintala, P., Rimmer, J. H., Millar, A. L., & De Silva, A. (2001). Prediction of maximal heart rate in individuals with mental retardation. *Medicine and Science in Sports and Exercise*, *33*(10), 1655-1660.

Fernhall, B., Millar, A. L., Pitetti, K. H., Hensen, T., & Vukovich, M. D. (2000). Cross validation of the 20-m shuttle run test for children and adolescents with mental retardation. *Adapted Physical Activity Quarterly*, *17*(4), 402-412.

Fernhall, B., Pitetti, K., Stubbs, N., & Stadler, L. (1996). Validity and reliability of the 1/2 mile run-walk as an indicator of aerobic fitness in children with mental retardation. *Pediatric Exercise Science, 8*, 130-142.

Field, A. (2011). *Discovering statistics using SPSS*. London: Sage.

Finn, K., Johannsen, N., & Specker, B. (2002). Factors associated with physical activity in preschool children. *The Journal of Pediatrics, 140*(1), 81-85.

Foley, J. T., Bryan, R. R., & McCubbin, J. A. (2008). Daily physical activity levels of elementary school-aged children with and without mental retardation. *Journal of Developmental and Physical Disabilities, 20*, 365-378.

Foley, J. T., & McCubbin, J. A. (2009). An exploratory study of after-school behaviour in elementary school-age children with intellectual disability. *Journal of Intellectual and Developmental Disability, 34*(1), 3-9.

Freedson, P. (1991). Electronic motion sensors and heart rate as measures of physical activity in children. *Journal of School Health*, *61*(5), 220-223.

Freedson, P., Bowles, H. R., Troiano, R., & Haskell, W. (2012). Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Medicine and Science in Sports and Exercise, 44*(1), S1-S4.

Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of accelerometer output for children. *Medicine and Science in Sports and Exercise, 37*(11), S523-S530.

Frey, G. C., Stanish, H. I., & Temple, V. A. (2008). Physical activity of youth with intellectual disability: review and research agenda. *Adapted Physical Activity Quarterly, 25*(2), 95-117.

Gando, Y., & Muraoke, I. (2015). Health impact of light-intensity physical activity and exercise. In K. Kanosue, S. Oshima, Z-B. Cao, & K. Oka (Eds.), *Physical activity, exercise, sedentary behaviour and health* (pp. 51-61). Tokyo, Japan: Springer.

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.

Graham, A., & Reid, G. (2000). Physical fitness of adults with an intellectual disability: a 13-year follow-up study. *Research Quarterly for Exercise and Sport*, *71*(2), 152-161.

Grydeland, M., Hansen, B. H., Ried-Larsen, M., Kolle, E., & Anderssen, S. A. (2014). Comparison of three generations of ActiGraph activity monitors under free-living conditions: do they provide comparable assessments of overall physical activity in 9-year old children? *BMC Sports Science, Medicine and Rehabilitation*, *6*(1), 26.

Guinhouya, B. C., Apete, G. K., & Hubert, H. (2009). Diagnostic quality of Actigraph-based physical activity cut-offs for children: what overweight/obesity references can tell? *Pediatrics International, 51*(4), 568-573.

Guinhouya, C. B., & Hubert, H. (2008). Incoherence with studies using actigraph mti among children age 6-12 yr. *Medicine and Science in Sports and Exercise*, *40*(5), 979.

Guinhouya, C. B., Hubert, H., Soubrier, S., Vilhelm, C., Lemdani, M., & Durocher, A. (2006). Moderate-to-vigorous physical activity among children: discrepancies in accelerometry-based cut-off points. *Obesity*, *14*(5), 774-777.

Guinhouya, B. C., Hubert, H. & Zitouni, D. (2011). Need for unbiased computation of the moderate-intensity physical activity of youth in epidemiological studies. *American Journal of Preventive Medicine, 41*(1), e1-e3.

Guralnick, M. J. (2005). Early intervention for children with intellectual disabilities: current knowledge and future prospects. *Journal of Applied Research in Intellectual Disabilities, 18*(4), 313-324.

Hänggi, J. M., Phillips, L. R., & Rowlands, A. V. (2013). Validation of the GT3X ActiGraph in children and comparison with the GT1M ActiGraph. *Journal of Science and Medicine in Sport*, *16*(1), 40-44.

Hallal, P. C., Victora, C. G., Azevedo, M. R., & Wells, J. C. (2006). Adolescent physical activity and health. *Sports Medicine*, *36*(12), 1019-1030.

Harris, J. C. (2006). *Intellectual disability: understanding its development, causes, classification, evaluation, and treatment*. New York: Oxford University Press.

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.

Harris, L., Hankey, C., Murray, H., & Melville, C. (2015). The effects of physical activity interventions on preventing weight gain and the effects on body composition in young adults with intellectual disabilities: systematic review and meta-analysis of randomized controlled trials. *Clinical Obesity*, *5*(4), 198-210.

Heil, D. P., Brage, S., & Rothney, M. P. (2012). Modelling physical activity outcomes from wearable monitors. *Medicine and Science in Sports and Exercise, 44*(1), S50-S60.

Hemingway, P., & Brereton, N. (2009). *What is a systematic review? What is series*. Retrieved from http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/syst-review.pdf

Higgins, J. P. T. & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions.* The Cochrane Collaboration. Retrieved from www.cochrane-handbook.org.

Hinckson, E. A., & Curtis, A. (2013). Measuring physical activity in children and youth living with intellectual disabilities: a systematic review. *Research in Developmental Disabilities, 34*(1), 72-86.

Hinckson, E. A., Dickinson, A., Water, T, Sands, M., & Penman, L. (2013). Physical activity, dietary habits and overall health in overweight and obese children and youth with intellectual disability or autism. *Research in Developmental Disabilities, 34*(4), 1170-1178.

Hislop, J. F., Bulley, C., Mercer, T. H., & Reilly, J. J. (2012). Comparison of epoch and uniaxial versus triaxial accelerometers in the measurement of physical activity in preschool children: a validation study. *Pediatric Exercise Science, 24*(3), 450-460.

Honas, J. J., Washburn, R. A., Smith, B. K., Greene, J. L., Cook-Wiens, G., & Donnelly, J. E. (2008). The System for Observing Fitness Instruction Time (SOFIT) as a measure of energy expenditure during classroom-based physical activity. *Pediatric Exercise Science*, *20*(4), 439.

Hoos, M. B., Plasqui, G., Gerver, W. J. M., & Westerterp, K. R. (2003). Physical activity level measured by doubly labeled water and accelerometry in children. *European Journal of Applied Physiology, 89*(6), 624-626.

Horvat, M., & Franklin, C. (2001). The effects of the environment on physical activity patterns of children with mental retardation. *Research Quarterly for Exercise and Sport, 72*(2), 189-195.

Howie, E. K., & Pate, R. P. (2012). Physical activity and academic achievement in children: a historical perspective. *Journal of Sport and Health Sciences, 1*(3), 160-169.

Howley, E. T., Bassett JR, D. R., & Welch, H. G. (1995). Criteria for maximal oxygen uptake: review and commentary. *Medicine and Science in Sports and Exercise, 27*(9), 1292-1301.

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., & 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), 1-16.

Janssen, X., Cliff, D., Okely, T., Hinkley, T., Reilly, J. J., Jones, R., Batterham, M.,… Brage, S. (2013). Comparison of ActiGraph cut-points for predicting physical activity intensity in preschool children. *Journal of Science and Medicine in Sport, 15,* S66.

Janz, K. F. (1994). Validation of the CSA accelerometer for assessing children's physical activity. *Medicine & Science in Sports & Exercise, 26*(3), 369-375.

Janz, K. F. (2002). Use of heart rate monitors to assess physical activity. In G. Welk (Ed.), *Physical activity assessments for health-related research* (pp. 143-161). Champaign, IL: Human Kinetics.

Janz, K. F., Witt, J., & Mahoney, L. T. (1995). The stability of children's physical activity as measured by accelerometry and self-report. *Medicine & Science in Sports & Exercise, 27*(9), 1326-1332.

Jimmy, G., Seiler, R., & Mäder, U. (2013). Development and validation of GT3X accelerometer cut-off points in 5-to 9-year-old children based on indirect calorimetry measurements. *Schweizerische Zeitschrift für Sportmedizin und Sporttraumatologie*, *61*(4), 37-43.

John, D., & Freedson, P. (2012). ActiGraph and Actical physical activity monitors: a peek under the hood. *Medicine and Science in Sports and Exercise, 44*(1), S86-S89.

John, D., Tyo, B., & Bassett, D. R. (2010). Comparison of four Actigraph accelerometers during walking and running. *Medicine and Science in Sports and Exercise, 42*(2), 368-374.

Johnson, C. C. (2009). The benefits of physical activity for youth with developmental disabilities: a systematic review. *American Journal of Health Promotion, 23*(3), 157-167.

Johnston, T. E., Moore, S. E., Quinn, L. T., & Smith, B. T. (2004). Energy cost of walking in children with cerebral palsy: relation to the Gross Motor Function Classification System. *Developmental Medicine & Child Neurology*, *46*(01), 34-38.

Kahan, D., Nicaise, V., Reuben, K. (2013). Convergent validity of four accelerometer cutpoints with direct observation of preschool children's outdoor physical activity. *Research Quarterly for Exercise and Sport, 84*(1), 59-67.

Katch, V. L., McArdle, W. D., & Katch, F. I. (2011). *Essentials of exercise physiology*. Philadelphia: Lippincottt Williams & Wilkins.

Keating, X. D., Kulinna, P. H., & Silverman, S. (1999). Measuring teaching behaviors, lesson context, and physical activity in school physical education programs: comparing the SOFIT and the C-SOFIT instruments. *Measurement in Physical Education and Exercise Science*, *3*(4), 207-220.

Kelly, L. A. (2005). *Objectively measured physical activity and sedentary behaviour in young children* (unpublished doctoral dissertation). University of Glasgow, UK.

Kelly, L. A., Reilly, J. J., Fairweather, S. C., Barrie, S., Grant, S., & Paton, J. Y. (2004). Comparison of two accelerometers for assessment of physical activity in preschool children. *Pediatric Exercise Science*, *16*(4), 324-333.

Kim, Y., Beets, M. W., & Welk, G. J. (2012). Everything you wanted to know about selecting the "right" Actigraph accelerometer cut-points for youth, but…: a systematic review. *Journal of Science and Medicine in Sport*, *15*(4), 311-321.

King, B. H., Toth, K. E., Hodapp, R. M., & Dykens, E. M. (2009). *Intellectual disability*. In B. J. Saddock, V. A. Saddock, & P. Ruiz (Eds.). *Comprehensive textbook of Psychiatry* (pp. 3444-3474). Philadelphia: Lippincott Williams & Wilkins.

Kochersberger, G., McConnell, E., Kuchibhatla, M. N., & Pieper, C. (1996). The reliability, validity, and stability of a measure of physical activity in the elderly. *Archives of physical medicine and rehabilitation*, *77*(8), 793-795.

Kohl, H. W., Fulton, J. E., & Caspersen, C. J. (2000). Assessment of physical activity among children and adolescents: a review and synthesis. *Preventive Medicine, 31*, S54-S76.

Kozub, F. M. (2003). Explaining physical activity in individuals with mental retardation: an exploratory study. *Education and Training in Developmental Disabilities, 38*(3), 302-313.

Krzanowski, W. J., & Hand, D. J. (2009). *ROC curves for continuous data*. London: Taylor & Francis.

Lafleur, M., Strongin, S., Cole, B. L., Bullock, S. L., Banthia, R., Craypo, L.,…Garcia, R. (2013). Physical education and student activity: evaluating implementation of a new policy in Los Angeles public schools. *Annals of Behavioral Medicine, 45*(1), S122-S130.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

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

Lee, S. J., & Hidler, J. (2008). Biomechanics of overground vs. treadmill walking in healthy individuals. *Journal of Applied Physiology, 104*(3), 747-755.

Lennox, N., Taylor, M., Rey-Conde, T., Bain, C., Purdie, D. M., Boyle, F. (2005). Beating the barriers: recruitment of people with intellectual disability to participate in research. *Journal of Intellectual Disability Research, 49*(4), 296-305.

Leonard, H., & Wen, X. (2002). The epidemiology of mental retardation; challenges and opportunities in the new millennium. *Mental Retardation and Developmental Disabilities Research Reviews, 8*(3), 117-134.

Leung, M. M., Agaronov, A., Grytsenko, K., & Yeh, M. C. (2011). Intervening to reduce sedentary behaviors and childhood obesity among school-age youth: a systematic review of randomized trials. *Journal of Obesity*, *2012*, 14pp.

Lin, J. D., Lin, P. Y., Lin, L. P., Chang, Y. Y., Wu, S. R., & Wu, J. L. (2010). Physical activity and its determinants among adolescents with intellectual disabilities. *Research in Developmental Disabilities, 31*(1), 263-269.

Linacre, J. M. (2000). New approaches to determining reliability and validity. *Research Quarterly for Exercise and Sport, 71*(2), 129-136.

Lloyd, M. C. (2008). *Physical activity of preschool age children with and without Down syndrome* (Unpublished doctoral dissertation). University of Michigan, MI, USA.

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

Lorenzi, D. G., Horvat, M., & Pellegrini, A. D. (2000). Physical Activity of children with and without mental retardation in inclusive recess settings. *Education and Training in Developmental Disabilities, 35*, 160–167.

MacDonald, M., Esposito, P., & Ulrich, D. (2011). The physical activity patterns of children with autism. *BMC Research Notes, 4*, 422.

Macfarlane, D. J. (2001). Automated metabolic gas analysis systems. *Sports Medicine, 31*(12), 841-861.

Mackintosh, K. A., Fairclough, S. J., Stratton, G., & Ridgers, N. D. (2012) A calibration protocol for population-specific accelerometer cut-points in children. *PLoS ONE 7*(5): e36919.

Mahar, M. T., & Rowe, D. A. (2002). Construct validity in physical activity research. In G. J. Welk (Ed.), *Physical activity assessments for health-related research* (pp. 51-72). Champaign, IL: Human Kinetics.

Maiano, C. (2010). Prevalence and risk factors of overweight and obesity among children and adolescents with intellectual disabilities. *Obesity Reviews, 12*(3), 189-197.

Maiano, C., Normand, C. L., Aime, A., & Bergarie, J. (2014). Lifestyle interventions targeting changes in body weight and composition among youth with an intellectual disability: a systematic review. *Research in Developmental Disabilities, 35*(8), 1914-1926.

Marshall, S. J., Biddle, S. J., Gorely, T., Cameron, N., & Murdey, I. (2004). Relationships between media use, body fatness and physical activity in children and youth: a meta-analysis. *International Journal of Obesity*, *28*(10), 1238-1246.

Masse, L. C., Fuemmeler, B. F., Anderson, C. B., Matthews, C. E., Trost, S. G., Catellier, D. J., & Treuth, M. (2005). Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. *Medicine and Science in Sports and Exercise*, *37*(11), S544-S554.

Massersmith, N. V., Slifer, K. J., Gomez-Cabello, A., Pullbrook-Vetter, V., & Bellipanni, K. (2008). Interdisciplinary behavioral intervention for life-threatening obesity in an adolescent with Prader-Willi syndrome – a case report. *Journal of Developmental & Behavioral Pediatrics, 29*, 129-134.

Matthews, C. E. (2002). Use of self-report instruments to assess physical activity. In G. Welk (Ed.), *Physical activity assessments for health-related research* (pp. 107-124). Champaign, IL: Human Kinetics.

Matthews, C. E. (2005). Calibration of accelerometer output for adults. *Medicine and Science in Sports and Exercise, 37*(11), S512-S522.

Matthews, C. E., Hagströmer, M., Pober, D. M., & Bowles, H. R. (2012). Best practices for using physical activity monitors in population-based research. *Medicine and Science in Sports and Exercise*, *44*(1), S68-S76.

Mattocks, C., Leary, S., Ness, A., Deere, K., Saunders, J., Tilling, K.,…Riddoch, C. (2007). Calibration of an accelerometer during free-living activities. *International Journal of Pediatric Obesity, 2*, 218-226.

Mattocks, C., Ness, A. R., Leary, S. D., Tilling, K., Blair, S. N., Shield, J.,… & Riddoch, C. J. (2008). 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*(1), S98-S111

Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T., & Saxena, S. (2011). Prevalence of intellectual disability: a meta-analysis of population-based studies. *Research in Developmental Disabilities, 32*(2), 419-436.

McArdle, W. D., Katch, F. I., & Pechar, G. S. (1973). *Comparison of continuous and discontinuous treadmill and bicycle tests for max* VO_2. *Medicine and Science in Sports, 5*(3), 156-160.

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

McClain, J. J., Abraham, T. L., Brusseau Jr, T. A., & 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.

McClave, S. A., Spain, D. A., Skolnick, J. L., Lowen, C. C., Kleber, M. J., Wickerham, P. S., Vogt, J. R., & Looney, S. W. (2003). Achievement of steady state optimizes results when performing indirect calorimetry. *Journal of Parenteral and Enteral Nutrition, 27*(1), 16-20.

McDermott, S., Durkin, M. S., Schupf, N., & Stein, Z. A. (2007). Epidemiology and etiology of mental retardation. In J. W. Jacobson, J. A. Mulick, & J. Rojahn (Eds.), *Handbook of intellectual and developmental disabilities* (pp. 3-40). New York: Springer.

McDonald, M., Esposito, P., & Ulrich, D. (2011). The physical activity patterns of children with autism. *BMC Research Notes, 4*(1), 422.

McGarty, A. M., Penpraze, V., & Melville, C. A. (2014). Accelerometer use during field-based physical activity research in children and adolescents with intellectual disabilities: a systematic review. *Research in Developmental Disabilities, 35*(5), 973-981.

McGarty, A. M., Penpraze, V., & Melville, C. A. (2015). Feasibility of a laboratory-based accelerometer calibration protocol for children with intellectual disabilities. *Pilot and Feasibility Studies, 1*(1), 18.

McKenzie, T. L. (2002). 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.

McKenzie, T. L. (2009). *System for observing fitness instruction time: generic description and procedures manual*. San Diego: San Diego State University.

McKenzie, T. L., Sallis, J. F., & Armstrong, C. A. (1994). Association between direct observation and accelerometer measures of children's physical activity during physical education and recess. *Medicine and Science in Sports and Exercise, 6*(5), S143.

McKenzie, T. L., Sallis, J. F., & Nader, P. R. (1991). SOFIT: system for observing fitness instruction time. *Journal of Teaching in Physical Education, 11*, 195-205.

McMurray, R. G., Ring, K. B., Treuth, M. S., Welk, G. J., Pate, R. R., Schmitz, K. H.,… & Sallis, J. F. (2004). Comparison of two approaches to structured physical activity surveys for adolescents. *Medicine and Science in Sports and Exercise, 36*(12), 2135-2143.

Melville, C. A., Cooper, S. A., Morrison, J., Allan, L., Smiley, E., & Williamson, A. (2008). The prevalence and determinants of obesity in adults with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, *21*(5), 425-437.

Mendonca, G. V., Pereira, F. D., & Fernhall, B. (2010). Reduced exercise capacity in person with Down syndrome: cause, effect, and management. *Therapeutics and Clinical Risk Management, 6*, 601-610.

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

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine, 8*(4), 283-298.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine, 151*(4), 264–269.

Montoye, A. H., Pfeiffer, K. A., Suton, D., & Trost, S. G. (2014). Evaluating the responsiveness of accelerometry to detect change in physical activity. *Measurement in Physical Education and Exercise Science*, *18*(4), 273-285.

Montoye, H. J., Washburn, R., Servais, S., Ertl, A., Webster, J. G., & Nagle, F. J. (1983). Estimation of energy expenditure by a portable accelerometer. *Medicine and Science in Sports and Exercise, 15*(5), 403-407.

Morris, J. N., Kagan A., Pattison, D. C., & Gardiner, M. J. (1966). Incidence and prediction of ischaemic heart-disease in London busmen. *Lancet, 288*(7463), 553-559.

Murphy, S. L. (2009). Review of physical activity measurement using accelerometers in older adults: considerations for research design and conduct. *Preventive Medicine, 48*(2), 108-114.

Must, A., & Tybor, D. J. (2005). Physical activity and sedentary behavior: a review of longitudinal studies of weight and adiposity in youth. *International Journal of Obesity, 29*, S84-S96.

Nilsson, A., Ekelund, U., Yngve, A., & Sjöström, M. (2002). Assessing physical activity among children with accelerometers using different time sampling intervals and placements. *Pediatric Exercise Science, 14*(1), 87-96.

O'Dowd, L. (2003). Ecological fallacy. In R. L. Miller & J. D. Brewer (Eds.), *The AZ of social research: a dictionary of key social science research concepts* (pp. 84-86). London: Sage.

O'hara, N. M., Baranowski, T., Wilson, B. S., Parcel, G. S., & Simons-Morton, B. G. (1989). Validity of the observation of children's physical activity. *Research Quarterly for Exercise and Sport, 60*(1), 42-47.

Oftedal, S., Bell, K. L., Davies, P. S. W., Ware, R. S., & Boyd, R. N. (2014). Validation of accelerometer cut points in toddlers with and without cerebral palsy. *Medicine and Science in Sports and Exercise*, *46*(9), 1808-1815.

Oortwijn, A. W. M. C., Plasqui, G., Reilly, J. J., & Okely, A. D. (2009). Feasibility of an activity protocol for young children in a whole room indirect calorimeter: a proof-of-concept study. *Journal of Physical Activity and Health, 6*, 633-637.

Owen, N., Healy, G. N., Matthews, C. E., & Dunstan, D. W. (2010). Too much sitting: the population-health science of sedentary behavior. *Exercise and Sport Sciences Reviews, 38*(3), 105-113.

Paffenbarger, R. S., Hyde, R. T., Wing, A. L., & Hsieh, C. (1986). Physical activity, all-cause mortality, and longevity of college alumni. *New England Journal of Medicine, 314*(10), 605-613.

Pate, R. R., O'Neill, J. R., & Lobelo, F. (2008). The evolving definition of "sedentary". *Exercise and Sport Sciences Reviews, 36*(4), 173-178.

Penpraze, V., Reilly, J. 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.

Pfeiffer, K. A., McIver, K. L., Dowda, M., Almeida, M. J., & Pate, R. R. (2006). Validation and calibration of the Actical accelerometer in preschool children. *Medicine and Science in Sports and Exercise, 38*(1), 152-157.

Phillips, A. C., & Holland, A. J. (2011). Assessment of objectively measured physical activity levels in individuals with intellectual disabilities with and without Down's syndrome. *PLoS ONE, 6*(12), e28618.

Pillemer, D. B. (1984). Conceptual issues in research synthesis. *Journal of Special Education*, *18(27)*, 27-40.

Pitetti, K. H., Beets, M. W., & Combs, C. (2009). Physical activity levels of children with intellectual disabilities during school. *Medicine and Science in Sports and Exercise, 41*(8), 1580-1586.

Pitetti, K. H., Beets, M. W., & Flaming, J. (2009). Accuracy of pedometer steps and time for youth with intellectual disabilities during dynamic movements. *Adapted Physical Activity Quarterly, 26*, 336-351.

Pitetti, K. H., Jongmans, B., & Fernhall, B. (1999). Feasibility of a treadmill test for adolescents with multiple disabilities. *Adapted Physical Activity Quarterly, 16*(4), 362-371.

Pitetti, K. H., Yarmer, D. A., & Fernhall, B. (2001). Cardiovascular fitness and body composition of youth with and without mental retardation. *Adapted Physical Activity Quarterly, 18*(2), 127–141.

Plasqui, G., Joosen, A. M., Kester, A. D., Goris, A. H., & Westerterp, K. R. (2005). Measuring free-living energy expenditure and physical activity with triaxial accelerometry. *Obesity Research, 13*(8), 1363-1369.

Pope, R. P., Coleman, K. J., Gonzalez, E. C., Barron, F., & Heath, E. M. (2002). Validity of a revised system for observing fitness instruction time (SOFIT). *Pediatric Exercise Science, 14*, 135-146.

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.

Powers, S. K., & Howley, E. T. (2007). *Exercise physiology: theory and application to fitness and performance*. Dubuque: William C. Brown.

Puhl, J., Greaves, K., Hoyt, M., & Baranowski, T. (1990). Children's activity rating scale (CARS): description and calibration. *Research Quarterly for Exercise and Sport, 61*(1), 26–36.

Pulakka, A., Cheung, Y. B., Ashorn, U., Penpraze, V., Maleta, K., Phuka, J. C., & Ashorn, P. (2013). Feasibility and validity of the ActiGraph GT3X accelerometer in measuring physical activity of Malawian toddlers. *Acta Paediatrica*, *102*(12), 1192-1198.

Pulsford, R. M., Cortina-Borja, B. M., Rich, C., Kinnafick, F. E., Dezateux, C., & Griffiths, L. J. (2011). Actigraph accelerometer-defined boundaries for sedentary behaviour and physical activity intensities in 7 year old children. *PLoS One, 6*(8), e21882.

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., 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.

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*(1), 45.

Ried-Larsen, M., Brønd, J. C., Brage, S., Hansen, B. H., Grydeland, M., Andersen, L. B., & Møller, N. C. (2012). Mechanical and free living comparisons of four generations of the Actigraph activity monitor. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 113.

Rich, C., Geraci, M., Griffiths, L., Sera, F., Dezateux, C., & Cortina-Borja, M. (2013). Quality control methods in accelerometer data processing: defining minimum wear time. *PLoS One, 8*(6), e67206.

Riddoch, C. (1998). Relationships between physical activity and physical health in young people. In S. Biddle, J. F. Sallis, & N. A. Cavill (Eds.), *Young and Active?* (pp. 17-48). London: Health Education Authority.

Riddoch, C., & Boreham, C. (1995). The health-related physical activity of children. *Sports Medicine, 19*, 252–266.

Ridely, 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.

Rikli, R. E. (1997). Measurement challenges in assessing special populations: implications for behavioral research in adapted physical activity. *Measurement in Physical Education and Exercise Science, 1*(2), 117-126.

Rimmer, J. H., Rowland, J. L., & Yamaki, K. (2007). Obesity and secondary conditions in adolescents with disabilities: addressing the needs of an underserved population. *Journal of Adolescent Health*, *41*(3), 224-229.

Rintala, P., McCubbin, J. A., & Dunn, J. M. (1995). Familiarization process in cardiorespiratory fitness testing for persons with mental retardation. *Sports Medicine, Training, and Rehabilitation: an international Journal, 6*(1), 15-27.

Robergs, R. A., Dwyer, D., & Astorino, T. (2010). Recommendations for improved data processing from expired gas analysis indirect calorimetry. *Sports Medicine, 40*(2), 95-111.

Robusto, K. M., & Trost, S. G. (2012). Comparison of three generations of ActiGraph activity monitors in children and adolescents. *Journal of Sports Sciences, 30*(13), 1429-1435.

Rodriguez, D.A., Brown, A. L., & Troped, P. J. (2005). Portable global positioning units to complement accelerometry-based physical activity monitors. *Medicine and Science in Sports and Exercise, 37*(11), S572-S581.

Rothney, M. P., Apker, G. A., Song, Y., & Chen, K. Y. (2008). Comparing the performance of three generations of ActiGraph accelerometers. *Journal of Applied Physiology, 105*(4), 1091-1097.

Rowe, D. A. (2011). Back to the future? Algorithms and equipment vs. simplicity and common sense in physical activity measurement. *International Journal of Human Movement Science, 5*(2), 25-45.

Rowe, P. J., Schuldheisz, J. M., & Van der Mars, H. (1997). Measuring physical activity in physical education: validation of the SOFIT direct observation instrument for use with first to eighth grade students. *Pediatric Exercise Science*, *9*(2), 136-149.

Rowe, P. J, van Der Mars, H., Schuldheisz, J., & Fox, S. (2004). Measuring students' physical activity levels: validating SOFIT for use with high-school students. *Journal of Teaching in Physical Education*, *23*(3), 235-251.

Rowland, T. (1994). Children as exercise subjects; their recruitment, care, and nurture. *Pediatric Exercise Science, 6*(4), 205-207.

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

Rowlands, A. V., Ingledew, D. K., & Eston, R. G. (2000). The effect of type of physical activity measure on the relationship between body fatness and habitual physical activity in children: a meta-analysis. *Annals of Human Biology, 27*, 479-497.

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

Rowlands, A. V., Thomas, P. W., Eston, R. G., & Topping, R. (2004). Validation of the RT3 triaxial accelerometer for the assessment of physical activity. *Medicine and Science in Sports and Exercise, 36*(3), 518-524.

Ruotsalainen, H., Kyngäs, H., Tammelin, T., & Kääriäinen, M. (2015). Systematic review of physical activity and exercise interventions on body mass indices, subsequent physical activity and psychological symptoms in overweight and obese adolescents. *Journal of Advanced Nursing*, 71(11), 2461-2477.

Russell, S., Mammen, P., & Russell, P. S. S. (2005). Emerging trends in accepting the term intellectual disability in the world disability literature. *Journal of Intellectual Disabilities, 9*(3), 187-192.

Ryan, J. M., Forde, C., Hussey, J. M., & Gormley, J. (2015). Comparison of patterns of physical activity and sedentary behavior between children with cerebral palsy and children with typical development. *Physical Therapy*, *95*(12), 1609-1616.

Sallis, J. F. (1991). Self-report measures of children's physical activity. *Journal of School Health*, 61, 215–219.

Sallis, J. F., & Patrick, K. (1994). Physical activity for adolescents: consensus statement. *Pediatric Exercise Science, 6*, 302-314.

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), S1-S14.

Sallis, J. F., Buono, M. J., & Freedson, P. S. (1991). Bias in estimating caloric expenditure from physical activity in children. *Sports Medicine, 11*, 203-209.

Salmon, J., & Okely, A. D. (2009). Physical activity in young people – assessment and methodological issues. *Journal of Science and Medicine in Sport, 12*, 513-514.

Santos-Lozano, A., Marín, P. J., Torres-Luque, G., Ruiz, J. R., Lucía, A., & Garatachea, N. (2012). Technical variability of the GT3X accelerometer. *Medical Engineering & Physics*, *34*(6), 787-790.

Schalock, R. L., Borthwick-Duffy, S. A., Bradley, V. J., Buntinx, W. H. E., Coulter, D. L., Craig, E. M.,…Yeager, M. H. (2010). *Intellectual disability: definition, classification, and systems of support*. Washington, DC: American Association on intellectual and developmental disabilities.

Schalock, R. L., Luckasson, R. A., & Shogren, K. A. (2007). The renaming of mental retardation: understanding the change to the term intellectual disability. *Intellectual and Developmental Disabilities, 45*(2), 116-124.

Schoeller, D. A., & Webb, P. (1984). Five-day comparison of the doubly labelled water method with respiratory gas exchange. *American Journal of Clinical Nutrition, 40*, 153-158.

Schofield, W. N. (1985). Predicting basal metabolic rate, new standards and review of previous work. *Human Nutrition – Clinical Nutrition, 39*(1), 5-41.

Scottish National Physical Activity Task Force. (2003). *Let's make Scotland more active: a strategy for physical activity*. Edinburgh: Scottish Executive.

Sekine, M., Tamura, T., Togawa, T., & Fukui, Y. (2000). Classification of waist-acceleration signals in a continuous walking record. *Medical Engineering and Physics*, *22*(4), 285-291.

Shields, N., Dodd, K. J., & Abblitt, C. (2009). Do children with Down syndrome perform sufficient physical activity to maintain good health? A pilot study. *Adapted Physical Activity Quarterly, 26*, 307-320.

Sharma, S. V., Chaung, R. J., Skala, K., & Atteberry, H. (2011). Measuring physical activity in pre-schoolers: reliability and validity of the system for observing fitness instruction time for preschoolers (SOFIT-P). *Measurement in Physical Education and Exercise Science, 5*(4), 257-273.

Shine, R., Perry, A., & Weiss, J. (2012, July). *Health and well-being of a sample of Canadian children with severe/profound/multiple disabilities*. Presented at the 14th World Conference of the International Associated for the Scientific Study of Disabilities, Halifax, NS.

Sibley, B. A., & Etnier, J. L. (2003). The relationship between physical activity and cognitive function: a meta-analysis. *Pediatric Exercise Science, 15*, 243-256.

Singh, N. N., Lancioni, G. E., Singh, A. N., Winton, A. S. W., Singh, J., McAleavey, K. M., Adkins, A. D. (2008). A mindfulness-based health wellness program for an adolescent with Prader-Willi syndrome. *Behaviour Modification, 32*(2), 167-181.

Sirard, J. R., & Pate, R. R. (2001). Physical activity assessment in children and adolescents. *Sports Medicine, 31*, 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, 2*(3), 345–357.

Sit, C. H., McKenzie, T. L., Lian, J. M., & McManus, A. (2008). Activity levels during physical education and recess in two special schools for children with mild intellectual disabilities. *Adapted Physical Activity Quarterly*, *25*(3), 247-259.

Spruijt-Metz, D., Belcher, B., Anderson, D., Lane, C. J., Chou, C. P., Salter-Venzon, D., ... & Weigensberg, M. J. (2009). A high-sugar/low-fiber meal compared with a low-sugar/high-fiber meal leads to higher leptin and physical activity levels in overweight Latina females. *Journal of the American Dietetic Association, 109*(6), 1058-1063.

Stanish, H. I., & Mozzochi, M. (2000). Participation of preschool children with developmental delay during gross motor activity sessions. *Research Quarterly for Exercise and Sport, 71*(1), S111-S112.

Staudenmayer, J., Zhu, W., & Catellier, D. J. (2012). Statistical considerations in the analysis of accelerometry-based activity monitor data. *Medicine and Science in Sports and Exercise*, *44*(1), S61-S67.

Stein, Z., Belmont, L., & Durkin, M. (1987). Mild mental retardation and severe mental retardation compared: experiences in eight less developed countries. *Journal of Medical Science, 44*, 89–96.

Steinbeck, K. S. (2001). The importance of physical activity in the prevention of overweight and obesity in childhood: a review and an opinion. *Obesity Reviews,* *2*(2), 117-130.

Stewart, A., Marfell-Jones, M., Olds, T. & de Ridder, H. (2011). *International standards for anthropometric assessment*. Lower Hutt, New Zealand: International Society for the Advancement of Kinanthropometry.

Strath, S. J., Brage, S., & Ekelund, U. (2005). Integration of physiological and accelerometer data to improve physical activity assessment. *Medicine and Science in Sports and Exercise, 37*(11), S571-S489.

Strauss, R. S., Rodzilsky, D., Burack, G., & Colin, M. (2001). Psychosocial correlates of physical activity in healthy children. *Archives of Pediatrics and Adolescent Medicine*, *155*(8), 897-902.

Stromme, P., & Valvatne, K. (1998). Mental retardation in Norway; prevalence and sub-classification in a cohort of 30037 children born between 1980 and 1985. *Acta Paediatrica, 6*(3), 291-296.

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.

Suzuki, M., Saitch, S., Tasaki, Y., Shimomura, Y., Makishima, R., & Hosoy, N. (1991). Nutritional status and daily physical activity of handicapped students in Tokyo metropolitan schools for deaf, blind, mentally retarded and physically handicapped individuals. *American Journal of Clinical Nutrition, 54*, 1101–1111.

Tanaka, C., Tanaka, S., Kawahara, J., & Midorikawa, T. (2007). Triaxial accelerometry for assessment of physical activity in young children. *Obesity, 15*(5), 1233-1241.

Taylor, C. A., & Yun, J. (2006). Psychometric properties of two systematic observation techniques for assessing physical activity levels in children with mental retardation. *Pediatric Exercise Science, 18*, 446-456.

Telama, R. (2009). Tracking of physical activity from childhood to adulthood: a review. *Obesity Facts*, *2*(3), 187-195.

Telama, R., Yang, X., Viikari, J., Välimäki, I., Wanne, O., & Raitakari, O. (2005). Physical activity from childhood to adulthood: a 21-year tracking study. *American Journal of Preventive Medicine*, *28*(3), 267-273.

Teo-Koh, S. M., & McCubbin, J. A. (1999). Relationship between peak VO_2 and 1-mile walk test performance of adolescents with mental retardation. *Pediatric Exercise Science, 11*(2), 144–157.

Thomas, J. R., Nelson, J. K., & Silverman, S. J. (2005). *Research methods in physical activity* (5th ed.). Champaign, IL: Human Kinetics.

Thomas, S. S., Buckon, C. E., Russman, B. S., Sussman, M. D., & Aiona, M. D. (2011). A comparison of the changes in the energy cost of walking between children with cerebral palsy and able-bodied peers over one year. *Journal of Pediatric Rehabilitation Medicine, 4*(3), 225-233.

Tremblay, M. S., LeBlanc, A. G., Kho, M. E., Saunders, T. J., Larouche, R., Colley, R. C.,… & Gorber, S. C. (2011a). Systematic review of sedentary behaviour and health indicators in school-aged children and youth. *International Journal Behavioral Nutrition and Physical Activity*, *8*(1), 98.

Tremblay, M. S., Warburton, D. E. R., Janssen, I., Paterson, D. H., Latimer, A. E., Rhodes, R. E.,… & Dugga, M. (2011b). New Canadian physical activity guidelines. *Applied Physiology Nutrition and Metabolism, 36*(1), 47-58.

Treuth, M. S., Schmitz, K., Catellier, D. J., McMurray, R. G., Murray, D. M., Almeida, M. J.,… & Pate, R. (2004). Defining accelerometer thresholds for activity intensities in adolescent girls. *Medicine and Science in Sports and Exercise*, *36*(7), 1259-1266.

Treuth, M. S., Sherwood, N. E., Butte, N. F., McClanahan, B.., Obarzanek, E., Zhou, A., ... & Rochon, J. (2003). Validity and reliability of activity measures in African-American girls for GEMS. *Medicine and Science in Sports and Exercise*, *35*(3), 532-539.

Troiano, R. P. (2005). A timely meeting: objective measurement of physical activity. *Medicine and Science in Sports and Exercise*, *37*(11), S487.

Troiano, R. P. & Bucher, D. M. (2012). National guidelines for physical activity. In B. E. Ainsworth & C. A. Macera (Eds.), *Physical activity and public health practice* (pp. 195-210). London: Taylor & Francis.

Trost, S. G. (2007a). Measurement of physical activity in children and adolescents. *American Journal of Lifestyle Medicine, 1*(4), 299–314.

Trost, S. G. (2007b). 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. (2005). Conducting accelerometer-based activity assessments in field-based research. *Medicine and Science in Sports and Exercise, 37*(11), S531-S543.

Trost, S. G., Pate, R. R., Freedson, P. S., Sallis, J. F., & Taylor, W. C. (2000). Using objective physical activity measures with youth: how many days of monitoring are needed? *Medicine and Science in Sports and Exercise*, *32*(2), 426-431.

Trost, S. G., Pate, R. R., Sallis, J. F., Freedson, P. S., Taylor, W. C., Dowda, M., & Sirard, J. (2002). Age and gender differences in objectively measured physical activity in youth. *Medicine and Science in Sports and Exercise*, *34*(2), 350-355.

Trost, S. G., Ward, D. S., Moorehead, S. M., Watson, P. D., Riner, W., & Burke, J. R. (1998). Validity of the computer science and applications (CSA) activity monitor in children. *Medicine and Science in Sports and Exercise*, *30*(4), 629-633.

Trost, S. G., Way, R., & Okely, A. D. (2006). Predictive validity of three ActiGraph energy expenditure equations for children. *Medicine and Science in Sports and Exercise*, *38*(2), 380-387.

Tudor-Locke, C. E., & Myers, A. M. (2001). Methodological considerations for researchers and practitioners using pedometers to measure physical (ambulatory) activity. *Research Quarterly for Exercise and Sport, 72,* 1–12.

Turk, V., Leer, G., Burchell, S., Khattram, S., Corney, R., & Rowlands, G. (2012). Adults with intellectual disabilities and their carers as researchers and participants in a RCT. *Journal of Applied Research in Intellectual Disabilities*, *25*(1), 1-10.

Tyler, K., MacDonald, M., Menear, K. (2014). Physical activity and physical fitness of school-aged children and youth with autism spectrum disorders. *Autism Research and Treatment, 2014,* 1-6.

Ulrich, D. A., Burghardt, A. R., Lloyd, M., Tiernan, C., & Hornyak, J. E. (2011). Physical activity benefits of learning to ride a two-wheel bicycle for children with Down syndrome: a randomized trial. *Physical Therapy, 91*(10), 1463–1473.

Ulrich, B. D., Haehl, V., Buzzi, U. H., Kubo, M., & Holt, K. G. (2004). Modeling dynamic resource utilization in populations with unique constraints:

preadolescents with and without Down syndrome. *Human Movement Science*, *23*(2), 133-156.

Unnithan, V. B., Clifford, C., & Bar-Or, O. (1998). Evaluation by exercise testing of the child with cerebral palsy. *Sports Medicine*, *26*(4), 239-251.

U.S. Department of Health and Human Services. (1996). *Physical activity and health: a report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services.

Vanhelst, J., Beghin, L., Duhamel, A., Bergman, P., Sjöström, M., & Gottrand, F. (2012). Comparison of uniaxial and triaxial accelerometry in the assessment of physical activity among adolescents under free-living conditions: the HELENA study. *BMC Medical Research Methodology*, *12*(1), 26.

Vanhelst, J., Beghin, L., Rasoamanana, P., Theunynck, D., Meskini, T., Iliescu, C.,… Gottrand, F. (2010). Calibration of the RT3 accelerometer for various patterns of physical activity in children and adolescents. *Journal of Sports Sciences, 28*(4), 381-387.

Vanhelst, J., Beghin, L., Turck, D., & Gottrand, F. (2011). New validated thresholds for various intensities of physical activity in adolescents using the Actigraph accelerometer. *International Journal of Rehabilitation Research*, *34*(2), 175-177.

van Mil, E. G., Westerterp, K. R., Kester, A. D., Curfs, L. M., Gerver, W. J., Schrander-Stumpel, C. T., Saris, W. H. (2000). Activity related energy expenditure in children and adolescents with Prader-Willi syndrome. *International Journal of Obesity, 24*(4), 429-434.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Family Medicine*, *37*(5), 360-363.

Walmsley, J. (2004). Involving users with learning difficulties in health improvement: lessons from inclusive learning disability research. *Nursing Inquiry*, *11*(1), 54-64.

Ward, D. S., Evenson, K. R., Vaughan, 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), S582-S588.

Warms, C. (2005). Physical activity measurement in persons with chronic and disabling conditions: methods, strategies, and issues. *Family Community Health, 29*(5), 785-885.

Warolin, J., Carrico, A. R., Whitaker, L. E., Wang, L., Chen, K. Y., Acra, S., & Buchowski, M. S. (2012). Effect of BMI on prediction of accelerometry-based energy expenditure in youth. *Medicine and Science in Sports and Exercise*, *44*(12), 2428-2435.

Warren, J. M., Ekelund, U., Besson, H., Mezzani, A., Geladas, N., & Vanhees, L. (2010). Assessment of physical activity – a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular prevention and rehabilitation. *European Journal of Preventive Cardiology, 17*(2), 127-139.

Waters, E., de Silva Sanigorski, A., Hall, B. J., Brown, T., Campbell, K. J., Gao, Y., ... & Summerbell, C. D. (2011). Interventions for preventing obesity in children (review). *Cochrane Collaboration,* 12, 1-212.

Welk, G. J. (2005). Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Medicine and Science in Sports and Exercise, 37*(11), S501-S511.

Welk, G. J., & Corbin, C. B. (1995). The validity of the Tritrac-R3D activity monitor for the assessment of physical activity in children. *Research Quarterly for Exercise and Sport*, *66*(3), 202-209.

Welk, G. J., Corbin, C. B., & Dale, D. (2000). Measurement issues in the assessment of physical activity in children. *Research Quarterly for Exercise and Sport*, *71*(2), S59-S73.

Welk, G. J., Dale, D., & Schaben, J. (2002). Application of direct observation techniques for the calibration of activity monitors in children. *Research Quarterly for Exercise and Sport, 73*(1), A16-A17.

Welk, G. J., Eisenmann, J. C., Schaben, J., Trost, S. G., & Dale, D. (2007). Calibration of the Biotrainer Pro activity monitor in children. *Pediatric Exercise Science, 19*(2), 145-158.

Welk, G. J., McClain, J., & Ainsworth, B. E. (2012). Protocols for evaluating equivalency of accelerometry-based activity monitors. *Medicine and Science in Sports and Exercise, 44*(1), S39-S49.

Welk, G. J., Schaben, J. A., & Morrow Jr, J. R. (2004). Reliability of accelerometry-based activity monitors: a generalizability study. *Medicine and Science in Sports and Exercise, 36*(9), 1637-1645.

Welk, G. J., Wood, K., & Morss, G. (2003). Parental influences on physical activity in children: an exploration of potential mechanisms. *Pediatric Exercise Science, 15*, 19-33.

Whitt-Glover, M. C., O'Neill, K., Stettler, N. (2006). Physical activity patterns in children with and without Down syndrome. *Pediatric Rehabilitation, 9*(2), 158-164.

World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. Geneva: WHO. Retrieved from http://apps.who.int/classifications/icd10/browse/2015/en#/F70-F79

World Health Organization. (2007). *Atlas: global resources for persons with intellectual disabilities*. Geneva: WHO.

World Health Organization. (2009). *Global health risks: mortality and burden of disease attributable to selected major risks*. Geneva: World Health Organization.

World Health Organization. (2010). *Global recommendations on physical activity for health*. Geneva: World Health Organization.

World Health Organization. (2015). *Physical activity strategy for the WHO European region 2016-2025*. Geneva: World Health Organization.

Yang, C. C., & Hsu, Y. L. (2009). Development of a wearable motion detector for telemonitoring and real-time identification of physical activity. *Telemedicine and e-Health*, *15*(1), 62-72.

Yang, C. C., & Hsu, Y. L. (2010). A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors, 10*(8), 7772-7788.

Yun, J., & Ulrich, D. A. (2002). Estimating measurement validity: a tutorial. *Adapted Physical Activity Quarterly, 19*, 32-47.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39*(4), 561-577.